Ecological Modelling 381 (2018) 46–77

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

A beta distribution-based moment closure enhances the reliability of trait-based aggregate models for natural populations and communities

Toni Klauschies[a,*], Renato Mendes Coutinho[b], Ursula Gaedke[a,c]

[a] Department of Ecology and Ecosystem Modeling, Institute for Biochemistry and Biology, University of Potsdam, Am Neuen Palais 10, D-14469 Potsdam, Germany
[b] Universidade Federal do ABC, Centro de Matemática, Computação e Cognição, 09210-580, Santo André, Brazil
[c] Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), D-14195 Berlin, Germany

ARTICLE INFO

Keywords:
Moment closure
Normal and beta distribution
Skewed and peaked trait distributions
Fitness landscape and frequency-dependent selection
Eco-evolutionary dynamics
Modelling functional diversity

ABSTRACT

Ecological communities are complex adaptive systems that exhibit remarkable feedbacks between their biomass and trait dynamics. Trait-based aggregate models cope with this complexity by focusing on the temporal development of the community's aggregate properties such as its total biomass, mean trait and trait variance. They are based on particular assumptions about the shape of the underlying trait distribution, which is commonly assumed to be normal. However, ecologically important traits are usually restricted to a finite range, and empirical trait distributions are often skewed or multimodal. As a result, normal distribution-based aggregate models may fail to adequately represent the biomass and trait dynamics of natural communities. We resolve this mismatch by developing a new moment closure approach assuming the trait values to be beta-distributed. We show that the beta distribution captures important shape properties of both observed and simulated trait distributions, which cannot be captured by a *Gaussian*. We further demonstrate that a beta distribution-based moment closure can strongly enhance the reliability of trait-based aggregate models. We compare the biomass, mean trait and variance dynamics of a full trait distribution (*FD*) model to the ones of beta (*BA*) and normal (*NA*) distribution-based aggregate models, under different selection regimes. This way, we demonstrate under which general conditions (stabilizing, fluctuating or disruptive selection) different aggregate models are reliable tools. All three models predicted very similar biomass and trait dynamics under stabilizing selection yielding unimodal trait distributions with small standing trait variation. We also obtained an almost perfect match between the results of the *FD* and *BA* models under fluctuating selection, promoting skewed trait distributions and ongoing oscillations in the biomass and trait dynamics. In contrast, the *NA* model showed unrealistic trait dynamics and exhibited different alternative stable states, and thus a high sensitivity to initial conditions under fluctuating selection. Under disruptive selection, both aggregate models failed to reproduce the results of the *FD* model with the mean trait values remaining within their ecologically feasible ranges in the *BA* model but not in the *NA* model. Overall, a beta distribution-based moment closure strongly improved the realism of trait-based aggregate models.

1. Introduction

Ecological communities can respond to altered environmental conditions by adapting their trait composition, which may in turn alter the strength of ecological interactions and thus the selection pressure of the environment. A recently advocated approach to unravel such feedbacks between biomass and trait dynamics is to consider communities as complex adaptive systems (Norberg, 2004; Merico et al., 2009). Building on fundamental theory in evolutionary biology (Lande, 1982; Turelli and Barton, 1994; Day and Proulx, 2004) and spatial ecology (Iwasa et al., 1987; Dieckmann and Law, 1996; Bolker and Pacala, 1997), community ecologists developed trait-based aggregate models in order to focus on key functional attributes of the communities (Wirtz and Eckhardt, 1996; Norberg et al., 2001). Unlike discrete or full trait distribution models that consider the population dynamics of several or all functionally different species separately (Bruggeman and Kooijman, 2007; Tirok and Gaedke, 2010; Acevedo-Trejos et al., 2016), aggregate models describe the temporal dynamics of the community's aggregate properties such as its total biomass, mean trait and trait variance (Tirok et al., 2011; Merico et al., 2014). The mean and variance correspond to the first and second central moments of the corresponding trait distribution, and therefore represent the average strategy and functional

* Corresponding author.
E-mail addresses: tklausch@uni-potsdam.de (T. Klauschies), renato.coutinho@ufabc.edu.br (R.M. Coutinho), gaedke@uni-potsdam.de (U. Gaedke).

https://doi.org/10.1016/j.ecolmodel.2018.02.001
Received 5 April 2017; Received in revised form 30 January 2018; Accepted 1 February 2018
Available online 08 May 2018

diversity of a community. The corresponding reduction in model complexity may enhance the computational efficiency and allow the achievement of analytical results, which may increase the mechanistic understanding of biomass-trait feedbacks in natural systems.

The derivation of an aggregate model from a full trait distribution model involves a moment expansion, which determines how the rates of change of the aggregate properties depend on the shape of the trait distribution (Norberg et al., 2001; Coutinho et al., 2016). In general, the temporal development of lower moments such as the mean and variance depend on higher moments such as the skewness and kurtosis, which quantify how skewed and peaked a distribution is. In order to get self-contained expressions of the aggregate model equations, one has to conduct a moment closure where higher moments are expressed in terms of lower ones (Kuehn, 2015). This is commonly done by assuming a particular shape of the underlying distribution.

In community ecology, previous studies often used a normal distribution-based moment closure approach and thus assumed the shape of the trait distribution to be a *Gaussian* (Terseleer et al., 2014; Acevedo-Trejos et al., 2015; Smith et al., 2016). This assumption implies that trait distributions should be unimodal and symmetric, and that traits are defined on an infinite range, without a lower or upper boundary. The assumption of a unimodal trait distribution seems to be valid under stabilizing selection, in which intermediate trait values are favored over extreme ones. However, recent empirical studies showed that trait distributions are often heavily skewed or multimodal (Havlicek and Carpenter, 2001; Thibault et al., 2011; Downing et al., 2014; Gaedke and Klauschies, 2017). This is likely a result of fluctuating or disruptive selection and questions the strong assumption of normally distributed trait values (Coutinho et al., 2016). Under fluctuating selection, trait distributions are expected to show skewness because different trait values are favored at different times, whereas under disruptive selection extreme trait values are favored over intermediate ones, promoting bimodal trait distributions. In addition, important traits such as body size, reproduction rate or edibility are clearly restricted to a finite range. For example, a prey species cannot be more than entirely edible or inedible, constraining the trait edibility to the range between zero and one (Klauschies et al., 2016). In general, this limited range of trait values will lead to specific correlations between lower and higher moments (Gaedke and Klauschies, 2017): a distribution whose mean ($\bar{x}$) is close to the lower or upper limit will most probably be right- ($S > 0$) or left-skewed ($S < 0$), promoting a negative correlation between S and $\bar{x}$. Furthermore, when the trait range is limited, distributions with high variances (v) have to be rather flat, or even bimodal, leading to low values of kurtosis (K). Conversely, peaked distributions with pronounced tails resulting in high values of K must exhibit low values of v. This results in a negative correlation between K and v. In contrast, a normal distribution presents no such correlations because both S and K are constant, and thus independent of $\bar{x}$ and v. Hence, a *Gaussian* does not capture such properties of natural trait distributions.

Inspired by recent developments in stochastic population biology (Krishnarajah et al., 2005; Clancy and Mendy, 2011), we explore in this study a novel moment closure approach for trait-based aggregate models, by assuming that the shape of the trait distribution follows a beta distribution. In contrast to a normal distribution, the beta distribution is defined only over a finite range with lower and upper limits, and can be unimodal or bimodal, symmetric or skewed, depending on its mean and variance. Moreover, the correlations between its higher and lower moments conform to the pattern described above. We corroborate these qualitative relations by comparing the relationships between lower and higher moments of the beta distribution with those observed for phytoplankton size distributions of the Baltic Sea (Klais et al., 2016a, b) and for simulated trait distributions of a full trait distribution (*FD*) model. The *FD* model comprises the full trait distribution of a prey community that competes for a common carrying capacity and that is grazed upon by three different predator species that are specialized on different parts of the prey's trait spectrum. Hence, the *FD* model entails key features of natural food webs and generates different selection regimes, which determine the shape of the prey community's trait distribution (see Fig. 1). Finally, we show how a beta distribution-based moment closure approach improves trait-based aggregate models by comparing the biomass and trait dynamics of the *FD* model to a beta distribution-based aggregate (*BA*) model and a normal distribution-based aggregate (*NA*) model under different selection regimes.

2. Materials and methods

To evaluate if the beta distribution represents a more suitable basis for a moment closure of trait-based aggregate models than the normal distribution, we compare the relationships between lower and higher central moments of observed phytoplankton size distributions, simulated trait distributions of a full trait distribution (*FD*) model, and the beta distribution. Subsequently, we compare the dynamics of biomass, and mean and variance of the trait distribution under the *FD* model to that of aggregate models based on a beta or normal distribution moment closure.

2.1. Analyzing the shape of observed phytoplankton size distributions

We analyzed the correlations among lower and higher central moments of more than 7000 phytoplankton size distributions of the Baltic Sea (Klais et al., 2016a, b). Size, i.e. cell volume, is the most important trait, influencing many eco-physiological processes of phytoplankton species such as nutrient uptake, maximum growth rates, sedimentation rates, and susceptibility to zooplankton grazing (Weithoff, 2003). The classification of the different phytoplankton species was done according to their individual cell volume or colony size (for details see Klais et al., 2016a, b). Missing values of the size of a phytoplankton species at particular sampling dates or locations were replaced by the average value for that species over all sampling times and locations. We log-transformed the size values of the phytoplankton species to account for their large range (covering about eleven orders of magnitude) and for the inherent skewness in size distributions. This is in line with the general assumption of normal distribution-based aggregate models that size is log-normally distributed (Terseleer et al., 2014; Acevedo-Trejos et al., 2015). To reduce the sampling error, we grouped the biovolume of the species into 85 equidistant size classes on a $\log_2$ scale. To account for individual growth and variability within each species, we assumed that the minimum and maximum size values of each species differed by a factor of 4 prior to $\log_2$-transformation (Gaedke, 1992). Assuming that the species-specific size distribution is log-normal (see Lande, 1982), we allocated 50% of the biovolume to the class corresponding to the species' measured mean size value, x, 20% to the two adjacent classes corresponding to the ranges within $2^{1/2}$ and $2^{-1/2}$ of x and 5% to the classes corresponding to the ranges 2 and 2^{-1} away from x. All subsequent calculations are based on the mean values of the classes, ordered on a $\log_2$ scale.

2.2. Full trait distribution model (FD model)

In order to achieve different selection regimes, we consider a predator-prey model with a set of prey species with continuous trait values, grazed upon by three predators, each feeding on a specific part of the prey's trait range. The rate of change of the different prey (R) and predator (C) species are given by the following set of equations:

$$\frac{\partial R(x)}{\partial t} = r\left(1-\frac{\int_0^1 R(x')dx'}{\kappa}\right)R(x)-\sum_{i=1}^{3}\frac{a_i(x)R(x)C_i}{1+h\int_0^1 a_i(x')R(x')dx'}+I \tag{1}$$

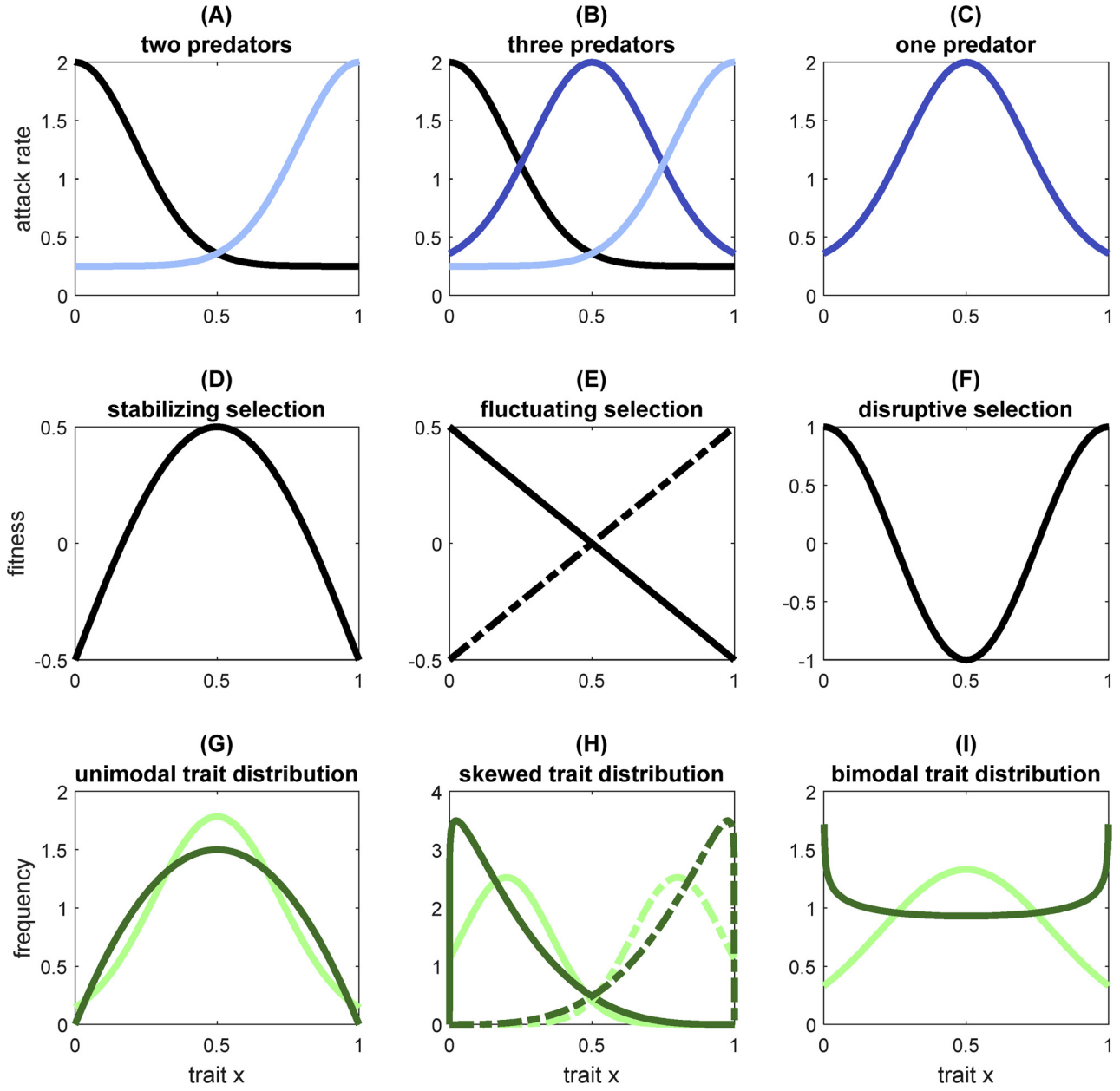

Fig. 1. (A–C) Predator-specific attack rates in dependence of the prey's trait value x. (D–F) Sketch of resulting fitness landscapes of the prey community imposing stabilizing, fluctuating or disruptive selection. (G–I) Expected shape of the beta distribution (dark green) and the normal distribution (light green) in response to selection. The presence or absence of the different predators determines the overall selection regime and the shape of the trait distribution of the prey community. The presence of only two predators specialized on the prey's extreme trait values (A) imposes stabilizing selection on the prey community (D) and thus promotes unimodal trait distributions of the prey community ($\bar{x} = 0.5$; $\nu = 0.05$) (G). With all three predators present (B), fluctuating selection is imposed, in which different trait values are favored at different times depending on the relative importance of the three predators (E), promoting temporally left (dashed) or right (solid) skewed trait distributions ($\bar{x} = 0.2$; $\nu = 0.025$) (H). The presence of a single predator specialized on the intermediate mean trait values (C) provokes disruptive selection (F) and thus bimodal trait distributions ($\bar{x} = 0.5$; $\nu = 0.09$) (I). Note that in contrast to the beta distribution, the shape of the normal distribution does not reflect the expected shape properties of the trait distribution under fluctuating or disruptive selection regimes.

$$\frac{dC_i}{dt} = \left(\frac{\varepsilon \int_0^1 a_i(x) R(x) dx}{1 + h \int_0^1 a_i(x) R(x) dx} - \mathrm{d} \right) C_i \tag{2}$$

The prey are structured by their trait value x, bounded between 0 and 1. The parameters r and κ are respectively the intrinsic growth rate and carrying capacity of the prey. Hence, the prey species are competing for a shared resource such as light. The net growth rates of the three predators, which are identified by $i \in \{1, 2, 3\}$, depend on their common handling time (h), efficiency (ε) and death rate (d), and their species-specific attack rates (a_i). The parameter values used in the simulations are given in Table 1. All parameters are independent of x except for a_i. These are functions of the prey phenotype x (Fig. 1A–C):

$$a_i(x) = (a_{max} - a_{min}) \cdot e^{-\left(\frac{x - \left(\frac{i-1}{2} \right)}{\sigma} \right)^2} + a_{min} \tag{3}$$

with a_{max} and a_{min} being the maximum and minimum attack rates of the different predators. The attack rates of the three different predators are maximal at their species-specific optimal prey phenotypes with trait values of 0, 0.5 or 1. The parameter σ determines the width of the predators'

Table 1
Description of parameters used in the model for the figures. The parameters are estimated by considering the biomasses of phytoplankton (prey) and zooplankton (predators) in units of carbon in the upper most 20 m of the water column of Lake Constance corresponding approximately to the euphotic zone and the epilimnion (Tirok and Gaedke, 2007). Hence, the units g/m^2 of biomasses refer to the biomass in the water column of the top levels (m·g/m^3 = g/m^2).

Parameter	Description	Value	Unit
r	intrinsic growth rate of the prey	2	1/day
κ	carrying capacity of the prey	10	g C/m^2
a_{max}	maximum attack rate of the predators	2	g C/(m^2 day)
a_{min}	minimum attack rate of the predators	0.25	g C/(m^2 day)
h	handling time of the predators	0.5	day
σ	standard deviation of the feeding kernel	0.3	–
ε	conversion efficiency of the predators	0.3	–
d	death rate of the predators	0.2	1/day
I	immigration rate	0.0001	g C/(m^2 day)

feeding kernel. In addition, we added a small immigration rate (I) to Eq. (1) to prevent extinction of any particular trait value in the *FD* model and to promote short transients in its dynamics.

We compared the results of the *FD* model with those of the two different aggregate models by calculating the first three aggregate properties of the prey community's trait distribution, i.e. its total biomass (R_T), trait mean ($\overline{x}$) and trait variance (ν), which are given for the *FD* model by:

$$R_T = \int_0^1 R(x)dx \tag{4}$$

$$\overline{x} = \frac{1}{R_T}\int_0^1 xR(x)dx \tag{5}$$

$$\nu = \frac{1}{R_T}\int_0^1 (x-\overline{x})^2 R(x)dx \tag{6}$$

The *FD* model was numerically integrated using the method of lines, where a system of coupled ordinary differential equations is obtained through a discretization along the trait axis. Hence, we approximated the *FD* model by the following set of equations:

$$\frac{dR_j}{dt} = r\left(1-\frac{\sum_{k=1}^n R_k}{\kappa}\right)R_j - \sum_{i=1}^3 \frac{a_i(x_j)R_jC_i}{1+h\sum_{k=1}^n a_i(x_k)R_k} + \frac{I}{n} \tag{7}$$

$$\frac{dC_i}{dt} = \left(\frac{\varepsilon\sum_{j=1}^n a_i(x_j)R_j}{1+h\sum_{j=1}^n a_i(x_j)R_j} - \mathrm{d}\right)C_i \tag{8}$$

with $n = 101$ being the number of different prey species. Their corresponding trait values are distributed evenly along the trait axis of the entire prey community, i.e. $x_j \in \{0,\ 0.01,\ 0.02, ...,\ 1\}$. Accordingly, the aggregate properties were estimated from the biomass values obtained for the different prey species:

$$R_T = \sum_{j=1}^n R_j \tag{9}$$

$$\overline{x} = \frac{1}{R_T}\sum_{j=1}^n x_jR_j \tag{10}$$

$$\nu = \frac{1}{R_T}\sum_{j=1}^n (x_j-\overline{x})^2R_j \tag{11}$$

2.3. Aggregate models (AM)

Instead of describing the temporal development of the biomasses of different species with particular trait values separately, one can focus on a system of differential equations that describe the rate of change of the aggregate properties directly, using moment approximation methods. We derived the trait-based aggregate models from the corresponding *FD* model as shown in Appendix A. Consequently, the temporal development of the aggregate properties is described by the following set of approximate equations:

$$\frac{dR_T}{dt} = \left(F(\overline{x}) + \frac{\nu}{2}\frac{\partial^2F(x)}{\partial x^2}\bigg|_{x=\overline{x}}\right)R_T + I \tag{12}$$

$$\frac{d\overline{x}}{dt} = \nu\frac{\partial F(x)}{\partial x}\bigg|_{x=\overline{x}} + \frac{S\nu^{\frac{3}{2}}}{2}\frac{\partial^2F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) \tag{13}$$

$$\frac{d\nu}{dt} = S\nu^{\frac{3}{2}}\frac{\partial F(x)}{\partial x}\bigg|_{x=\overline{x}} + \frac{(K\nu^2-\nu^2)}{2}\frac{\partial^2F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{I}{R_T}\left(\frac{1}{12}-\nu+\left(\frac{1}{2}-\overline{x}\right)^2\right) \tag{14}$$

$$\frac{dC_i}{dt} = \left(\frac{\varepsilon\left(a_i(\overline{x}) + \frac{\nu}{2}\frac{\partial^2a_i(x)}{\partial x^2}\Big|_{x=\overline{x}}\right)R_T}{1+h\left(a_i(\overline{x}) + \frac{\nu}{2}\frac{\partial^2a_i(x)}{\partial x^2}\Big|_{x=\overline{x}}\right)R_T} - d\right)C_i \tag{15}$$

with F being the per capita net growth rate (fitness), and S and K being the skewness and the kurtosis of the trait distribution of the prey community. When conducting a moment closure based on the assumption of normally-distributed trait values, Eqs. (13) and (14) simplify to

$$\frac{d\overline{x}}{dt} = \nu\frac{\partial F(x)}{\partial x}\bigg|_{x=\overline{x}} + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) \tag{16}$$

$$\frac{d\nu}{dt} = \nu^2\frac{\partial^2F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{I}{R_T}\left(\frac{1}{12}-\nu+\left(\frac{1}{2}-\overline{x}\right)^2\right) \tag{17}$$

because S equals zero and K equals 3. Hence, S and K are independent of the mean $\overline{x}$ and variance ν. In contrast, S and K of the beta distribution strongly vary with both $\overline{x}$ and ν (for details see Appendix A; cf. Fig. 1G–I):

$$S(\overline{x},\nu) = \frac{2(1-2\overline{x})\sqrt{\nu}}{(1-\overline{x})\overline{x}+\nu} \tag{18}$$

$$K(\overline{x},\nu) = \left(\frac{(1-\overline{x})\overline{x}-\nu}{4\nu+2(1-\overline{x})\overline{x}}+1\right)S(\overline{x},\nu)^2 - \frac{6\nu}{2\nu+(1-\overline{x})\overline{x}} + 3 \tag{19}$$

In line with Coutinho et al. (2016) and Eq. (1), the second-order approximation of F is given by:

$$F(\mathrm{x}) = r\left(1-\frac{R_T}{\kappa}\right) - \sum_{i=1}^3 \frac{a_i(x)C_i}{1+h\left(a_i(\overline{x}) + \frac{\nu}{2}\frac{\partial^2a_i(x)}{\partial x^2}_{x=\overline{x}}\right)R_T} \tag{20}$$

In accordance with the *FD* model, the small immigration rate I in the aggregate model equations prevents a complete loss of the standing trait variation in the prey community.

2.4. Numerical simulations and selection regimes

We conducted numerical simulations of our models to compare their biomass and trait dynamics under different selection regimes. The type of selection pressure on the prey community was altered by choosing which of the three predators were present (see Fig. 1). With only the two predators specialized on the two extreme trait values present, the overall selection pressure was stabilizing towards the intermediate trait value of the prey. In contrast, with only one predator specialized on the intermediate trait values present, the prey community exhibits disruptive selection. Finally, if all three predators are present, selection is on average rather neutral, potentially leading to fluctuating and thus recurrent directional selection, in which different trait values are favored at different times. We initialized all simulations with equally abundant predators, $C_i = \frac{1}{N}\frac{\kappa}{10}$ with N being the number of predators, and total biomass and trait variance of the prey community of $\frac{\kappa}{2}$ and

10^{-3}, respectively. To investigate the possibility of multi-stability in our models we ran them for different initial values of the mean trait, i.e. $\bar{x} \in \{0.05,\ 0.07,\ ...,\ 0.93,\ 0.95\}$. We assumed the initial trait distribution of the prey community in the *FD* model to follow the shape of a beta distribution, though that was not relevant for the long-term solutions obtained. All simulations lasted for 10^4 time units.

Simulations and analyses were performed in MATLAB, version 7.13, using solver ode23 for ODEs (The MathWorks Inc., Natick, MA, 2011). We increased the precision of the solver by reducing the absolute and relative tolerances to 10^{-9} and 10^{-7} and the maximum step size to 0.1.

3. Results

3.1. Analyzing the shape of the empirical and simulated trait distributions

The observed phytoplankton size distributions and the simulated trait distributions both revealed strong relationships between lower and higher moments. The skewness (S) and the mean ($\bar{x}$) were strongly negatively correlated (Fig. 2A ($r^2 = 0.41$) and 2C) because the trait distribution tends to become skewed as $\bar{x}$ approaches one of the extreme trait values, for a given value of ν. We also found a strong negative correlation between the log of kurtosis ($ln(K)$) and the log of the variance ($ln(\nu)$) (Fig. 2B ($r^2 = 0.42$) and 2D). In fact, a trait distribution defined on a finite trait range can only exhibit very high variances when its shape is close to the one of a flat or even bimodal distribution, corresponding to small values of K. In contrast, peaked trait distributions with pronounced tails, i.e. distributions with high values of K, will usually have rather low values of ν. These relationships between lower and higher central moments are qualitatively well approximated by a beta distribution (Eqs. (13) and (14); Fig. 2E and F).

3.2. Comparing the biomass and trait dynamics of the FD, BA and NA models

Our model comparison revealed further support for the suitability of a beta distribution-based moment closure. The presence or absence of distinct predator species imposed either stabilizing, fluctuating or disruptive selection on the prey community (Fig. 1). Under stabilizing selection, the predator biomasses and the three aggregate properties of the prey community, i.e. its total biomass, mean trait value and variance, showed stasis in the full trait distribution (*FD*) model at the end of the simulation (Fig. 3A, D and G). The lack of oscillations is explained by the comparably low attack rates of the predators at equilibrium conditions. In line with that, the two aggregate models predicted very similar values of the predator biomasses and the aggregate properties of the prey community (Fig. 3) because the symmetric unimodal trait distribution of the *FD* model exhibited a very small variance ($\nu \approx 2 \cdot 10^{-6}$) that could be captured well by both the beta (*BA*) and the normal (*NA*) distribution-based approaches. This result was independent of the initial conditions. The biomass and trait dynamics of the *FD*, *BA* and *NA* models also agreed very well when the system exhibited predator-prey oscillations, e.g. for $K = 20$ rather than $K = 10$. This is in line with additional results from a chemostat model, in which a trade-off between the prey's nutrient uptake efficiency and its vulnerability to predation imposes stabilizing selection on the trait distribution of the prey community (for details see Appendix B). As expected, the numerical integration of both aggregate models was up to 4 times faster than that of the *FD* model (Table 2). Interestingly, there was no difference between the simulation times of the *BA* model and the *NA* model.

We also observed an almost perfect match between the cyclic biomass and trait dynamics of the *FD* and *BA* models under fluctuating selection, favoring different trait values of the prey community and different predators at different times (Fig. 4). The pronounced predator-prey oscillations were caused by the temporarily high attack rates of the predators. This result was independent of the initial conditions (Fig. 5) and robust to changes in the carrying capacity, minimum attack rate and immigration rate (Appendix C). The *BA* model could not reproduce the biomass and trait dynamics of the *FD* model when the immigration rate was very low (Fig. C3). In contrast to the *BA* model, the results of the *NA* model deviated considerably from the ones of the *FD* model (Fig. 4C, F, I). Additionally, it exhibited several alternative states and thus a high sensitivity to initial conditions. Depending on the initial mean trait value of the prey community, different equilibria for the biomass and trait dynamics were obtained (Fig. 5C, F and I) when either one or two predators went extinct (Fig. 4C). Extinctions of two predators occurred when the equilibrium value of $\bar{x}$ was outside the ecologically feasible range, which occurred frequently. Depending on the strength of the predator-prey interactions at equilibrium conditions, the predator-prey dynamics showed either stasis or oscillations. These results were insensitive to deviations from the exact parameter values (Appendix C). In the *BA* model, $\bar{x}$ always remained within its ecologically feasible range. This result was to be expected because the beta distribution is naturally defined on the finite range between 0 and 1. A more detailed analysis revealed that the boundedness of $\bar{x}$ is realized by a dynamical feedback between $\bar{x}$ and S. In fact, the absolute value of S strongly increased when $\bar{x}$ approached one of the extremes, leading to a strong decline of ν. As a result, the rate of change of $\bar{x}$ strongly slowed down preventing any further change of $\bar{x}$ close to the extremes (for details see Appendix D). This mechanism of self-regulation is lacking in the *NA* model. In line with stabilizing selection, the numerical integration of both aggregate models was about three times faster than that of the *FD* model under fluctuating selection, whereas the simulation times of the *BA* model and the *NA* model were similar (Table 2). The general features of our results, that is, the match between the *FD* and *BA* models but not with the *NA* model, as well as the gains in computation times, do not depend on details of the model structure (Appendix B). We demonstrate this with a species competition model, in which the recurrent changes in an environmental factor impose fluctuating selection on the trait distribution of the community.

Under disruptive selection, both aggregate models failed to match the results of the *FD* model, which predicted a symmetric bimodal trait distribution exhibiting maximum variance (Fig. 6). The *NA* model even failed to reproduce the predator and prey biomasses. The disagreement between the results of the *FD* and *NA* models was to be expected, because a normal distribution cannot represent a bimodal trait distribution that is dominated by two very different strategies (cf. Fig. 1). In contrast, the reason for the mismatch between the results of the *BA* and *FD* models is more intriguing, since the beta distribution is in principle able to capture the shape of bimodal trait distributions (see Fig. 1). Given the symmetry of the feeding interactions in our model, the *BA* model predicted only the biomasses of the prey and predator communities correctly but not $\bar{x}$ and ν of the prey community's trait distribution. This mismatch is only explicable by the restrictions imposed by the second-order approximation of the shape of the fitness landscape, i.e. it is independent of the type of moment closure. In fact, an extended version of the *BA* model that accounted for higher derivatives of the fitness function was able to reproduce the mean and variance but not the biomass dynamics of the prey community of the *FD* model under disruptive selection (for details see Appendix E). Furthermore, and in contrast to the *FD* model, both aggregate models showed bistability, where the prey community was dominated by a species possessing either a high or low $\bar{x}$. As expected, the latter was still within the ecologically feasible range in the *BA* model (Fig. 7E) but not in the *NA* model making it fundamentally inconsistent (Fig. 7F). Hence, the inability of the aggregate models to represent bimodal trait distributions with large variances was compensated by the emergence of two alternative stable states. In case of the *BA* model, the average distribution of these alternative stable states captured the bimodal distribution of the *FD* model well. The numerical integration of both aggregate models was about three times faster than that of the *FD* model (Table 2). The mismatch between the results of the *FD* and aggregate models under disruptive selection was independent of the exact model structure (Appendix B). We show this with a modified version of the predator-prey model used in the main text, in which a trade-off between the

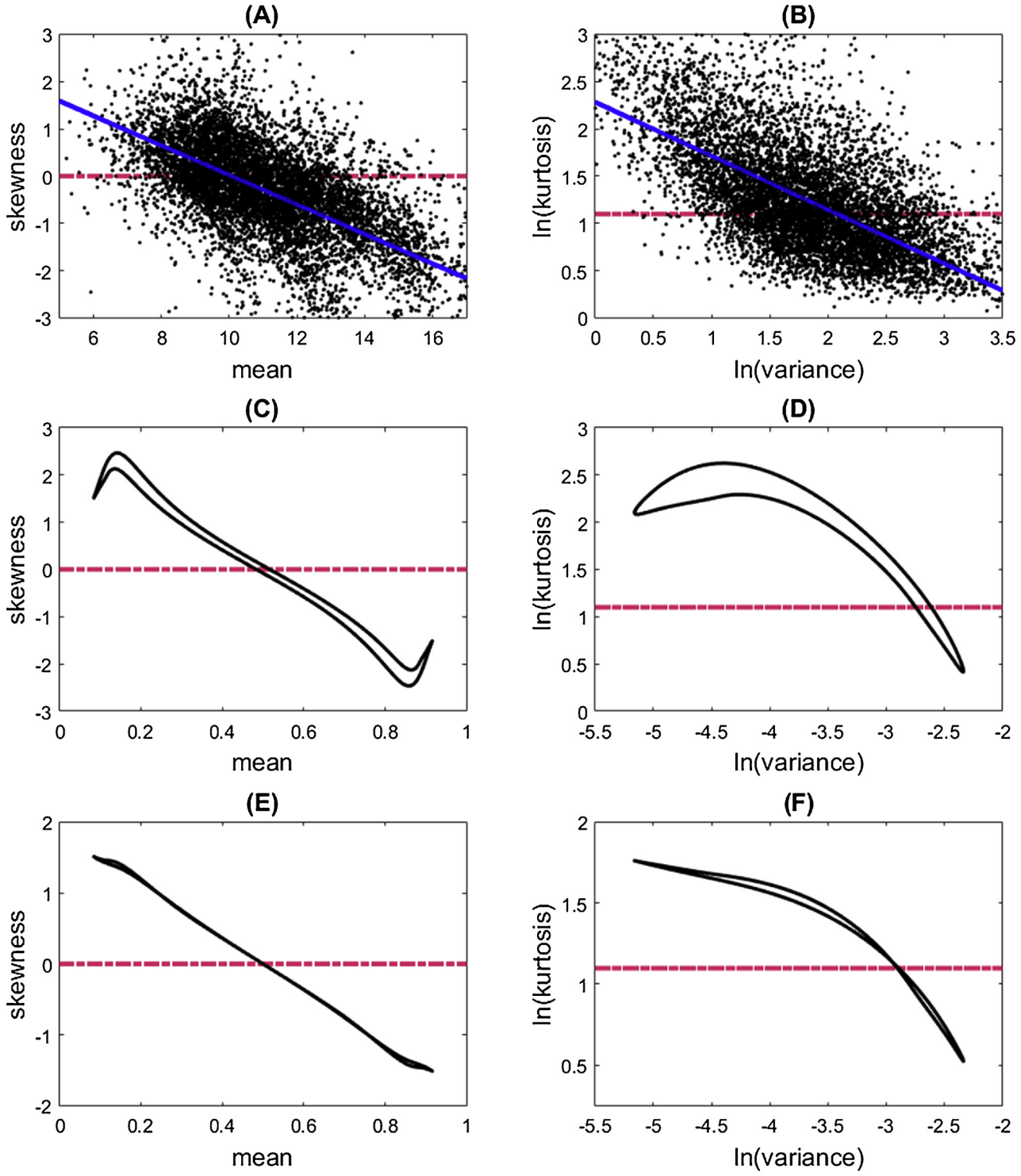

Fig. 2. Skewness (S), i.e. the standardized third central moment, and Kurtosis (K), i.e. the standardized fourth central moment, in dependence of the mean ($\bar{x}$) and variance (v) for the observed phytoplankton size distributions (A, B), the simulated trait distributions of the full trait distribution model (C, D) and the beta distribution (E, F). The red dotted-dashed lines indicate the constant values of $S$and K for the normal distribution. We fitted linear regressions (blue lines) to the empirical data in panels A and B. We enhanced the visibility of the relationships between lower and higher central moments found in the observed phytoplankton size distributions by showing only 95% of the data points in panel A and 96.5% of the data points in panel B. The relationships between the lower and higher central moments of the simulated trait distributions shown in panels C and D are the result of the corresponding simulation of the full trait distribution model shown in Fig. 4A–C. The relationships between the lower and higher central moments of the beta distribution shown in panels E and F follow from Eqs. (18) and (19) of the methods section. We calculated the values of S and K for the beta distribution based on $\bar{x}$ and v values that resulted in the corresponding simulation of the full trait distribution model shown in Fig. 4A–C. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

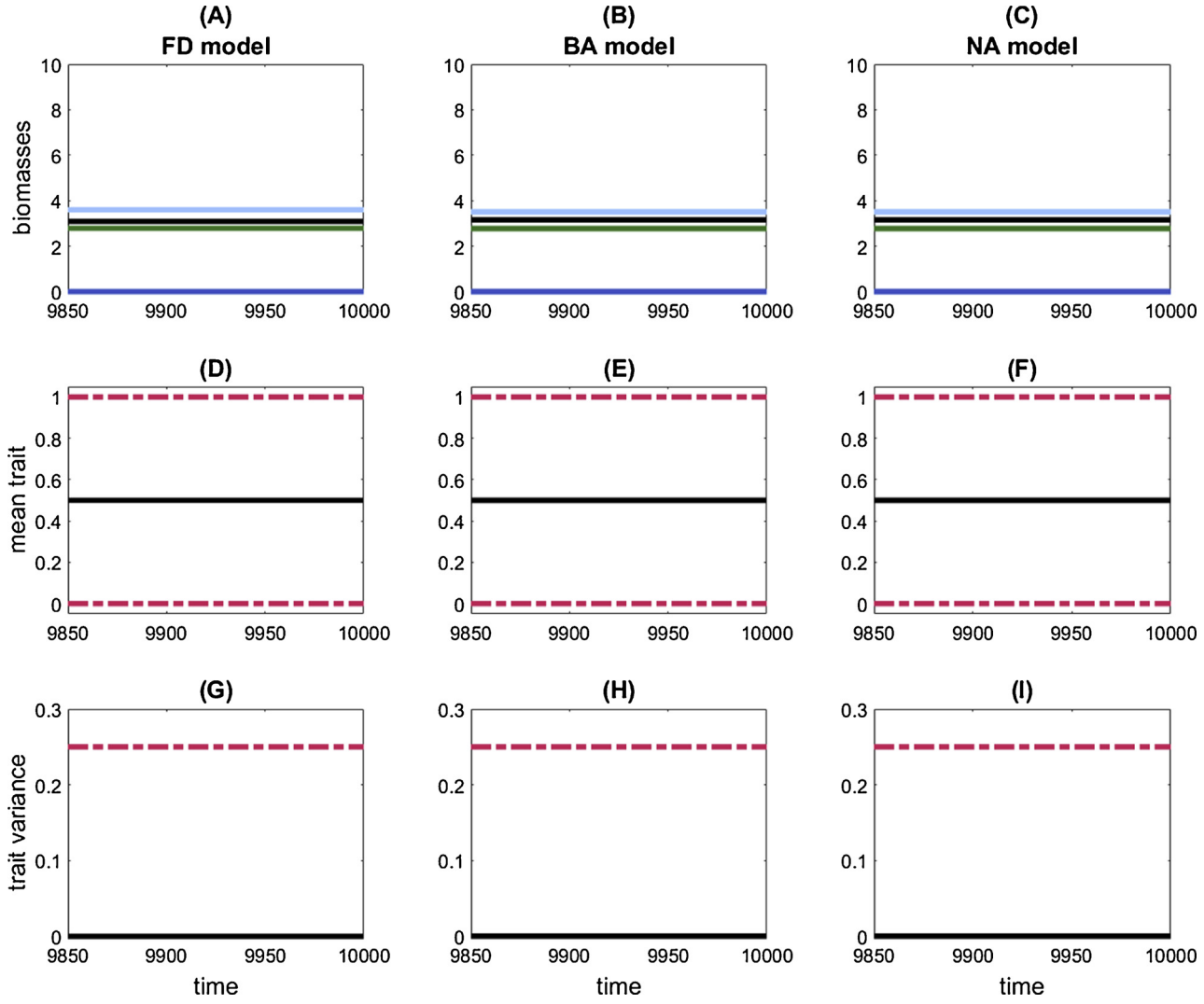

Fig. 3. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (green) and the 3 predator species that are specialized on low (black), intermediate (dark blue) and high (light blue) trait values of the prey community under stabilizing selection. In order to distinguish the lines, we added some biomass to the equilibrium values of the two remaining predators. Results shown for the full trait distribution model (*FD* model, A, D and G), the beta distribution based aggregate model (*BA* model, B, E and H) and the normal distribution based aggregate model (*NA* model, C, F and I). The dashed-dotted red lines in (D)–(I) indicate the ecological feasible minima and maxima of the mean and variance. The initial mean trait value of the prey community was set to 0.25. Note a value of $\nu \approx 0$ implies that the prey community is hardly adaptive anymore.

Table 2
Average simulation time (mean ± standard deviation in seconds, number of repetitions = 10) of the full trait distribution model (*FD*) and normal distribution based (*NA*) and beta distribution based (*BA*) aggregate models for stabilizing (cf. Fig. 3), fluctuating (cf. Fig. 4) and disruptive (cf. Fig. 6) selection when numerically integrating the corresponding model versions for 10,000 time steps.

Selection Pressure	*FD* model	*BA* model	*NA* model
Stabilizing	7.82 ± 0.04	2.30 ± 0.04	2.27 ± 0.05
Fluctuating	38.31 ± 0.12	12.22 ± 0.06	15.58 ± 0.18
Disruptive	8.02 ± 0.54	2.58 ± 0.34	2.34 ± 0.1

prey's vulnerability to predation by the two specialist predators imposes disruptive selection on the trait distribution of the prey community.

4. Discussion

Trait-based aggregate models are a promising tool to incorporate the effects of functional diversity of natural communities into an efficient mathematical framework. The corresponding equations directly inform about the properties of the fitness landscape and trait distribution that drive the temporal development of the aggregate properties and, thus, may enable a detailed understanding of the biomass-trait feedback of complex adaptive systems (CAS) under simplifying assumptions. Their derivation from a corresponding full trait distribution model demonstrates that the temporal development of the lower moments generally depends on higher moments (for details see Appendix A). Obtaining a closed system of differential equations requires a moment closure so that the higher moments are expressed in terms of the lower ones (Norberg et al., 2001; Coutinho et al., 2016). A frequently made assumption is that the higher moments such as skewness (*S*) and kurtosis (*K*) of the trait distribution follow a normal distribution (Wirtz and Eckhardt, 1996; Tirok et al., 2011; Smith et al., 2016). In accordance with previous results (Coutinho et al., 2016), we show that this assumption is likely to be valid under stabilizing or weakly fluctuating selection (Savage et al., 2007; Merico et al., 2009; Acevedo-Trejos et al., 2015), promoting rather unimodal trait distributions with small variances. However, observed size distributions of phytoplankton communities in various habitats were often heavily

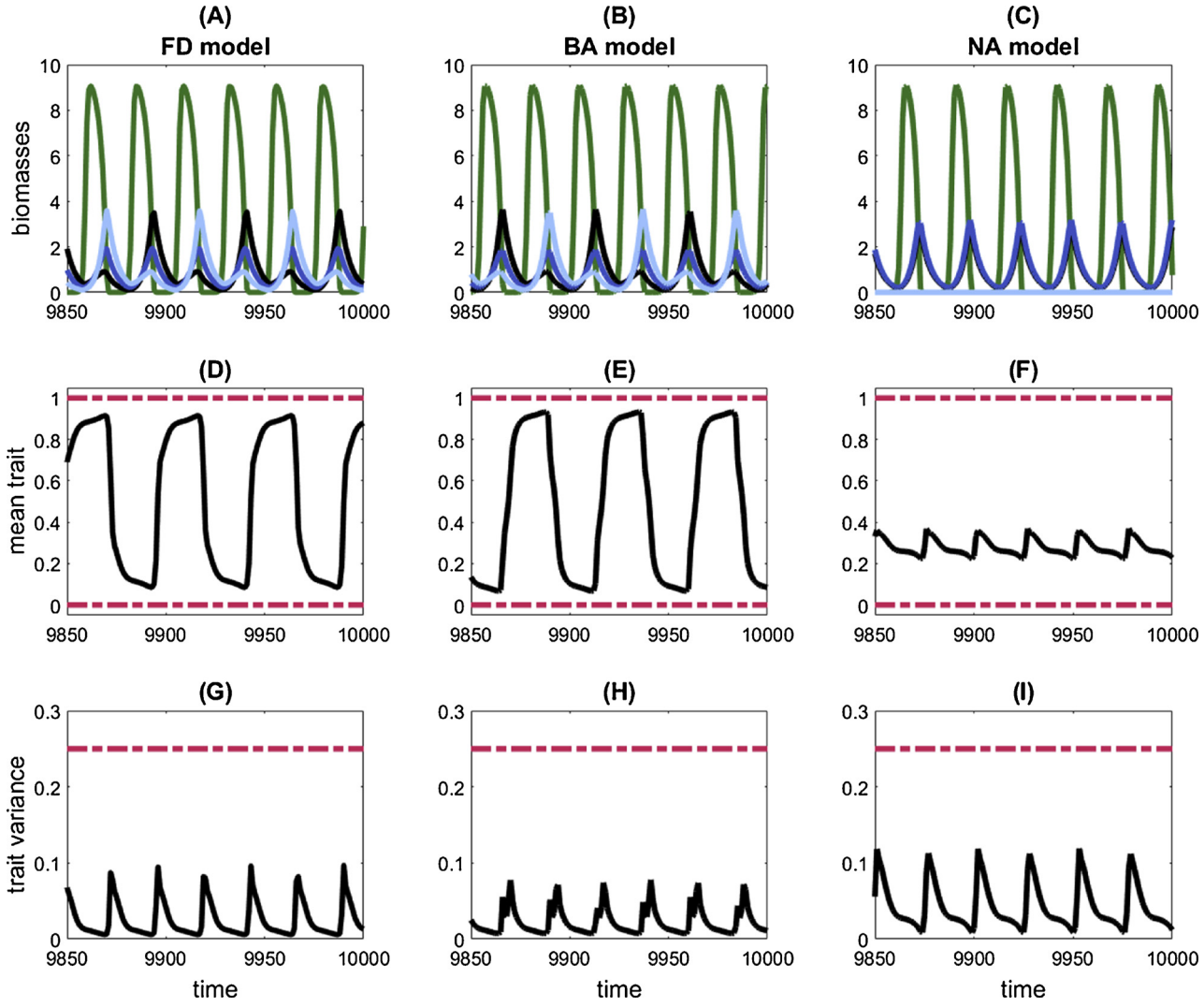

Fig. 4. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (green) and the 3 predator species that are specialized on low (black), intermediate (dark blue) and high (light blue) trait values of the prey community under fluctuating selection. For further details see Fig. 3.

skewed or flat, likely caused by strongly fluctuating or disruptive selection. Skewed and bimodal size distributions were also frequently observed in other aquatic (Tackx et al., 1994; Downing et al., 2014) and terrestrial (Maurer et al., 2004; Thibault et al., 2011) communities. Therefore, normal distribution-based aggregate (*NA*) models are likely to be inappropriate for many natural systems (Coutinho et al., 2016).

We aimed to overcome these limitations by investigating the behavior of a new trait-based aggregate model, based on the assumption that the trait values are beta-distributed. In contrast to the normal distribution, the beta distribution is defined on a finite range, as are presumably all ecologically relevant traits such as body size, generation times or feeding rates. As a result, the beta distribution can be unimodal (and symmetric), heavily skewed, or bimodal, depending on its mean and variance. Additionally, it exhibits strong negative correlations between the lower and higher moments in line with the observed phytoplankton size distributions.

4.1. Performance of trait-based aggregate models under different selection regimes

A comparison of the biomass and trait dynamics of a full trait distribution (*FD*) model with the ones of a beta distribution-based aggregate (*BA*) model revealed that a corresponding moment closure can strongly improve the reliability of trait-based aggregate models. We also applied the method to other models found in the literature (Appendix B), showing that our results are not specific to particular model choices. Under stabilizing and fluctuating selection, we observed an almost perfect match between the biomass and trait dynamics of the *FD* and *BA* models. This result was independent of the initial conditions, and relies mainly on the fact that the skewness (S) of the trait distribution and the curvature of the fitness landscape at the mean trait value ($\overline{x}$) ($\partial^2 F/\partial x^2|_{x=\overline{x}}$) jointly prevent $\overline{x}$ from getting locked at a local fitness maximum (Appendix E). In contrast to previous results where the existence of convergent stable fitness minima or convergent unstable fitness maxima relies on frequency-dependent selection (e.g. Abrams et al., 1993b), the mechanism described above can stabilize fitness minima or destabilize fitness maxima even in the absence of frequency-dependent selection (Appendix E). Nevertheless, a certain amount of standing trait variation (ν), and thus immigration, was still needed to maintain ongoing changes in $\overline{x}$ under fluctuating selection (for details see Appendix C). The close correspondence between the biomass and trait dynamics of the *FD* and *BA* models is in line with a statistical comparison of different aggregate models linked to empirical size distributions of phytoplankton communities (Gaedke and Klauschies, 2017). The best approximation of the rate of change of the aggregate properties of the corresponding *FD* model was obtained with a moment closure approach based on the observed negative correlations between the higher and lower central moments of the phytoplankton size distributions.

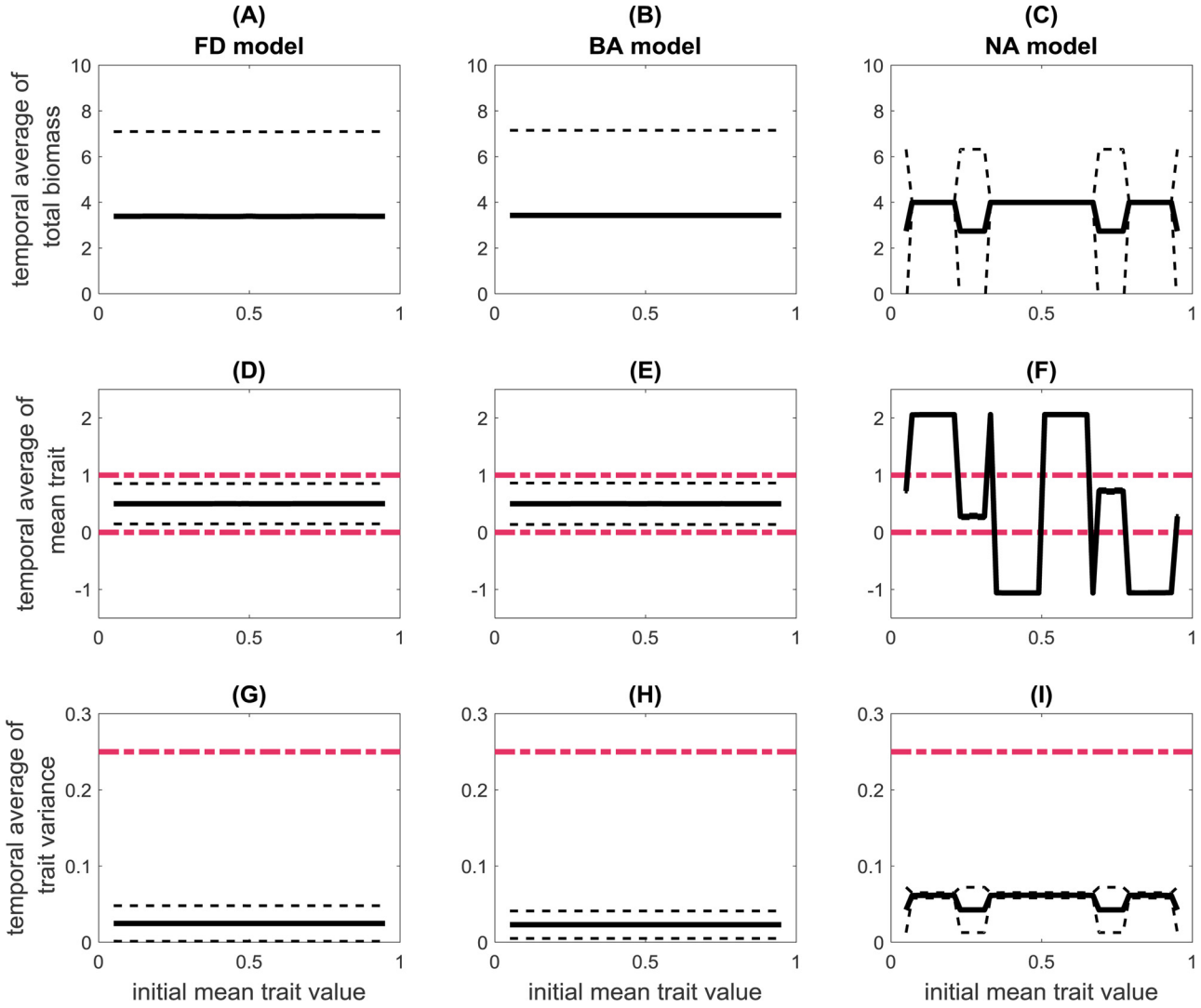

Fig. 5. Temporal mean (solid line) ± standard deviation (dashed line) of the total biomass (A–C), mean trait (D–F) and trait variance (G–I) of the prey community in dependence of the initial mean trait value for the full trait distribution (*FD*) model (A, D, G), beta distribution based aggregate (*BA*) model (B, E, H) and normal distribution based aggregate (*NA*) model (C, F, I) under fluctuating selection. The dashed-dotted red lines in (D)–(I) indicate the ecological feasible minima and maxima of the mean and variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In contrast, the *NA* model predicted biomass and trait dynamics that differed strongly from the ones of the *BA* and *FD* models. For example, the changes in $\bar{x}$ did not cover the entire trait range, but exhibited small-amplitude oscillations. Consequently, one or two of the three predators went extinct. In addition, the *NA* model showed a high sensitivity to initial conditions under fluctuating selection. In many cases, $\bar{x}$ surpassed its natural limits, because the *NA* model lacks an inherent mechanism to prevent this. In contrast, the *BA* model entails a skewness-mediated feedback between $\bar{x}$ and ν close to the extremes which arises from the upper and lower limits of the beta distribution, keeping $\bar{x}$ within its ecologically feasible range. When $\bar{x}$ approaches one of the extremes, S will always increase in modulus and have the opposite sign of the fitness gradient (see Eq. (18) and Fig. 2E), leading to a strong decline of ν (see Eq. (14)). As a result, the rate of change of $\bar{x}$ slows down quickly, given that ν scales $d\bar{x}/dt$, eventually stopping $\bar{x}$ at or close to one of the extremes (for details see Appendix D). This shows that temporal changes in S are very important for restricting $\bar{x}$ to its ecologically feasible range by promoting a reduction of ν when $\bar{x}$ approaches one of the extremes.

Investigating the moment closure equations of the beta distribution in more detail also reveals that a feedback between ν and K ensures ν to be bounded to its ecologically feasible range in the *BA* model. For example, assuming the trait distribution to be symmetric, an increase of ν leads to a decline of K (Eq. (19) and Fig. 2F) because only rather flat distributions can accommodate large values of ν. When ν approaches its maximum value of 0.25, K decreases towards one, indicating a Bernoulli distribution, which is a special (extreme) case of a symmetric bimodal distribution where all biomass is allocated to the shoulders of the trait distribution and none to the tails or the mean. Hence, the factor in front of the second term of the rate of change of ν becomes zero (see Eq. (14)), so ν cannot grow further. In contrast, the normal distribution does not exhibit lower and upper limits and, in addition, its kurtosis does not depend on ν, allowing ν to surpass its natural limit in *NA* models (cf. Tirok et al., 2011).

The good approximation of the *FD* model by the *BA* model under fluctuating selection agrees with recent findings of a stochastic population model for endemic infections. A moment closure based on the discrete beta-binomial distribution improved the reliability of a corresponding aggregate model to adequately represent the temporal development of $\bar{x}$ and ν of the probability distribution of infected individuals of a susceptible-infected (*SI*) logistic model (Krishnarajah et al., 2005). Furthermore, the beta-binomial distribution-based moment closure was found to clearly outperform other moment closure approaches based on e.g. the normal or binomial distribution in

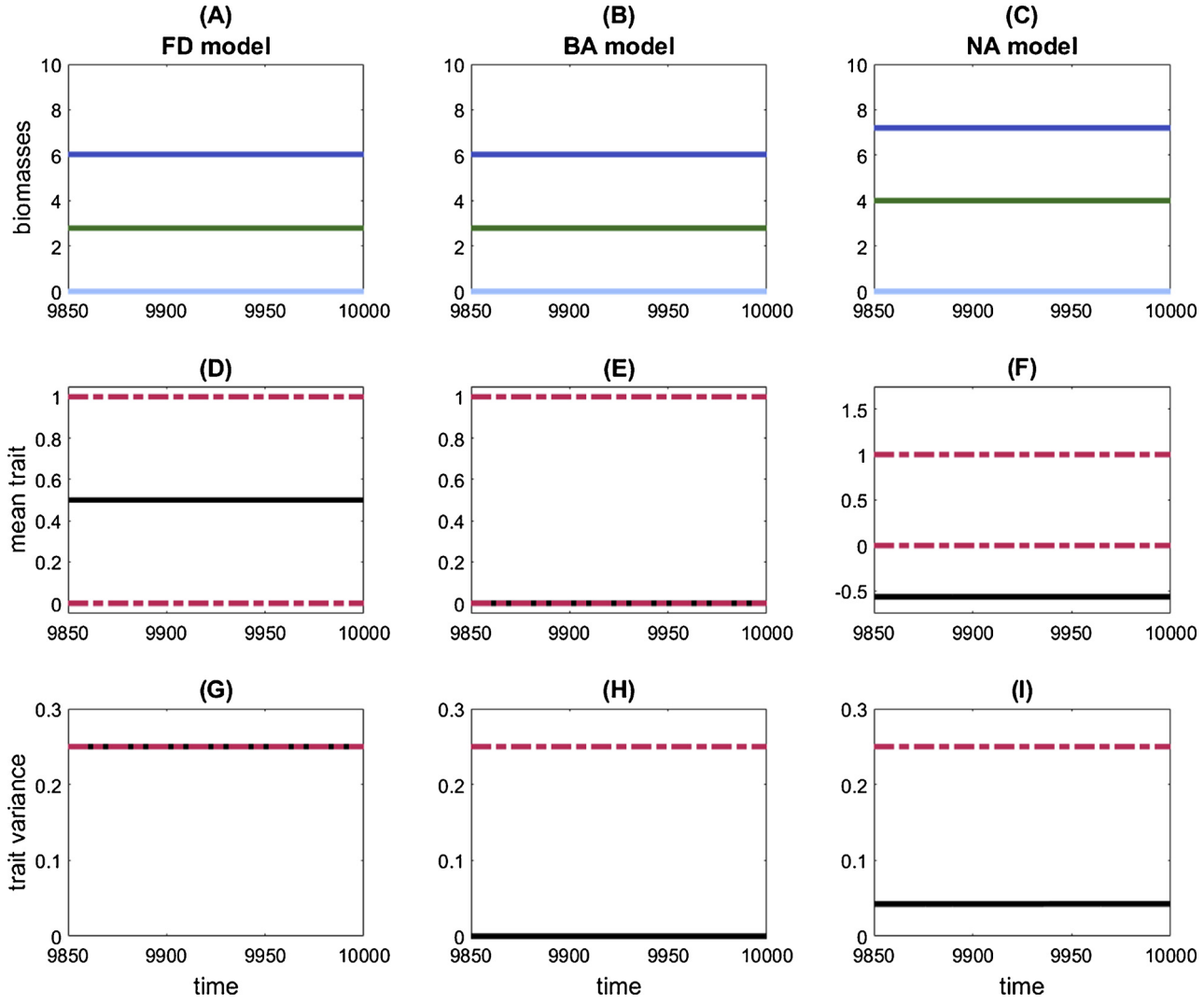

Fig. 6. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (green) and the 3 predator species that are specialized on low (black), intermediate (dark blue) and high (light blue) trait values of the prey community under disruptive selection. For further details see Fig. 3. Note the different scale of the y-axis in panel F.

matching the overall shape properties of the quasi-stationary distribution of a *SIS* logistic model (Clancy and Mendy, 2011).

Under disruptive selection both aggregate models failed to reproduce the biomass and trait dynamics of the *FD* model, exhibiting a bimodal distribution with maximum ν. Despite the fact that our model gives rise to frequency-dependent selection and thus may enable the emergence of convergent stable fitness minima in aggregate models (cf. Abrams et al., 1993b), our analysis presented in Appendix E shows that such minima do not exist in *NA* models with a single predator, leading to the same stability conditions for $\overline{x}$ with or without frequency-dependent selection. Essentially, the normal distribution cannot represent a bimodal distribution where $\overline{x}$ is located at the fitness minimum, and thus is no longer representative of the entire community. Consequently, $\overline{x}$ always moves towards a local fitness maximum in the *NA* model at least under frequency-independent selection, where $\partial^2 F/\partial x^2|_{x=\overline{x}}$ is negative, so ν tends to zero over time when no other source of standing trait variation exists (Appendix E). This is very important because high functional diversity may reduce the overall performance of the community because of suboptimal phenotypes (i.e., due to negative second derivatives at the optimal mean trait value) but at the same time may enable the community to adapt to new fitness optima when the environment is changing (Norberg et al., 2001).

In general, a strong reduction in the trait variance of aggregate models due to the fitness gradient approximation may lead to a mismatch between the timescales of the aggregate model and a corresponding *FD* model (Coutinho et al., 2016) which likely hampers the application of aggregate models to natural systems (but see Terseleer et al., 2014 for an example). Furthermore, in the *NA* model, $\overline{x}$ cannot overcome a fitness minimum separating different peaks from each other, which strongly contrasts with *FD* models (Coutinho et al., 2016) and potentially also with *BA* models (Appendix E). Hence, considering local fitness gradients instead of entire fitness landscapes necessarily assumes the absence of multiple concurrent peaks in the fitness landscape. In contrast, our results demonstrate that higher derivatives of the fitness function and higher moments of the trait distribution are required to drive changes of the aggregate properties into the right direction under disruptive selection. As shown in Appendix E, the negative correlations between the lower and higher moments, entailed in the beta distribution and observed in the size distributions of the phytoplankton communities, may contribute to the maintenance of ν and thus functional diversity over time. In fact, ν exhibits a potential stable fixed point larger than zero if $\overline{x}$ is located at a fitness minimum, where the curvature of the fitness function is positive (see Fig. E1). Since skewness can stabilize a fixed point of $\overline{x}$ at a fitness minimum, a stable solution of a *BA* model can exist with $\nu > 0$ even without immigration (for details see Appendix E). Indeed, when the skewness and the local fitness gradient exhibit the same sign, an increase in the

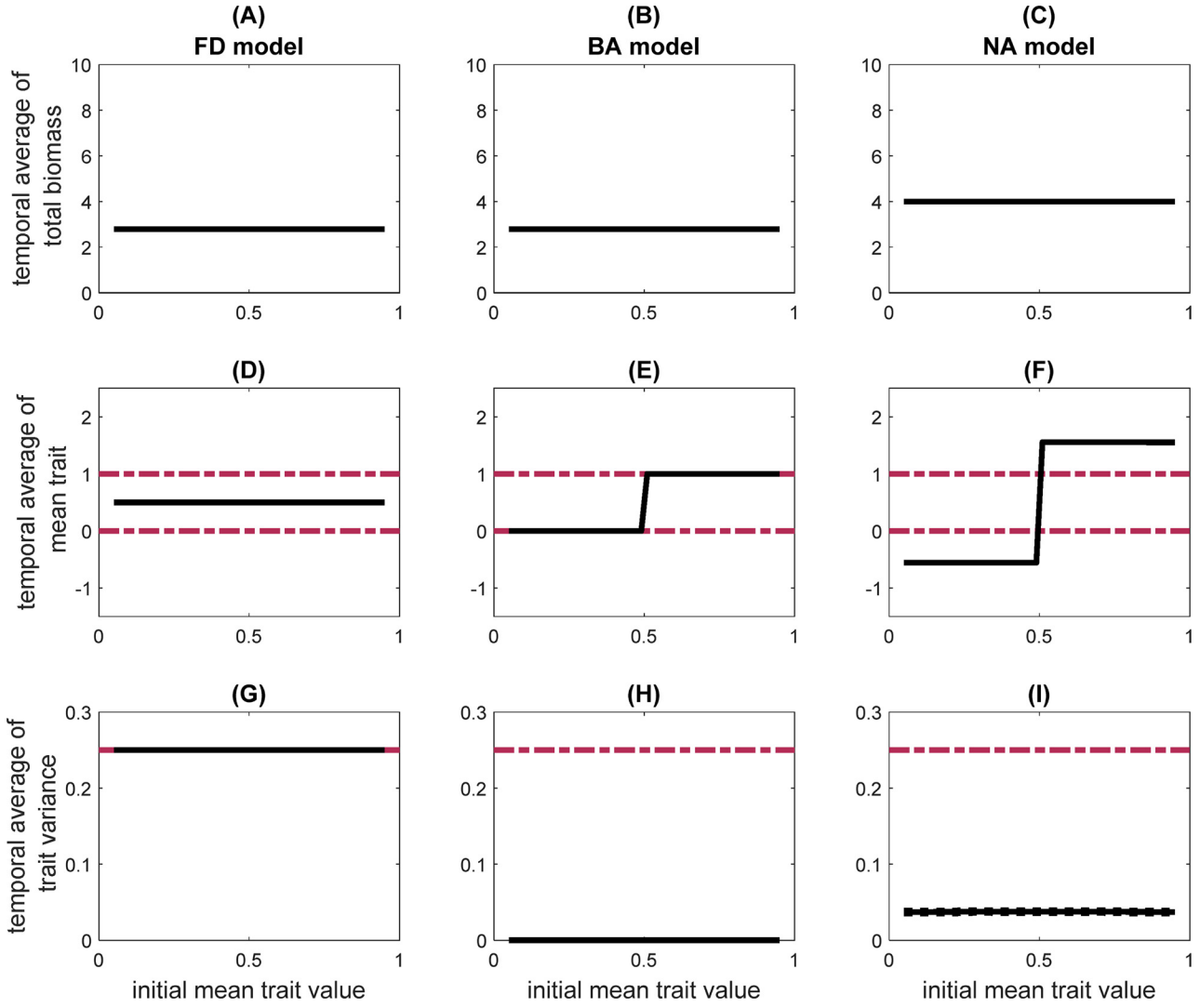

Fig. 7. Temporal mean (solid line) ± standard deviation (dashed line) of the total biomass (A–C), mean trait (D–F) and trait variance (G–I) of the prey community in dependence of the initial mean trait value for the *FD* model (A, D, G), *BA* model (B, E, H) and *NA* model (C, F, I) under disruptive selection. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

variance is promoted (Norberg et al., 2001). However, in our case the second order approximation of the fitness function entailed in the *BA* model was not sufficient to maintain $\nu > 0$. As a result, additional processes, like immigration or mutation, may be required to ensure $\nu > 0$ over time (Merico et al., 2014; Coutinho et al., 2016).

Various mechanisms have been used to maintain ν (cf. Merico et al., 2014). For example, while mutation is often viewed to have overall small effects on $d\overline{x}/dt$ and $d\nu/dt$ (Day, 2005), immigration strongly influences them at low densities (Appendix E). Furthermore, in the absence of mutational bias, mutation has little impact on $d\overline{x}/dt$ while preventing ν from getting vanishingly small, a feature which can be derived when modelling the mutation process as trait diffusion (Coutinho et al., 2016). As a result, *NA* models are sensitive to the added source of standing trait variation and, thus, to the choice of mechanism used to maintain ν (Coutinho et al., 2016). This holds also for *BA* models because the sum of a beta distribution and another trait distribution is not necessarily beta-distributed, that is, the moment closure equations might interfere with the other mechanism used to maintain ν. For example, adding standing trait variation to $d\nu/dt$ of a *BA* model may strongly interfere with $d\overline{x}/dt$ when $\overline{x}$ is close to one of the extremes, because the moment closure equations can no longer represent adequately the actual values of S and K under such conditions. In fact, ν has to become very small when $\overline{x}$ approaches one of the extremes. Hence, accounting for a process that only leads to an increase in ν without pushing $\overline{x}$ towards intermediate values will generally not work within the framework of a *BA* model. Consequently, we could not derive a *BA* model that incorporates mutation as diffusion with a reflecting boundary because the beta distribution is defined on an open interval. More generally, these results underline the need to understand more clearly the potential effects of immigration and mutation on biomass-trait feedbacks in trait-based aggregate models.

4.2. Relevance of our findings for evolutionary ecology

The aggregate model approach used in this study to describe the temporal dynamics of the aggregate properties of the trait distribution of a community is akin to various mathematical descriptions of trait adaptation in evolutionary ecology. In line with classical quantitative genetics (*QG*) models that describe rapid or short-term evolution within a population (Slatkin, 1979; Taper and Case, 1992; Day, 2005), aggregate models do not assume a separation of ecological and evolutionary times scales. Rather, they consider explicitly the dynamical feedback between trait and biomass dynamics within a community (Norberg et al., 2001; Merico et al., 2009). In addition, both model approaches explicitly allow for standing trait variation and often rely on the assumption that the trait distribution is (and remains) normally distributed, thereby enhancing numerical or analytical tractability of

the model (Taylor and Day, 1997; Coutinho et al., 2016). This assumption is valid at the population level when considering the evolution of a quantitative trait in which numerous loci determine the individual's phenotype, with each locus having a small additive effect on the phenotype (Slatkin, 1979; Kirkpatrick and Rousset, 2005). Essentially, in case of random mating, sexual reproduction and recombination, the trait distribution of a population can often be well approximated by a normal distribution with a particular $\overline{x}$ and ν (Lande, 1976; Barton and Turelli, 1989). In contrast, there exists hardly any theoretical or empirical foundation for the *Gaussian* approximation at the community level, where trait changes mainly rely on species sorting (Gaedke and Klauschies, 2017). Furthermore, at the population level, also various mechanisms exist that may promote non-normally distributed trait values, including assortative mating in sexual populations or clonal sorting in asexual populations (Doebeli et al., 2007; Sasaki and Dieckmann, 2011). Hence, the shape of trait distributions may often strongly deviate from a *Gaussian* including skewed or bimodal trait distributions. Our results show that corresponding changes in skewness and kurtosis may strongly influence the biomass and trait dynamics and, therefore, model predictions (cf. Fig. 4).

In contrast to many aggregate models that allow for simultaneous changes in $\overline{x}$ and ν (Norberg et al., 2001; Coutinho et al., 2016; but see Merico et al., 2014), *QG* models often focus on changes in mean trait values by assuming ν to be constant (Taper and Case, 1985, 1992; Day, 2005). The latter seems to be valid under weak selection in which case the selective advantage of different phenotypes is vanishingly small, meaning that the fitness landscape is rather flat and thus, the fitness function and its first derivative do not change much over the bulk of the trait distribution (Lande, 1976; Taylor, 1996; Lion, 2018). However, keeping ν constant may neglect a potentially very important feedback between $d\overline{x}/dt$ and $d\nu/dt$ because temporal changes in ν and $\overline{x}$ are tightly coupled to each other through their combined effects on the entire fitness landscape (Appendix A). In line with previous studies (Norberg et al., 2001; Smith et al., 2016), our results show that the speed of trait adaptation through higher ν comes at a cost of reduced fitness at the community level when $\overline{x}$ is located around a fitness maximum (Appendix E).

In contrast to *QG* and aggregate models, the approach of Adaptive Dynamics (*AD*) considers long-term evolution to be a mutation-limited process, giving rise to a separation of ecological and evolutionary time scales (Parker and Maynard Smith, 1990; Day, 2005). The key assumption of *AD* is that a new but rare mutant with a slightly different phenotype than the resident phenotype arises at a point where the monomorphic resident population has reached its demographic equilibrium (Geritz et al., 1998; McGill and Brown, 2007). If the phenotype of the rare mutant is mal-adapted, it will necessarily die out, whereas it will spread when it is advantageous, replacing the former phenotype of the resident population. Evolution of the resident phenotype is then driven by a sequence of rare mutation and strong selection (Day, 2005; Kirkpatrick and Rousset, 2005). *AD* focuses on finding the evolutionary endpoints, i.e. evolutionary singular strategies, which can be reached by gradual evolution starting from a strategy nearby (McGill and Brown, 2007). Such convergent stable strategies (CSS) can be either resistant (stable - ESS) or vulnerable (branching point) to invasion by other nearby strategies (Geritz et al., 1998; Doebeli et al., 2007). In the latter case, traits evolve to a point where selection becomes disruptive giving rise to dimorphism (Rueffler et al., 2006; McGill and Brown, 2007). At this point, it becomes necessary to follow the evolution of two distinct phenotypes. This contrasts with our BA model, which describes the temporal development of a continuous trait distribution that may become bimodal and thus exhibit two dominant but distinct phenotypes. Furthermore, aggregate models explicitly consider biomass and trait dynamics that occur on similar time scales and therefore account for the fact that changes in $\overline{x}$ or ν may strongly affect the stability and shape of biomass dynamics (Abrams and Matsuda, 1997; Coutinho et al., 2016; Cortez and Patel, 2017).

4.3. Future perspectives

To overcome the overall weak performance of trait-based aggregate models under disruptive selection, one may use a hybrid model. Here the mean trait and variance dynamics of two or more different species or functional groups are modeled separately, using an aggregate model formulation for each species (Norberg et al., 2012; Klauschies et al., 2016). Hence, the community is subdivided into a limited number of species or functional groups, whose mean trait values and trait variances can, in turn, adapt in response to selection (cf. Norberg et al., 2001; Terseleer et al., 2014). In the context of genetics, a similar approach was developed and advocated by Sasaki and Dieckmann (2011) where the trait distribution is decomposed into a sum of unimodal trait distributions corresponding to the individual morphs of a population. This is reasonable when the entire trait distribution is oligomorphic or multimodal, meaning that it comprises a finite number of distinct peaks. Each peak is associated with a different morph that can be viewed as a quasi-species characterized by its particular trait distribution. In line with *QG* models (Lande, 1976; Abrams et al., 1993b), $d\overline{x}_i/dt$ of morph i follows the product of ν_i and $\partial F_i/\partial x_i|_{x_i=\overline{x}_i}$ under the assumption that all ν_i are small for all i (Sasaki and Dieckmann, 2011). Although our derivation of an aggregate model is also based on the assumption that ν is sufficiently small (cf. Appendix A), our results show that this approximation may capture also more complex scenarios where ν is temporarily quite large when assuming the trait values to be beta-distributed (cf. Fig. 4). Hence, a promising avenue for future research would be to combine our novel moment-closure approximation based on the beta distribution with a hybrid or oligomorphic dynamics approach. The alternative, to model the temporal dynamics of higher moments or to extend the second-order approximation, appears less promising given that the corresponding equations and numerical simulations become very demanding.

We established that the choice of a suitable moment closure approach strongly depends on the prevailing selection regime. Hence, it is important to develop reliable tools to characterize the selection pressure on the trait composition of natural communities. In the approach of Adaptive Dynamics, the predominant selection pressure on a convergent stable strategy $\overline{x}^*$ of a population is estimated based on the sign of the second derivative of the mutant's invasion fitness with respect to its own strategy x evaluated at the resident strategy $\overline{x}$, i.e. $\partial^2 F/\partial x^2|_{x=\overline{x}^*}$. Negative values indicate stabilizing selection and, thus, an evolutionary stable strategy (*ESS*) whereas positive values indicate disruptive selection and, thus, an evolutionary branching point that may give rise to dimorphisms (Geritz et al., 1998; McGill and Brown, 2007). These constraints are valid when assuming weak selection, which implies that the distribution is tightly clustered around $\overline{x}$ (Lion, 2018). They also hold when selection is not weak, but a separation of time scales between ecology and evolution under rare mutations is assumed (Day, 2005). In this case, advantageous novel phenotypes are strongly and positively selected, so that a monomorphic resident community can be assumed at the time a new mutant arises (Geritz et al., 1998). Accordingly, a recently proposed approach to distinguish between the prevalence of stabilizing or disruptive selection in *NA* models is based on the sign of their $\partial^2 F/\partial x^2|_{x=\overline{x}}$ (Acevedo-Trejos et al., 2016). However, using a normal distribution-based moment closure relies on the assumption that the environment imposes stabilizing or weak directional selection on the trait considered. Hence, a verification of stabilizing selection cannot be attained by considering actual properties of the resulting equations of a *NA* model, because the result will not be independent of the assumption of normality, and thus a prevalence of stabilizing selection. As detailed in Appendix E, the temporal changes of $\overline{x}$ and ν promote negative values of $\partial^2 F/\partial x^2|_{x=\overline{x}}$ without necessarily implying stabilizing selection. In general, the sign of $\partial^2 F/\partial x^2|_{x=\overline{x}}$ does only inform reliably about the prevalence of stabilizing or disruptive selection when the fitness landscape can be well described by a parabolic function for which higher derivatives vanish (cf. Abrams et al.,

1993a). In all other cases, the view that $\partial^2 F/\partial x^2|_{x=\bar{x}} < 0$ provides clear evidence for stabilizing selection and $\partial^2 F/\partial x^2|_{x=\bar{x}} > 0$ for disruptive selection is oversimplified and potentially wrong. This happens because concomitant changes in $\bar{x}$ also promote negative values of $\partial^2 F/\partial x^2|_{x=\bar{x}}$ under clearly disruptive selection regimes where two distinct fitness maxima co-occur. Hence, approaches developed in population genetics (cf. Nelson et al., 2005; Siepielski et al., 2009; Fox et al., 2010) may deliver more reliable estimates of the prevailing selection pressure on the trait composition of natural communities.

Finally, in line with *AD* and optimality approaches (Parker and Maynard Smith, 1990), we may further enhance our understanding of trait-based aggregate models by considering quasi-equilibrium approximations based on a separation of time scales of the ecological (biomass) and the evolutionary (trait) dynamics (Geritz et al., 1998; Cortez and Patel, 2017). For example, when ν is sufficiently small, we can assume the evolutionary dynamics to be much slower than the ecological dynamics. In this case, the ecological variables will reach a quasi-equilibrium at the current, changing trait values (Day, 2005; Lion, 2018). In contrast, when organisms are able to quickly adapt their traits in response to selection, we can assume that the evolutionary dynamics occur on a much faster time scale than the ecological dynamics (Smith et al., 2009; Abrams, 2010; Pahlow and Prowe, 2010). Analyzing such fast-slow dynamics may allow for the derivation of general results under simplifying assumptions (Cortez, 2016; Cortez and Patel, 2017).

4.4. Final remarks

The small number of state variables of trait-based aggregate models may enable rigorous analyses of their equations (Appendix E; Norberg et al., 2001) and significantly reduce their simulation times when compared to corresponding full trait distribution (*FD*) models (Table 2; Acevedo-Trejos et al., 2016). However, as moment approximations do not come with an estimate of the approximation error, one usually has to resort to numerical simulations of the *FD* model to judge when the approximation is good enough to replicate the behavior of the *FD* model. In practice, this may question the general applicability of trait-based aggregate models, even though an aggregate model usually does not need to be validated for the entire parameter space nor the whole simulation time. Our work improves this situation in three ways. In line with Coutinho et al. (2016), we first present detailed evidence that a general agreement between *FD* models and their aggregate counterparts is closely linked to the predominant selection regime. Hence, comparing *FD* and corresponding aggregate models under different but well-defined environmental conditions revealed the structural components, i.e. the shape properties of the fitness landscape and trait distribution that are responsible for the success or failure of the approximations and, thus, which underlie the dynamics of the *FD* model. Therefore, knowledge about the shape of natural trait distributions or corresponding selection regimes may sufficiently inform about the behavior of trait-based aggregate models compared to corresponding *FD* models, and circumvent the need for a detailed model comparison based on numerical simulations. Second, we show that the beta distribution captures important shape properties of natural and simulated trait distributions, which cannot be reflected by a *Gaussian* but which are important to restrict changes in the mean trait and variance values to their ecologically feasible ranges. In contrast, *NA* models usually require additional functional terms to keep traits bounded. Hence, conducting a moment closure based on the well-defined relationships between the lower and higher moments of the beta distribution instead of the normal distribution greatly enhanced the reliability of aggregate models under fluctuating selection. This increases the range of situations that can be treated successfully using moment approximation methods. Finally, we obtained a more comprehensive understanding of the mechanisms that may promote the maintenance of functional diversity in trait-based aggregate models, by analyzing the aggregate model equations. We explicitly provide the connection between higher moments of the trait distribution and higher derivatives of the fitness landscape and the preservation of the trait variance for a quite general class of models (see Appendix E).

To conclude, our study shows that conducting a moment closure based on the beta distribution rather than the normal distribution may strongly improve the reliability of aggregate models when applied to natural communities, e.g. with respect to the temporal average and variation of the communities' biomass and trait composition. This holds in particular when mean trait values change over time, e.g. due to strong fluctuating selection caused by endogenous or environmental changes such as seasonal temperature variation.

Acknowledgments

We thank Ruben Ceulemans, Ellen van Velzen, Alice Boit and three anonymous reviewers for helpful comments and suggestions. TK was funded by the German Research Foundation (DFG, GA 401/26-1). RMC was supported through grant #2014/23497-0, São Paulo Research Foundation (FAPESP), Brazil.

Appendix A. Derivation of an aggregate model

Temporal changes in the aggregate properties of a community, i.e. its total biomass (R_T), mean trait ($\bar{x}$) and trait variance (ν), can be conveyed from the changes of the population biomasses of the individual species:

$$\frac{\partial R(x)}{\partial t} = F(x)R(x) + I, \tag{A1}$$

where $F(x)$ denotes the per capita net growth rate – the fitness function – of the species with trait value x, ranging between 0 and 1, whose biomass is $R(x)$. The species' fitness function $F(x, R_T, \bar{x}, \nu)$ generally depends on the entire trait distribution of the community such as R_T, $\bar{x}$ and ν. However, for the sake of brevity, we stick to $F(x)$ throughout the derivation. The small but constant immigration rate I ensures long-term persistence of all trait values, i.e. species. The aggregate properties of a community are calculated according to the following set of equations:

$$R_T = \int_0^1 R(x)dx \tag{A2}$$

$$\bar{x} = \frac{1}{R_T}\int_0^1 xR(x)dx \tag{A3}$$

$$\nu = \frac{1}{R_T}\int_0^1 (x-\bar{x})^2 R(x)dx \tag{A4}$$

Instead of describing the dynamics of the community with differential equations defined at the population level, one may focus directly on a

system of differential equations that describes the rates of change of the community's aggregate properties, thereby allowing insight into the properties of the fitness landscape and trait distribution that drive their dynamics. In general, such aggregate models approximate the temporal development of the aggregate properties by first truncating the Taylor series of the fitness function developed around $\overline{x}$ after some low-order, e.g. second-order, terms. Afterwards, the resulting set of equations has to be closed by expressing the higher central moments such as skewness (S) and kurtosis (K) in terms of the lower central moments modelled, such as $\overline{x}$ and ν. As all odd moments, skewness quantifies some degree of asymmetry of a distribution. Left-skewed distributions exhibit negative values of S whereas right-skewed distributions exhibit positive values of S. In contrast, kurtosis quantifies the amount of dispersion of mass around the shoulders ($\overline{x} \pm \sqrt{\nu}$) of a distribution. Based on this, high values of K are associated with distributions having high mass at their center ($\overline{x}$) and tails ($\overline{x} \pm 2\sqrt{\nu}$) compared to their shoulders.

A.1. Second-order approximations of the rates of change of the aggregate properties

The rate of change of R_T is given by the integral of the rate of change of $R(x)$ with respect to x over the trait range [0,1]:

$$\frac{dR_T}{dt} = \frac{d}{dt}\int_0^1 R(x)\mathrm{d}x = \int_0^1 \frac{\partial R(x)}{\partial t}\mathrm{d}x = \int_0^1 (R(x)F(x) + I)\mathrm{d}x = \int_0^1 R(x)F(x)\mathrm{d}x + I\int_0^1 \mathrm{d}x \tag{A5}$$

Noting that $\int_0^1 dx = 1$ and expanding the fitness function F in a Taylor series around $\overline{x}$, we get:

$$\frac{dR_T}{dt} = \int_0^1 R(x)\sum_{k=0}^{\infty}\frac{(x-\overline{x})^k}{k!}\frac{\partial^k F(x)}{\partial x^k}\bigg|_{x=\overline{x}}\mathrm{d}x + I \tag{A6}$$

We can rearrange the terms of Eq. (A6) to express the rate of change of R_T as an infinite sum of products of the k-th derivative of the fitness function and the k-th central moment (M_k) of the trait distribution:

$$\frac{dR_T}{dt} = R_T\sum_{k=0}^{\infty}\frac{1}{k!}\frac{\partial^k F(x)}{\partial x^k}\bigg|_{x=\overline{x}}\underbrace{\frac{1}{R_T}\int_0^1 R(x)(x-\overline{x})^k dx}_{M_k} + I \tag{A7}$$

To highlight the interaction between the higher-order central moments and the higher derivatives we can rewrite Eq. (A7) in the following way:

$$\frac{dR_T}{dt} = R_T\left(F(x)|_{x=\overline{x}} + \frac{\nu}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{M_3}{6}\frac{\partial^3 F(x)}{\partial x^3}\bigg|_{x=\overline{x}} + \frac{M_4}{24}\frac{\partial^4 F(x)}{\partial x^4}\bigg|_{x=\overline{x}} + \ldots\right) + I \tag{A8}$$

given that $M_0 = 1$, $M_1 = 0$ and $M_2 = \nu$.

Similarly, the rate of change of $\overline{x}$ is given by the quotient rule of differential calculus using Eq. (A3):

$$\frac{d\overline{x}}{dt} = \frac{1}{R_T^2}\left[\left(\frac{d}{dt}\int_0^1 xR(x)dx\right)R_T - \left(\int_0^1 xR(x)dx\right)\frac{dR_T}{dt}\right] \tag{A9}$$

Using the definition of $\overline{x}$, Eq. (A9) can be expressed as:

$$\frac{d\overline{x}}{dt} = \frac{1}{R_T}\left(\int_0^1 x\frac{\partial R(x)}{\partial t}dx - \overline{x}\frac{dR_T}{dt}\right) \tag{A10}$$

By extending the first term of Eq. (A10) with $\overline{x}$ we can eliminate its second term:

$$\frac{d\overline{x}}{dt} = \frac{1}{R_T}\left(\int_0^1 (x-\overline{x}+\overline{x})\frac{\partial R(x)}{\partial t}dx - \overline{x}\frac{dR_T}{dt}\right) = \frac{1}{R_T}\left(\int_0^1 (x-\overline{x})\frac{\partial R(x)}{\partial t}dx + \overline{x}\int_0^1 \frac{\partial R(x)}{\partial t}dx - \overline{x}\frac{dR_T}{dt}\right) = \frac{1}{R_T}\int_0^1 (x-\overline{x})\frac{\partial R(x)}{\partial t}dx = \frac{1}{R_T}\int_0^1 (x-\overline{x})(R(x)F(x)+I)dx \tag{A11}$$

Evaluating the effect of external immigration I leads to:

$$\frac{1}{R_T}\int_0^1 (x-\overline{x})R(x)F(x)dx + \frac{I}{R_T}\int_0^1 (x-\overline{x})dx = \frac{1}{R_T}\int_0^1 (x-\overline{x})R(x)F(x)dx + \frac{I}{R_T}\left(\frac{x^2}{2}-\overline{x}x\right)\bigg|_0^1 = \frac{1}{R_T}\int_0^1 (x-\overline{x})R(x)F(x)dx + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) \tag{A12}$$

Expanding again the fitness function F in a Taylor series around $\overline{x}$ and rearranging the terms thereafter we get:

$$\frac{d\overline{x}}{dt} = \frac{1}{R_T}\int_0^1 R(x)\sum_{k=0}^{\infty}\frac{(x-\overline{x})^k}{k!}\frac{\partial^k F(x)}{\partial x^k}\bigg|_{x=\overline{x}}(x-\overline{x})dx + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) = \sum_{k=0}^{\infty}\frac{1}{k!}\frac{\partial^k F(x)}{\partial x^k}\bigg|_{x=\overline{x}}\underbrace{\frac{1}{R_T}\int_0^1 R(x)(x-\overline{x})^{k+1}dx}_{M_{k+1}} + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) \tag{A13}$$

with M_{k+1} denoting the (k + 1)-th central moment of the trait distribution. Again, we rewrite Eq. (A13) to highlight the interaction between the higher central moments of the trait distribution and the higher derivatives of the fitness function:

$$\frac{d\overline{x}}{dt} = \left(\nu\frac{\partial F(x)}{\partial x}\bigg|_{x=\overline{x}} + \frac{M_3}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{M_4}{6}\frac{\partial^3 F(x)}{\partial x^3}\bigg|_{x=\overline{x}} + \ldots\right) + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) \tag{A14}$$

given that $M_1 = 0$ and $M_2 = \nu$.

Finally, the rate of change of ν is also given by the quotient rule of differential calculus:

$$\frac{d\nu}{dt} = \frac{1}{R_T^2}\left[\left(\frac{d}{dt}\int_0^1 (x-\bar{x})^2 \mathrm{R}(x)\mathrm{d}x\right)R_T - \left(\int_0^1 (x-\bar{x})^2 \mathrm{R}(x)\mathrm{d}x\right)\frac{dR_T}{dt}\right] \tag{A15}$$

Now we use the product rule of differential calculus and the definition of the variance to display Eq. (A15) as:

$$\frac{d\nu}{dt} = \frac{1}{R_T}\left(\int_0^1 \frac{\partial R(x)}{\partial t}(x-\bar{x})^2 dx + \int_0^1 R(x)\frac{d(x-\bar{x})^2}{dt}dx - \nu\frac{dR_T}{dt}\right) \tag{A16}$$

Applying the chain rule of differential calculus, we see that the second term of Eq. (A16) is zero:

$$\frac{1}{R_T}\int_0^1 R(x)\frac{d(x-\bar{x})^2}{dt}dx = \frac{1}{R_T}\int_0^1 R(x)2(x-\bar{x})\frac{d(x-\bar{x})}{dt}dx = 2\frac{1}{R_T}\int_0^1 R(x)(x-\bar{x})\underbrace{\frac{dx}{dt}}_{=0}dx - 2\frac{d\bar{x}}{dt}\underbrace{\frac{1}{R_T}\int_0^1 R(x)(x-\bar{x})dx}_{=M_1=0} \tag{A17}$$

Deploying this result into Eq. (A16) leads to:

$$\frac{d\nu}{dt} = \frac{1}{R_T}\left(\int_0^1 \frac{\partial R(x)}{\partial t}(x-\bar{x})^2 dx - \nu\frac{dR_T}{dt}\right) = \frac{1}{R_T}\left(\int_0^1 (F(x)R(x)+I)(x-\bar{x})^2 dx - \nu\frac{dR_T}{dt}\right) \tag{A18}$$

Again, evaluating the effect of external immigration I leads to:

$$\frac{I}{R_T}\int_0^1 (x-\bar{x})^2 dx = \frac{I}{R_T}\left(\frac{x^3}{3} - \bar{x}x^2 + \bar{x}^2 x\right)\Big|_0^1 = \frac{I}{R_T}\left(\frac{1}{12} + \left(\frac{1}{2}-\bar{x}\right)^2\right) \tag{A19}$$

Leading to:

$$\frac{d\nu}{dt} = \frac{1}{R_T}\int_0^1 F(x)R(x)(x-\bar{x})^2 dx - \frac{\nu}{R_T}\frac{dR_T}{dt} + \frac{I}{R_T}\left(\frac{1}{12} + \left(\frac{1}{2}-\bar{x}\right)^2\right) \tag{A20}$$

Conducting a Taylor expansion of the fitness function around $\bar{x}$ in the first term and applying the results from Eq. (A7) to its second term, $d\nu/dt$ can be written as:

$$\begin{aligned}\frac{d\nu}{dt} &= \sum_{k=0}^{\infty}\frac{1}{k!}\frac{\partial^k F(x)}{\partial x^k}\Big|_{x=\bar{x}}\underbrace{\int_0^1 \frac{R(x)}{R_T}(x-\bar{x})^{k+2}dx}_{M_{k+2}} - \nu\sum_{k=0}^{\infty}\frac{1}{k!}\frac{\partial^k F(x)}{\partial x^k}\Big|_{x=\bar{x}}\underbrace{\int_0^1 \frac{R(x)}{R_T}(x-\bar{x})^k dx}_{M_k} + \frac{I}{R_T}\left(\frac{1}{12} - \nu + \left(\frac{1}{2}-\bar{x}\right)^2\right)\\ &= \sum_{k=0}^{\infty}(M_{k+2}-\nu M_k)\frac{1}{k!}\frac{\partial^k F(x)}{\partial x^k}\Big|_{x=\bar{x}} + \frac{I}{R_T}\left(\frac{1}{12} - \nu + \left(\frac{1}{2}-\bar{x}\right)^2\right)\end{aligned} \tag{A21}$$

In analogy to Eq. (A14), we rewrite Eq. (A21) in the following way:

$$\frac{d\nu}{dt} = \left[\underbrace{\left(\nu - \nu\underbrace{M_0}_{=1}\right)}_{=0}F(x)|_{x=\bar{x}} + \underbrace{\left(M_3 - \nu\underbrace{M_1}_{=0}\right)}_{=M_3}\frac{\partial F(x)}{\partial x}\Big|_{x=\bar{x}} + \frac{(M_4-\nu^2)}{2}\frac{\partial^2 F(x)}{\partial x^2}\Big|_{x=\bar{x}} + \frac{(M_5-\nu M_3)}{6}\frac{\partial^3 F(x)}{\partial x^3}\Big|_{x=\bar{x}} + \ldots\right] + \frac{I}{R_T}\left(\frac{1}{12} - \nu + \left(\frac{1}{2}-\bar{x}\right)^2\right) \tag{A22}$$

Since skewness (S) and kurtosis (K) are defined as the standardized central moments of third $\left(M_3 = S\nu^{\frac{3}{2}}\right)$, and fourth ($M_4 = K\nu^2$) orders of a trait distribution, a second-order approximation in F of dR_T/dt (Eq. (A8)), $d\bar{x}/dt$ (Eq. (A14)) and $d\nu/dt$ (Eq. (A22)) leads to the aggregate model defined by the Eqs. (12)–(14) in the methods section. This approximation is good when ν is small or the non-linearity of the fitness function F is weak (Abrams et al., 1993a; Taylor and Day, 1997; Gaedke and Klauschies, 2017).

Finally, since the rates of change of the lower central moments depend on the higher central moments, one has to close the system of differential equations. This is usually achieved by expressing the higher moments in terms of the lower ones. There are different moment closure techniques available, each making distinct assumptions about the shape of the trait distributions (Wirtz and Eckhardt, 1996; Norberg et al., 2001; Merico et al., 2009). For simplicity, a normal distribution-based moment closure is often conducted (Tirok et al., 2011; Terseleer et al., 2014; Acevedo-Trejos et al., 2015) where $S = 0$ and $K = 3$ are independent of $\bar{x}$ and ν. This strongly reduces the complexity of the aggregate model defined by Eqs. (16) and (17) at the cost of assuming the trait to be defined on an infinite range and the trait distribution to be unimodal and symmetric. However, ecologically important traits such as body size or time spent for searching or handling prey items are restricted to a finite range. In addition, fluctuating or disruptive selection may promote trait distributions exhibiting skewness or bimodality (Coutinho et al., 2016; Gaedke and Klauschies, 2017). For this reason, we will derive an alternative moment closure based on the assumption that the trait distribution follows the shape of a beta distribution, which can accommodate such features (cf. Fig. 1).

A.2. Beta distribution-based moment closure

Assuming the trait value x to be beta-distributed, $R(x)$ can be written as:

$$R(x;\, a,\, b) = \frac{R_T}{B(a,\, b)}x^{a-1}(1-x)^{b-1} \tag{A23}$$

with $B = \int_0^1 x^{a-1}(1-x)^{b-1}dx$ being the *beta function* and $a > 0$ and $b > 0$ being the shape parameters of the beta distribution. The mean $\overline{x}$, variance ν, skewness S and kurtosis K of the beta-distribution are given by (Gupta and Nadarajah, 2004):

$$\overline{x} = \frac{a}{a+b} \tag{A24}$$

$$\nu = \frac{ab}{(a+b)^2(a+b+1)} \tag{A25}$$

$$S = \frac{2(b-a)(a+b+1)^{\frac{1}{2}}}{(a+b+2)(ab)^{\frac{1}{2}}} \tag{A26}$$

$$K = \frac{6(a-b)^2(a+b+1)-6ab(a+b+2)}{ab(a+b+2)(a+b+3)} \tag{A27}$$

Since the beta distribution only depends on the two shape parameters a and b, we can derive corresponding moment closure equations for an aggregate model in that S and K of the beta distribution are expressed in terms of $\overline{x}$ and ν only.

First we have to employ Eqs. (A24) and (A25) to express a and b as functions of $\overline{x}$ and ν. Substituting $a + b = \frac{a}{\overline{x}}$ and $b = a\frac{1-\overline{x}}{\overline{x}}$ (from Eq. (A24)) into Eq. (A25), we get:

$$\nu = \frac{a^2\frac{1-\overline{x}}{\overline{x}}}{\left(\frac{a}{\overline{x}}\right)^2\left(\frac{a}{\overline{x}}+1\right)} = \frac{\overline{x}(1-\overline{x})}{\left(\frac{a}{\overline{x}}+1\right)} \tag{A28}$$

Rewriting Eq. (A28) in terms of a as a function of $\overline{x}$ and ν leads to:

$$a = \frac{\overline{x}^2(1-\overline{x})-\overline{x}\nu}{\nu} \tag{A29}$$

Furthermore, substituting Eq. (A29) into an alternative form of Eq. (A24), i.e. $b = a\frac{1-\overline{x}}{\overline{x}}$, leads to an expression of b as a function of $\overline{x}$ and ν:

$$b = \frac{\overline{x}(1-\overline{x})^2-(1-\overline{x})\nu}{\nu} \tag{A30}$$

Now we use Eqs. (A24), (A25), (A26), (A29) and (A30) to express S as a function of $\overline{x}$ and ν. Rewriting Eq. (A24) as $a + b = \frac{a}{\overline{x}}$ or $b = a\frac{1-\overline{x}}{\overline{x}}$ and Eq. (A25) as $\nu^{-\frac{1}{2}}(a+b)^{-1} = (a+b+1)^{\frac{1}{2}}/(ab)^{\frac{1}{2}}$ and substituting them sequentially into Eq. (A26) leads to:

$$S = \frac{2(b-a)}{(a+b+2)(a+b)\nu^{\frac{1}{2}}} = \frac{2(b-a)}{\left(\frac{a}{\overline{x}}+2\right)\left(\frac{a}{\overline{x}}\right)\nu^{\frac{1}{2}}} = \frac{2a\left(\frac{1-\overline{x}}{\overline{x}}-1\right)}{\left(\frac{a}{\overline{x}}+2\right)\left(\frac{a}{\overline{x}}\right)\nu^{\frac{1}{2}}} = \frac{2\left(\frac{1-2\overline{x}}{\overline{x}}\right)}{\left(\frac{a}{\overline{x}}+2\right)\left(\frac{1}{\overline{x}}\right)\nu^{\frac{1}{2}}} = \frac{2(1-2\overline{x})}{\left(\frac{a}{\overline{x}}+2\right)\nu^{\frac{1}{2}}} \tag{A31}$$

Furthermore, substituting Eq. (A29) into Eq. (A31) now allows to write S as a function of $\overline{x}$ and ν:

$$S = \frac{2(1-2\overline{x})}{\left(\frac{\overline{x}^2(1-\overline{x})-\overline{x}\nu}{\nu\overline{x}}+2\right)\nu^{\frac{1}{2}}} = \frac{2(1-2\overline{x})}{\left(\frac{\overline{x}(1-\overline{x})}{\nu}+1\right)\nu^{\frac{1}{2}}} = \frac{2(1-2\overline{x})\nu^{\frac{1}{2}}}{(\overline{x}(1-\overline{x})+\nu)} \tag{A32}$$

Finally, we use Eqs. (A24), (A27), (A29) and (A32) to express K as a function of $\overline{x}$, ν and S. Rewriting Eq. (A27) leads to:

$$K = \frac{6(a-b)^2(a+b+1)}{ab(a+b+2)(a+b+3)} - \frac{6}{(a+b+3)} = \left(\frac{3(a+b+2)}{2(a+b+3)}\right)\left(\frac{4(b-a)^2(a+b+1)}{(a+b+2)^2ab}\right) - \frac{6}{(a+b+3)} \tag{A33}$$

Now we can simplify the first term of Eq. (A33) by rewriting the first factor and substituting the second factor by the square of S, i.e. $S^2 = \frac{4(b-a)^2(a+b+1)}{(a+b+2)^2ab}$, leading to:

$$K = \frac{2(a+b)+4(a+b)+12}{4(a+b+3)}S^2 - \frac{6}{a+b+3} = \left(\frac{(a+b)}{2(a+b+3)}+1\right)S^2 - \frac{6}{a+b+3} \tag{A34}$$

Rewriting Eq. (A24) as $a + b = \frac{a}{\overline{x}}$ and substituting it into Eq. (A34) leads to:

$$K = \left(\frac{\frac{a}{\overline{x}}}{2\frac{a}{\overline{x}}+6}+1\right)S^2 - \frac{6}{\frac{a}{\overline{x}}+3} \tag{A35}$$

Using Eq. (A29), we can express K as a function of $\overline{x}$ and ν leading to:

$$K = \left(\frac{\frac{\overline{x}(1-\overline{x})-\nu}{\nu}}{2\frac{\overline{x}(1-\overline{x})-\nu}{\nu}+6}+1\right)S^2 - \frac{6}{\frac{\overline{x}(1-\overline{x})-\nu}{\nu}+3} = \left(\frac{\overline{x}(1-\overline{x})-\nu}{2\overline{x}(1-\overline{x})-2\nu+6\nu}+1\right)S^2 - \frac{6\nu}{\overline{x}(1-\overline{x})-\nu+3\nu} = \left(\frac{\overline{x}(1-\overline{x})-\nu}{2\overline{x}(1-\overline{x})+4\nu}+1\right)S^2 - \frac{6\nu}{\overline{x}(1-\overline{x})+2\nu} \tag{A36}$$

Finally, rewriting Eqs. (A32) and (A36) in terms of the third and fourth central moments allows us to conduct a beta distribution-based moment closure plugging the following set of equations into expressions (A8), (A14) and (A22):

$$M_3 = S\nu^{\frac{3}{2}} = \frac{2(1-2\overline{x})\nu^2}{(\overline{x}(1-\overline{x})+\nu)} \tag{A37}$$

$$M_4 = K\nu^2 = \left(\left(\frac{(1-\bar{x})\bar{x}-\nu}{4\nu + 2(1-\bar{x})\bar{x}} + 1\right)S^2 - \frac{6\nu}{2\nu + (1-\bar{x})\bar{x}} + 3\right)\nu^2 \tag{A38}$$

Hence, trait distributions with smaller ν are unimodal but also often peaked (high K) and skewed (high $|S|$) depending on $\bar{x}$. In contrast, trait distributions with larger ν tend to be flat or in its extreme bimodal (low K) and symmetric ($|S| \approx 0$) (Fig. 1).

Appendix B. Sensitivity analysis I – changes in model structure

To test the robustness of our results with respect to changes in the model structure we conducted numerical simulations with three other ecological models partly used in recent studies demonstrating the suitability of aggregate model approaches. First, we investigated the predator-prey model from Merico et al. (2009) in which the trade-off between the nutrient uptake efficiency of the prey and its defense against predation imposes stabilizing selection on the trait distribution of the prey community. Second, we explored the species competition model from Norberg et al. (2001) in which recurrent changes of an environmental factor such as temperature impose fluctuating selection on different species. Third, we analyzed a modified version of our predator-prey model from the main text in which a trade-off between the vulnerability of the prey species with respect to grazing by the two specialist predators imposes disruptive selection on the trait distribution of the prey community. In all three cases, we first run numerical simulations of the full trait distribution (*FD*) model by discretizing the trait axes of the associated community (n = 101). Afterwards, we derived the equations of a corresponding aggregate model with a second-order approximation and performed a moment closure either based on a beta distribution or a normal distribution. Subsequently, we also conducted numerical simulations of the corresponding *BA* and *NA* models and compared their results to the ones of the *FD* model.

B.1. Full trait distribution model with stabilizing selection

Following Merico et al. (2009) we first consider a chemostat model with a set of prey species (C) that differ in their trait values influencing their uptake of nutrients (N) and vulnerability to predation by a single predator (B). The corresponding set of differential equations is given as follows:

$$\frac{dN}{dt} = -\frac{\mu_C}{\kappa}\int_0^1 \frac{N}{N + K_N(x)}C(x)dx + \frac{m_C}{\kappa}C_T + \delta(N_0-N) \tag{B1}$$

$$\frac{\partial C(x)}{\partial t} = \left(\mu_C \frac{N}{N + K_N(x)} - \frac{\mu_B}{\varepsilon}\frac{\varphi(x)B}{\int_0^1 \varphi(x')C(x')dx' + K_C} - m_C - \delta\right)C(\mathrm{x}) \tag{B2}$$

$$\frac{dB}{dt} = \left(\frac{\mu_B}{\kappa}\frac{\int_0^1 \varphi(x)C(x)dx}{\int_0^1 \varphi(x)C(x)dx + K_C} - m_B - \delta\right)B \tag{B3}$$

The nutrient replenishment in the chemostat depends on the nutrient concentration N_0 that enters the chemostat vessel, the dilution rate δ and the nutrient recycling through prey mortality m_C (cf. Merico et al., 2009). In contrast, the nutrient uptake of the prey follows a Michaelis-Menten Kinetic with the maximum uptake rate μ_C and the effective half-saturation constant $K_N(x) = \frac{\widetilde{K}_N}{1-\kappa-\varphi(x)^{-1}}$, which depends on the basal half-saturation constant $\widetilde{K}_N$, the fraction of energy invested in generic biomass κ and the prey's edibility $\varphi(x) = (x(1-\kappa))^{-1}$, which is a function of the prey's level of defense x. Given that $\varphi(x)(1-\kappa)$ can range from 1 to ∞ (for details see Merico et al., 2009) it is impossible to develop a beta distribution-based aggregate model that considers trait variation in the prey's edibility because the beta distribution is constrained to the open interval (0, 1). Hence, in contrast to Merico et al. (2009), we considered trait variation in the prey's defense level $x(\varphi) = (\varphi(1-\kappa))^{-1}$ that actually varies between 0 and 1. The prey species are assumed to exhibit a trade-off between their nutrient uptake efficiency $K_N(x)$ and defense level $x(\varphi)$, which determines the vulnerability to grazing by the predator. Hence, the partitioning of the remaining fraction of energy $(1-\kappa)$ between nutrient harvesting biomass and defense biomass is species-specific (for details see Merico et al., 2009). The corresponding shape of the trade-off imposes stabilizing selection on the trait distribution of the prey community. In addition, foraging on the prey follows a type II functional response with maximum growth rate μ_B, conversion efficiency ε and half-saturation constant K_C. Like the prey, the predator experiences natural mortality m_B and is washed out from the chemostat with the dilution rate δ.

In accordance with our previous results for stabilizing selection shown in the main text (Fig. 3), both aggregate models perfectly captured the macroscopic properties of the *FD* model (Fig. B1). The total biomass of the prey community and the predator biomass exhibited regular oscillations whereas the mean trait value reached an equilibrium after a short transient. As a result of stabilizing selection, the variance dropped towards zero, which strongly reduced the adaptive capacity of the prey community (cf. Merico et al., 2009).

The numerical integration of both aggregate models were up to 4 times faster than that of the *FD* model (Table B1). There was no significant difference between the simulation times of the *BA* model and the *NA* model. Hence, the results of both aggregate models are reliable under stabilizing selection and their numerical integration is computationally more efficient than running simulations of the *FD* model.

B.2. Full trait distribution model with fluctuating selection

In line with Norberg et al. (2001) we further consider a species competition model with a set of species (C) that compete for a shared resource and that exhibit differences in their trait values. The corresponding equations to describe the temporal changes in the biomasses of the different species are:

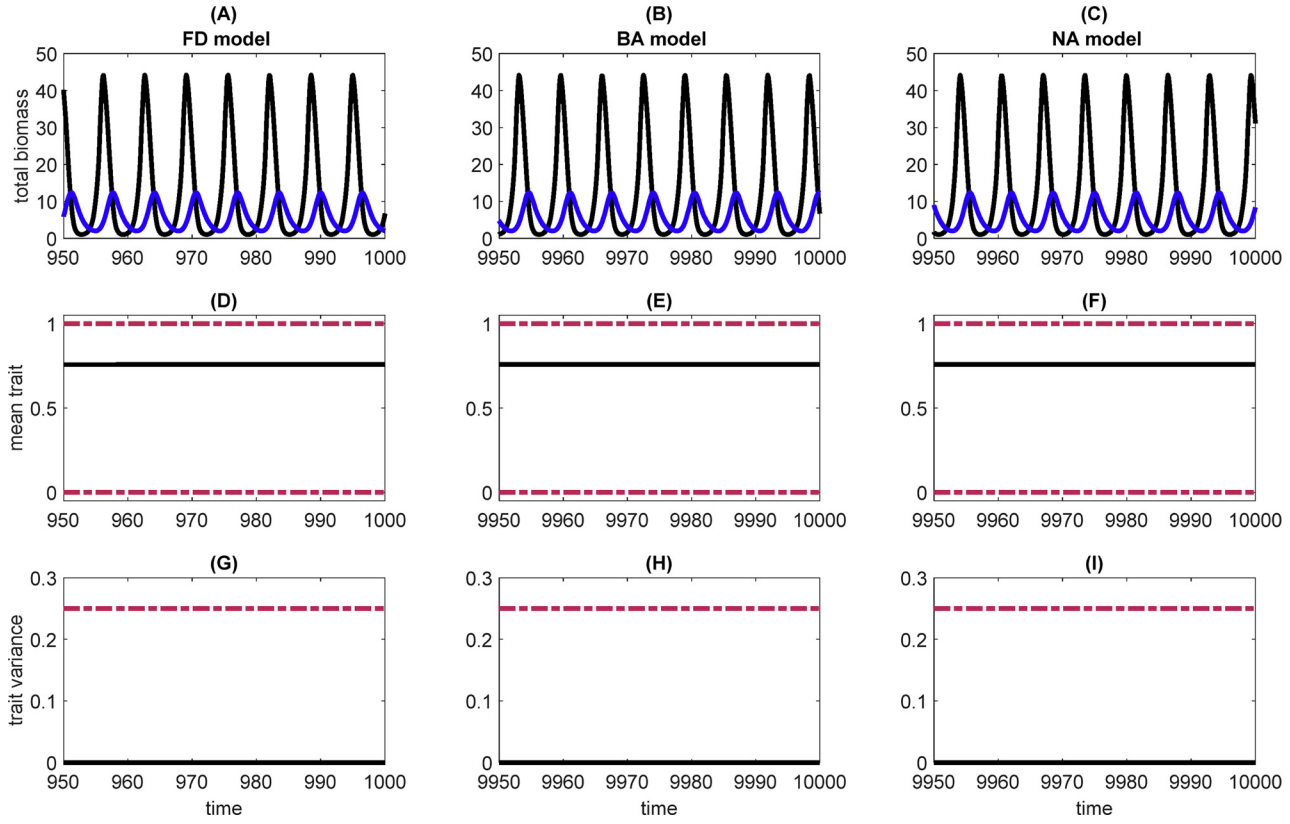

Fig. B1. Total biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (black) and the biomass dynamics (A–C) of the predator population (blue) under stabilizing selection. Results shown for the full trait distribution model (*FD* model, A, D and G), the beta distribution based aggregate model (*BA* model, B, E and H) and the normal distribution based aggregate model (*NA* model, C, F and I). The dashed-dotted red lines in (D)–(I) indicate the ecological feasible minima and maxima of the mean and variance. In line with Merico et al. (2009) we used the following parameter values: $\mu_C = 3.3$, $\mu_B = 2.25$, $\kappa = 0.8$, $m_C = 0.05$, $m_B = 0.1$, $\delta = 0.69$, $\varepsilon = 0.25$, $N_0 = 80$, $K_C = 150$, $\widetilde{K}_N = 0.4$. Note a value of $\nu \approx 0$ implies that the prey community is hardly adaptive anymore.

$$\frac{\partial C(x)}{\partial t} = \left(p \cdot e^{-\left(\frac{x-E}{\sigma}\right)^2} \cdot \left(1 - \frac{C_T}{K}\right) - d\right) C(x) + i \tag{B4}$$

The model extends logistic growth by including interspecific competition with K and C_T being the joint carrying capacity and the sum of all population biomasses, respectively. The growth rate of a species further depends on its specific phenotype x varying between 0 and 1, the maximum growth rate p and the state of a time-varying environmental factor $E = A \cdot \sin(\omega \cdot t) + M$, which oscillates around a fixed value M with amplitude A and frequency ω. To restrict E and thus the trait distribution to the ecologically feasible range (0, 1) we chose M and A to equal 0.5, respectively. The influence of the environmental changes on the species' growth rates depends on the scaling parameter σ. Hence, the recurrent changes in the environmental conditions E impose fluctuating selection on the underlying populations because different species are favored by the enviroment at different times. Furthermore, the biomasses of the populations are declining at the mortality rate d. To ensure the coexistence of all species, a constant amount of biomass of each species was assumed to enter the system through immigration. Hence, we added the immigration rate i to Eq. (B4).

In line with our previous results for fluctuating selection (Fig. 4), the *BA* model captured the overall shape of the dynamics of the total biomass, mean trait and trait variance of the *FD* model and their temporal averages quite well (Fig. B2). In contrast, the *NA* model agreed with the *FD* model only with respect to the time averages of the aggregate properties and the shape of the mean trait dynamics. Generally, the mean trait value of the community closely tracked the temporal changes of the optimal trait value currently selected by the environment (Fig. B2D–F). However, the sinusoidal shape of the environmental factor resulted in a time-averaged bimodal distribution of the optimal trait values (Kremer and Klausmeier, 2017). Consequently, the community was generally better adapted during times when the sign of the rate of change of the environmental factor was changing. This resulted in biomass dynamics where a larger peak in the total biomass at mean trait values close to the extremes followed a smaller peak in the total biomass at a mean trait value of 0.5 (cf. Norberg et al., 2001). This pattern was nicely encompassed in the results of the *BA* model but not in the ones of the *NA* model (Fig. B2A–C). In contrast to the *FD* and *BA* models, the biomass dynamics of the *NA* model exhibited a lower

Table B1
Average simulation time (mean ± standard deviation in seconds, number of repetitions = 10) of the full trait distribution model (*FD*) as well as beta distribution based (*BA*) and normal distribution based (*NA*) aggregate models for stabilizing (cf. Fig. B1), fluctuating (cf. Fig. B2) and disruptive (cf. Fig. B3) selection when numerically integrating the corresponding models for 10,000 time steps.

Selection Pressure	*FD* model	*BA* model	*NA* model
Stabilizing	116.63 ± 2.76	29.92 ± 0.14	29.73 ± 0.18
Fluctuating	3.79 ± 0.25	2.26 ± 0.04	2.15 ± 0.05
Disruptive	487.56 ± 39.58	14.84 ± 0.30	17.67 ± 0.78

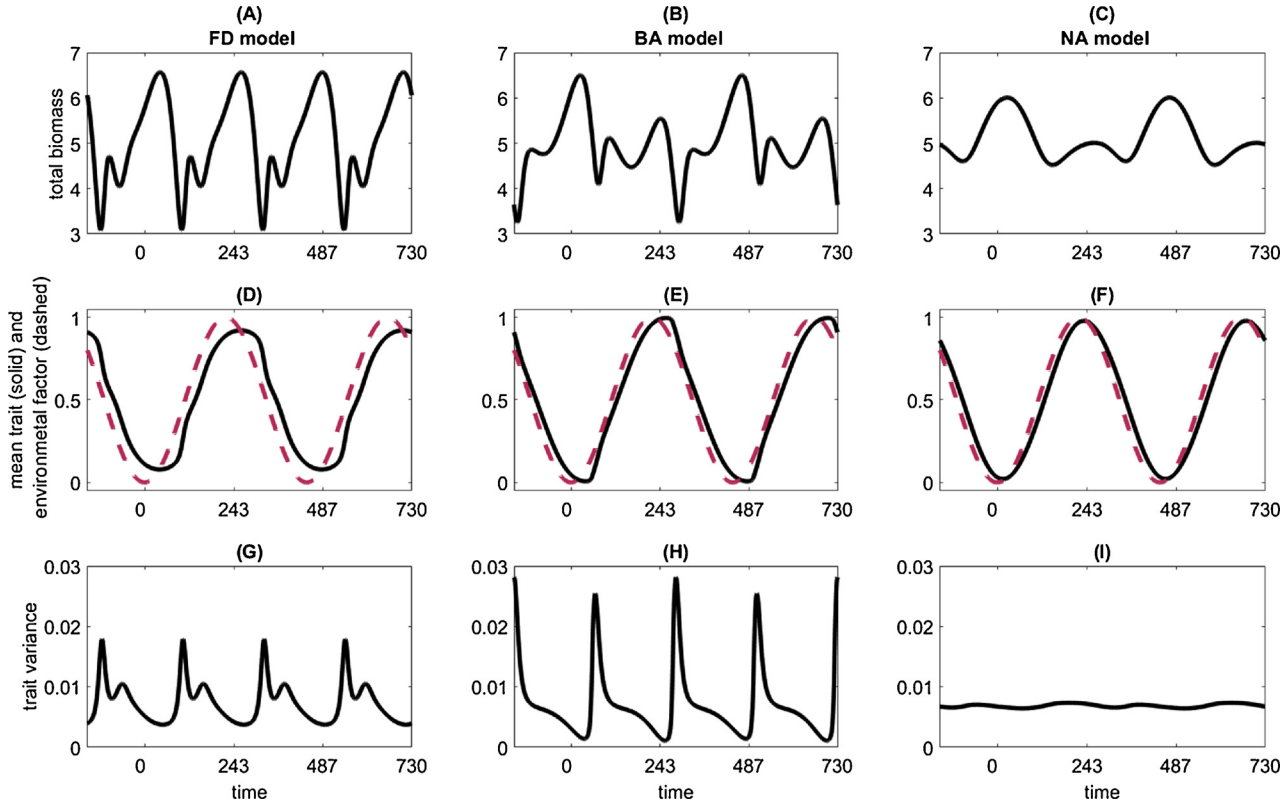

Fig. B2. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the ecological community (black) under fluctuating selection. Results shown for the full trait distribution model (*FD* model, A, D and G), the beta distribution based aggregate model (*BA* model, B, E and H) and the normal distribution based aggregate model (*NA* model, C, F and I). In line with Norberg et al. (2001) we used the following parameter values: $p = 0.6$, $\sigma = 0.2$, $K = 10$, $i = 0.01$, $d = 0.2$, $\omega = 2{\cdot}\pi/365$, $A = 0.5$, $M = 0.5$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

frequency and two differently-sized peaks, that were only associated with mean trait values close to the extremes. Furthermore, when the environmental conditions changed towards more intermediate optimal trait values, a larger range of trait values was favored promoting simultaneous growth of many different species. Consequently, in the *FD* model we observed an overall increase in the trait variance that was interrupted by a short period when the variance declined again. The latter resulted from a strong reduction in the biomasses of species exhibiting more extreme trait values. Hence, the variance dynamics of the *FD* model showed a larger peak in the trait variance that preceded a smaller peak (for details see Fig. 2 from Norberg et al., 2001). In contrast, when the environmental conditions selected for extreme trait values, we observed a strong reduction in the trait variance. Although the smaller peak of the trait variance was not well expressed in the *BA* model, the general course of the temporal dynamics of the trait variance was still very similar to the *FD* model. The *BA* model reflects the frequency and the amplitude of the dynamics as well as the presence of a pronounced peak in the trait variance at a particular moment in time. In contrast, the *NA* model showed hardly any temporal variation in the trait variance. Hence, the accuracy of our novel moment closure method based on the beta distribution strongly exceeded the one of a corresponding normal distribution-based moment closure approach under fluctuating selection. Our result is further supported by the good performance of a data-based aggregate predator-prey model from Norberg et al. (2001) that accounted for important statistical relationships between lower and higher moments.

The numerical integration of both aggregate models was almost twice as fast as that of the *FD* model whereas the simulation times of the *BA* model and the *NA* model were similar (Table B1).

B.3. Full trait distribution model with disruptive selection

Finally, we consider a model with two different predators (C) that are grazing upon a set of prey species (R) with continuous trait values. In contrast to our predator-prey model from the main text, we assumed the shape of the feeding functions to be sigmoidal instead of unimodal. Furthermore, we assumed the strength of resource competition to differ between the different prey species. The corresponding population biomasses thus change according to the following set of equations:

$$\frac{\partial R(x)}{\partial t} = r\left(1-\frac{\int_0^1 c(x',x)R(x')dx'}{\kappa}\right)R(x)-\sum_{i=1}^{2}\frac{a_i(x)R(x)C_i}{1+h\int_0^1 a_i(x')R(x')dx'}+I \tag{B5}$$

$$\frac{dC_i}{dt} = \left(\varepsilon\frac{\int_0^1 a_i(x)R(x)dx}{1+h\int_0^1 a_i(x)R(x)dx}-d\right)C_i \tag{B6}$$

$$c(x,x') = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x'-x}{\sigma}\right)^2} \tag{B7}$$

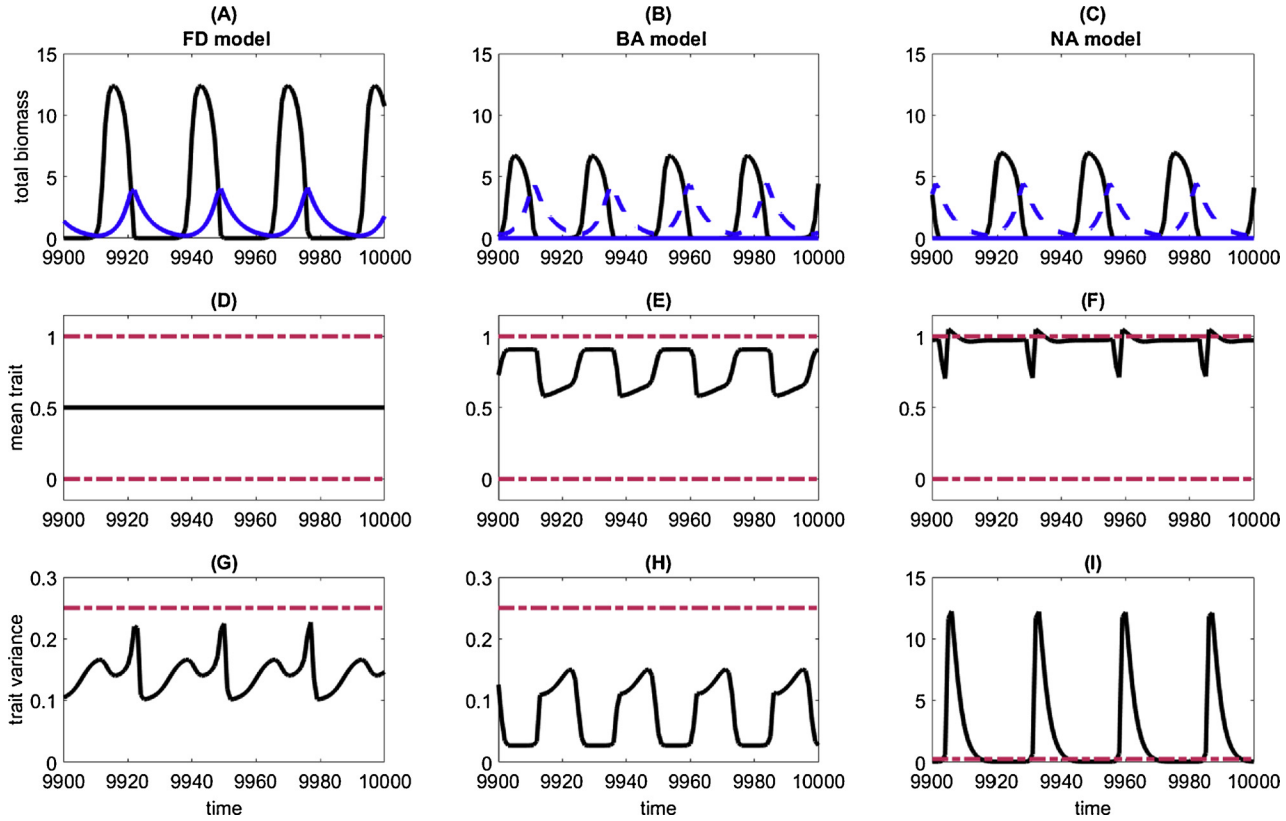

Fig. B3. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (black) and the biomass dynamics (A–C) of the two different predator populations (solid and dashed blue) under disruptive selection. Results shown for the full trait distribution model (*FD* model, A, D and G), the beta distribution based aggregate model (*BA* model, B, E and H) and the normal distribution based aggregate model (*NA* model, C, F and I). The dashed-dotted red lines in (D)–(I) indicate the ecological feasible minima and maxima of the mean and variance. We used the following parameter values: $r = 2$, $\kappa = 10$, $a_{max} = 2$, $a_{min} = 0, h = 0.5$, $w = 0.75$, $s = 8$, $I = 0.0001$, $d = 0.2$, $\varepsilon = 0.3, \sigma = 0.2$.

$$a_1(x) = (a_{max} - a_{min})(1 + e^{s(x-w)})^{-1} + a_{min} \tag{B8}$$

$$a_2(x) = (a_{max} - a_{min})(1 - (1 + e^{s(x-(1-w))})^{-1}) + a_{min} \tag{B9}$$

In the absence of predation, the prey species exhibit logistic growth with r and κ being the maximum growth rate and joint carrying capacity, respectively. Given that trait differences may result in niche differences between species, the strength of competition may depend on the species' trait similarity. Hence, we implemented the competition kernel c (MacArthur and Levins, 1967) and thus assumed that prey species with more similar trait values are competing more strongly with each other than species with very different trait values. This is in line with the assumption that intraspecific competition of the prey is larger than interspecific competition, which stabilizes species coexistence. The width of the competition kernel is determined by the scaling parameter σ. Furthermore, predation follows a type II functional response with a common handling time h and the predator-specific attack rates a_1 and a_2 with minimum and maximum values of a_{min} and a_{max}. Hence, the predation rates of the two different predators depend on the phenotype x of the individual prey species. In line with previous work from Abrams (2006) and Schreiber et al. (2011), we assumed a trade-off between the prey's vulnerability to predation by the two specialist predators, the shape of which depend on the parameters s and w. By using a value of w equal to 0.75, we generated a concave-shaped trade-off curve that imposes disruptive selection on the trait distribution of the prey community. Nevertheless, we ensured the persistence of all prey by adding the immigration rate I to Eq. (B5).

In line with our previous results for disruptive selection displayed in the main text (Fig. 6), both aggregate models failed to reproduce the dynamics of the aggregate properties of the *FD* model. Here, the total biomasses of the prey and predator communities showed regular predator-prey cycles because the population biomasses of the two predators were synchronized (Fig. B3A). In addition, the trait distribution of the prey community was symmetric around the mean trait value of 0.5 throughout time (Fig. B3D). Depending on the respective strength of competition and predation, the shape of the trait distribution of the prey community was either rather uniform or bimodal. At low prey biomasses, immigration strongly influenced the population dynamics promoting uniform trait distributions of the prey community with variance values down to 1/12 (Fig. B3A and G). In contrast, a high predation pressure on the prey species with intermediate trait values at high prey biomasses selected for bimodal trait distributions with variance values up to 0.2. Neither the *BA* model nor the *NA* model was able to reproduce these patterns. Despite showing regular predator-prey oscillations, both aggregate models underestimated the amplitude of the total biomass of the prey community (Fig. B3A–C). Furthermore and in contrast to the *FD* model, one of the predators went extinct. However, in line with our previous results (e.g. Fig. 5) the values of the mean trait and trait variance stayed within the ecologically feasible ranges in the *BA* model (Fig. B3E and H). In contrast, the aggregate properties surpassed their natural limits in the *NA* model (Fig. B3F and I).

The numerical integration of both aggregate models was about 30 times faster than that of the *FD* model, but this strong reduction in simulation time comes at a too high cost, given that the results of the aggregate models are not reliable. Unexpectedly, the simulation time of the *NA* model was somewhat longer than the one of the *BA* model (Table B1). While the Gaussian-based approximation is valid in a narrower parameter range than the beta approximation, the *NA* model may be more useful in practice if both model types are sufficiently accurate. This may follow from the fact that the *NA* model is

far simpler than the *BA* model and so it should be both computationally faster to calculate, and easier to work with analytically. However, the computation times depend not only on the complexity of the moment-closure terms themselves, but also on the stiffness of the resulting equations that strongly depends on the derivatives of the fitness landscape, and this favors in general the beta distribution when the trait range is finite. Furthermore, reliable trait-based aggregate models should not only provide shorter simulation times, but also be easier to fit to data (cf. Fig. 2).

Appendix C. Sensitivity analysis II – changes in parameter values

To test the robustness of our results with respect to parameter changes we conducted numerical simulations with different carrying capacities ($\varkappa \in \{10^{0.7}, 10^{0.71}, ...,10^{1.3}\}$) and minimum attack rates ($a_{min} \in \{2^{-1}, 2^{-1.1}, ...,2^{-3}\}$), thereby simulating changes in the bottom-up regulation and top-down control of the prey community under fluctuating selection. In addition, we tested for a potential dependency of the biomass and trait dynamics on immigration by running further simulations of the different models for very different immigration rates ($I \in \{10^{-2}, 10^{-2.1}, ...,10^{-6}\}$). All other parameter values were the same as before (see Table 1). For a few simulations, we had to increase the precision of the solver reducing the relative and absolute tolerances, and maximum step size (cf. Figs. C2 and C3, grey shaded areas).

The corresponding results show that the temporal mean and variance of the biomass and trait dynamics are very similar across the range of parameter values evaluated for the three different parameters (Figs. C1–C3). Furthermore, the exact shape of the trait and biomass dynamics of the beta distribution-based aggregate (*BA*) and full trait distribution (*FD*) models also agreed very well for most of the three parameter ranges considered, i.e. for $-1.3 \geq log_2(a_{min}) \geq -2.7$, for $20 \geq \varkappa \geq 8$ and for $-2 \geq log_{10}(I) \geq -5.2$ (for a representative example see Fig. 4). Hence, the general results described in the main text are fairly robust with respect to parameter changes. However, as with the results of the normal distribution-based aggregate (*NA*) model for the entire ranges of the three parameters, the biomass and trait dynamics of the *BA* model were very different from the

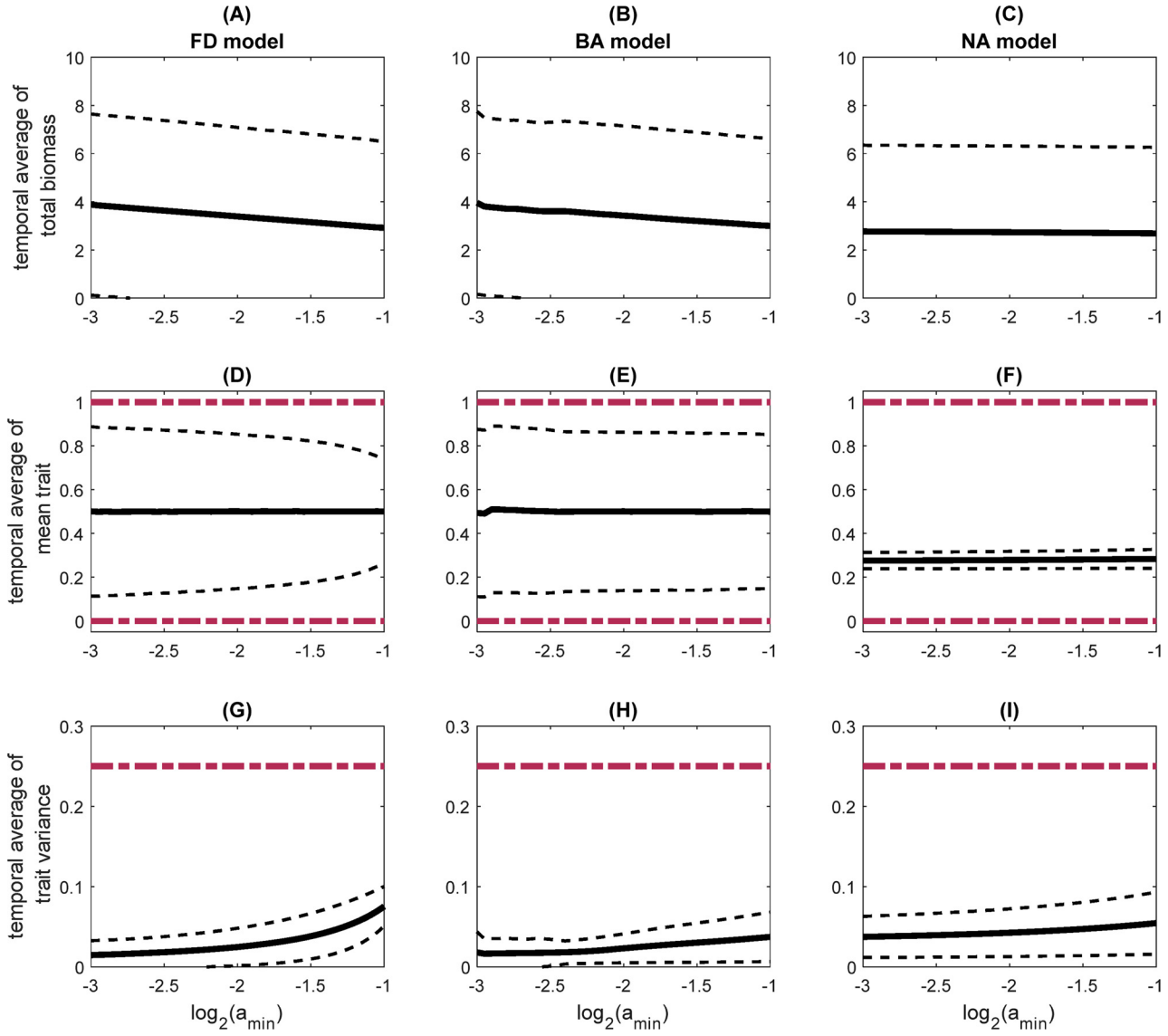

Fig. C1. Temporal mean (solid line) ± standard deviation (dashed line) of the total biomass (A–C), mean trait (D–F) and trait variance (G–I) of the prey community in dependence of the minimum attack rate a_{min} for the *FD* model (A, D, G), *BA* model (B, E, H) and *NA* model (C, F, I) under fluctuating selection. The dashed-dotted red lines in (D)–(I) indicate the ecologically feasible minima and maxima of the mean and variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

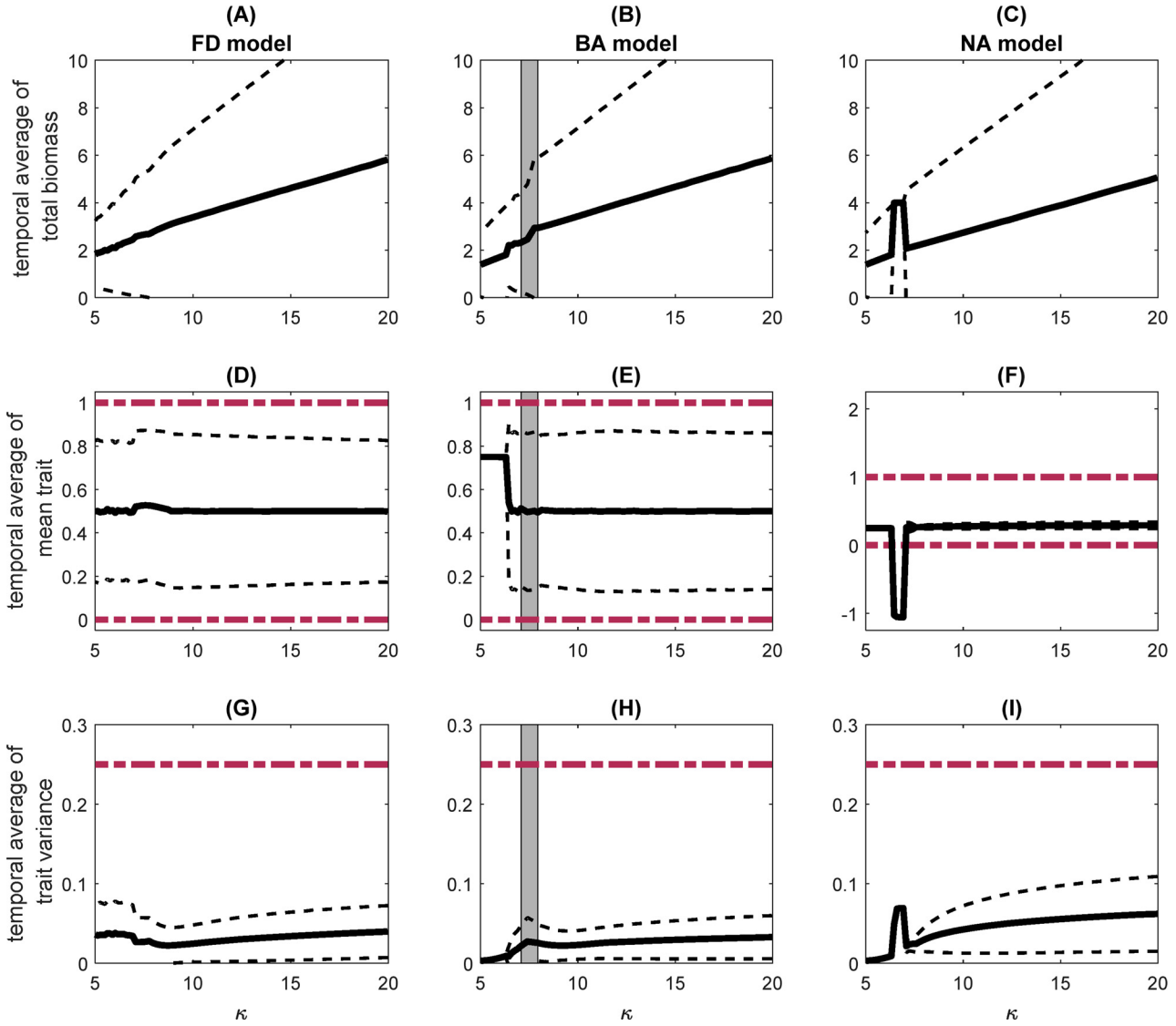

Fig. C2. As Fig. C1 but in dependence of the carrying capacity κ. Note that it was quite demanding ($7 < \kappa < 8$) or even impossible ($\kappa \approx 7.24$ and $\kappa \approx 7.59$) to solve the *BA* model numerically in a small region of the parameter space (indicated by the grey shaded area). Furthermore, the strong deviations in the general patterns of the *NA* model between $\kappa \approx 6.3$ and $\kappa \approx 7.1$ and of the *BA* model between $\kappa \approx 5$ and $\kappa \approx 6.3$ indicate the existence of alternative states as shown in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ones of the *FD* model in case of very low immigration rates and rather low carrying capacities. The former result implies that a small amount of immigration and thus standing trait variation is necessary to ensure ongoing oscillations in the biomass and trait values of the *BA* model and thus a good match with the *FD* model. This happens because the second order correction of $d\overline{x}/dt$ is not strong enough to prevent $\overline{x}$ from becoming stuck at a local fitness maximum when immigration, and thus standing trait variation, as well as the third central moment are too low. In addition, reducing the carrying capacity to rather low values strongly stabilized the predator-prey biomass dynamics (Fig. C2). As a result, changes in the fitness landscape were less pronounced so that it became more likely for $\overline{x}$ to get stuck at a persistent local fitness maximum, despite the second order correction. Based on the previous results (Fig. 5), we can expect that the *NA* model comprises alternative stable states throughout the ranges of the different parameters. In addition, because the food-web structure is symmetric, we also expect that the *BA* model will exhibit an alternative state for lower values of the carrying capacity and low values of the immigration rate.

The numerical integration of the *BA* model was quite demanding or even not feasible in small regions of the parameter space of κ and I (Figs. C2 and C3; grey shaded areas). A possible explanation is that the assumed shape of a trait distribution generally imposes some constraints on the conditions (including initial conditions and parameters) under which a corresponding aggregate model can provide an adequate representation of the biomass and trait dynamics. For example, in the *BA* model, the shape of the trait distribution of the immigrating community is assumed to follow a uniform distribution whereas the one of the resident community is assumed to follow a beta distribution. However, as is also the case with the normal distribution, the sum of a beta and a uniform distribution does not necessarily have to be beta-distributed. In such a case, the relationships between the lower and higher moments cannot be adequately described anymore by the moment closure equations of the beta distribution. Hence, the beta distribution-based moment closure may interfere with the mechanism assumed to maintain the trait variance. As a potential result, the *BA* model may not have a suitable solution, or numerical integration may get very laborious due to stiffness.

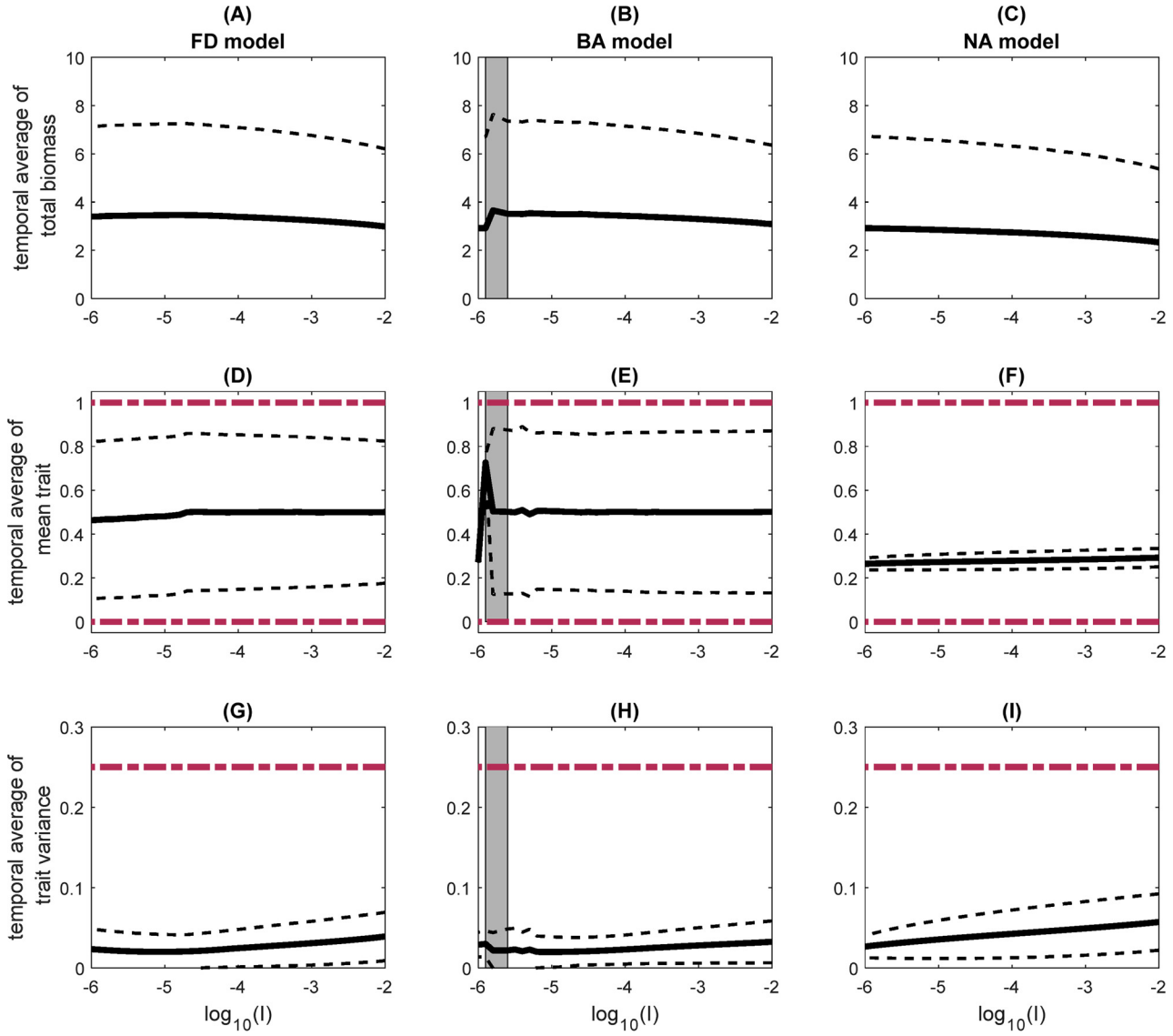

Fig. C3. As Fig. C1 but in dependence of the immigration rate I. Note that it was quite demanding ($10^{-5.9} < I < 10^{-5.6}$) or even impossible ($I = 10^{-5.6}$) to solve the *BA* model numerically in a small region of the parameter space (indicated by the grey shaded area). Furthermore, the strong deviation in the general pattern of the *BA* model between $I \approx 10^{-5.9}$ and $I \approx 10^{-6}$ indicates the emergence of alternative states as shown for the results of the *NA* model in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix D. Skewness restricts mean trait values to the ecologically feasible range

In contrast to the *NA* model, the mean trait value ($\overline{x}$) of the *BA* model always stayed within its ecologically feasible range because the beta distribution is intrinsically bounded. Here we show that this feature may generally arise from a strong increase in skewness (S) when $\overline{x}$ approaches one of the extremes. To focus on the effect of changes in S alone on the biomass and trait dynamics, we derive here an aggregate model based on the assumption that the shape of the trait distribution follows a special case of the *Johnson* SB distribution (Johnson, 1949) and, thus, becomes skewed when $\overline{x}$ approaches one of the extremes (for details see below). Nevertheless, the initial approximations made in the derivation ensure that no other terms arise that may reflect accompanying changes in the kurtosis (K) of the trait distribution. Hence, any effect that arises has to be attributed to S alone. In accordance with the analysis performed for the model comparison presented in the main text, we conducted numerical simulations of our *Johnson* SB distribution-based aggregate model (*JA* model) under different selection regimes. All corresponding simulations lasted for 10^4 time units.

Overall, the simulation results of the *JA* model were very similar to the ones of the *NA* model (compare Fig. D1 to Figs. 3, 5 and 7 of the main text). However, the mean trait value ($\overline{x}$) of the *JA* model never exceeded its ecologically feasible range, which contrasts with the results of the *NA* model. This feature was independent of the selection regime imposed (compare Fig. D1 to Figs. 5 and 7 of the main text). This is explained by the strong decline of ν due to increasingly large absolute values of S when $\overline{x}$ approaches one of the extremes, quickly slowing down $d\overline{x}/dt$. Hence, an increase in $|S|$ limits the changes in $\overline{x}$ indirectly via changes in ν. This feedback between $d\overline{x}/dt$ and $d\nu/dt$ suggests that the skewness and thus asymmetry of a trait distribution is very important for preventing the emergence of unrealistic values of the aggregate properties. We acknowledge that, in addition to the indirect effect of S, $\overline{x}$ may also be more directly limited in the *BA* model via the second term of Eq. (13). For example, in the simplest case of disruptive selection, the fitness gradient drives $\overline{x}$ towards one of the extreme values (cf. Fig. E1B, E). However, since the second derivative is positive, the sign and value of S operate in the opposite direction to the fitness gradient (cf. Fig. E1E, H) slowing down $d\overline{x}/dt$ and likely preventing $\overline{x}$ from overcoming its limits. Neither mechanism described can act in the *NA* model to restrict mean trait values to their ecologically

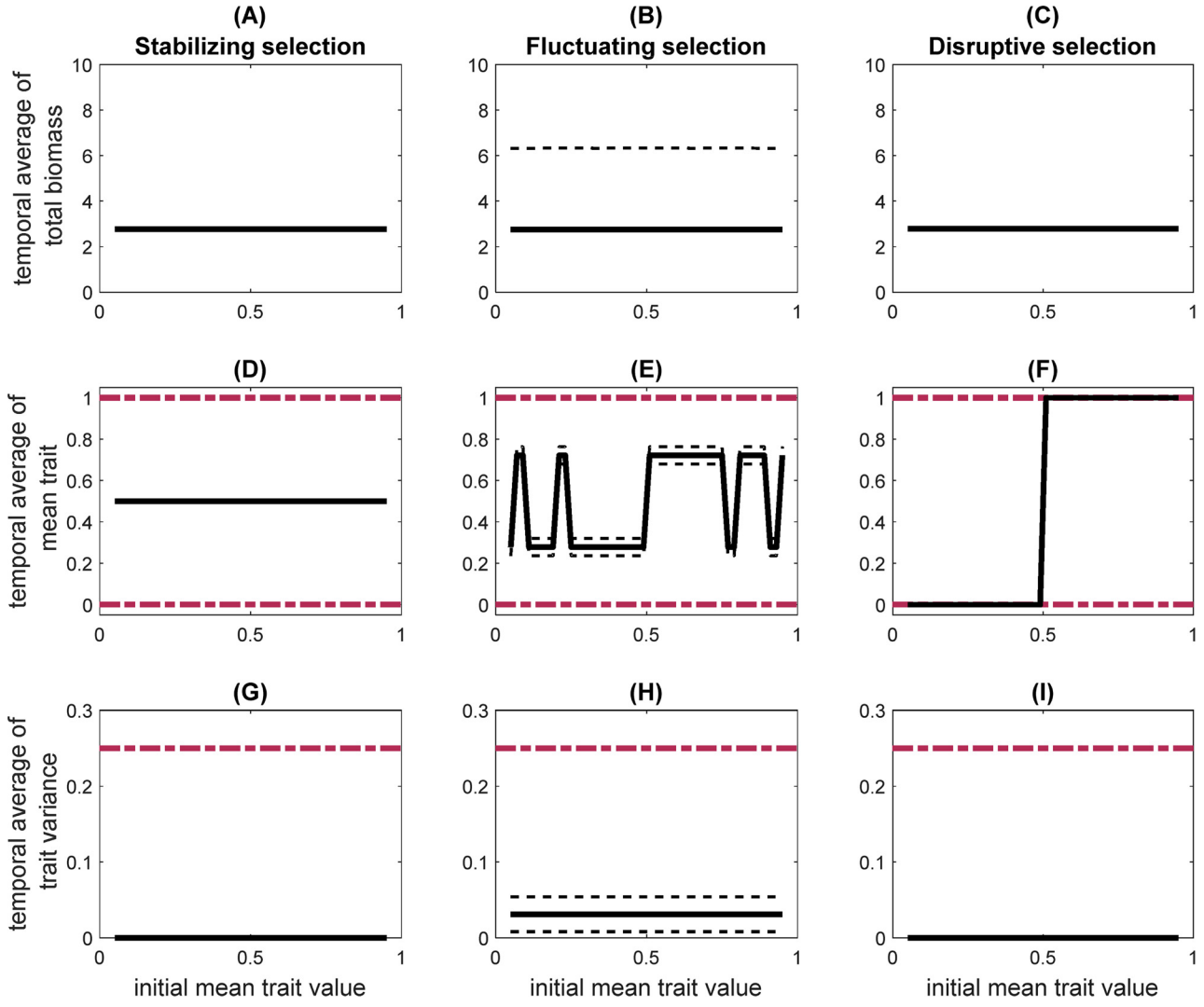

Fig. D1. Temporal mean (solid line) ± standard deviation (dashed line) of the total biomass (A–C), mean trait (D–F) and trait variance (G–I) of the prey community in dependence of the initial mean trait value for the *JA* model under stabilizing (A, D, G), fluctuating (B, E, H) and disruptive (C, F, I) selection. The dashed-dotted red lines in (D)–(I) indicate the ecological feasible minima and maxima of the mean and variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

feasible ranges.

Derivation of the Johnson SB distribution-based aggregate (JA) model

In contrast to Appendix A we now assume that the trait x is a function of another trait y, whose values are assumed to be normally distributed. Following Abrams (1999) we used a sigmoidal function to describe the transformation of the trait y (defined on a continuous infinite range) to the trait x (defined on a continuous finite range between zero and one):

$$x(y) = \frac{e^y}{1 + e^y} \tag{D1}$$

Hence, the distribution of x is a special case of the *Johnson* SB distribution, i.e., a logit-normal distribution with parameters $\gamma = 0$, $\delta = 1$, $\mu = \overline{y}$ and $\sigma = \sqrt{\nu_y}$ (Johnson, 1949). Although the *Johnson* SB distribution may in principle also be bimodal, our derivation below does not give rise to terms that may reflect excess kurtosis and thus bimodal shape properties. When conducting a moment closure based on the assumption of normally distributed trait values y, i.e. $y \sim N(\overline{y}, \nu_y)$, a second-order approximation of the equation describing the temporal dynamics of $\overline{y}$ is given by (see Eq. (16) of the methods section):

$$\frac{d\overline{y}}{dt} = \nu_y \frac{\partial F}{\partial y}\bigg|_{y=\overline{y}} \tag{D2}$$

where F is the per capita net growth rate (fitness) of the species as a function of y. The mean trait $\overline{x}$ is given by:

$$\overline{x} = \int N(y)x(y)dy \tag{D3}$$

A Taylor expansion of $x(y)$ around $\bar{y}$ yields:

$$\bar{x} = \int_{-\infty}^{\infty} N(y) \sum_{k=0}^{\infty} \frac{(y-\bar{y})^k}{k!} \frac{\partial^k x(y)}{\partial y^k}\bigg|_{y=\bar{y}} dy = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k x(y)}{\partial y^k}\bigg|_{y=\bar{y}} \underbrace{\int_{-\infty}^{\infty} N(y)(y-\bar{y})^k dy}_{M_k} \tag{D4}$$

For simplicity, we will use a first order approximation of $\bar{x}$:

$$\bar{x} = \left(x(y)|_{y=\bar{y}} + \frac{\nu_y}{2} \frac{\partial^2 x(y)}{\partial y^2}\bigg|_{y=\bar{y}} + \frac{M_3}{6} \frac{\partial^3 x(y)}{\partial y^3}\bigg|_{y=\bar{y}} + \frac{M_4}{24} \frac{\partial^4 x(y)}{\partial y^4}\bigg|_{y=\bar{y}} + \ldots \right) \approx x(\bar{y}) \tag{D5}$$

Applying twice the chain rule, the temporal dynamics of $\bar{x}$ is described by:

$$\frac{d\bar{x}}{dt} = \frac{d\bar{x}}{d\bar{y}} \frac{d\bar{y}}{dt} = \frac{d\bar{x}}{d\bar{y}} \nu_y \frac{\partial F}{\partial y}\bigg|_{y=\bar{y}} = \frac{d\bar{x}}{d\bar{y}} \nu_y \left(\frac{dx}{dy} \frac{\partial F}{\partial x} \right)\bigg|_{x=\bar{x},\, y=\bar{y}} = \nu_y \left(\frac{d\bar{x}}{d\bar{y}} \right)^2 \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \tag{D6}$$

Based on Eq. (D5), the derivative of $\bar{x}$ with respect to $\bar{y}$ is given by:

$$\frac{d\bar{x}}{d\bar{y}} = \frac{e^{\bar{y}}(1+e^{\bar{y}}) - e^{\bar{y}} e^{\bar{y}}}{(1+e^{\bar{y}})^2} = \frac{e^{\bar{y}}}{1+e^{\bar{y}}} \left(\frac{1+e^{\bar{y}} - e^{\bar{y}}}{1+e^{\bar{y}}} \right) = \bar{x}(1-\bar{x}) \tag{D7}$$

Inserting Eq. (D7) into Eq. (D6) yields:

$$\frac{d\bar{x}}{dt} = \nu_y (\bar{x}(1-\bar{x}))^2 \frac{\partial F}{\partial x}_{x=\bar{x}} \tag{D8}$$

In analogy to Eq. (16) (or Eq. (D2)), the product in front of the first derivative of Eq. (D8) is akin to the variance of the trait x. Hence, defining $\nu_x := \nu_y (\bar{x}(1-\bar{x}))^2$, Eq. (D8) becomes:

$$\frac{d\bar{x}}{dt} = \nu_x \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \tag{D9}$$

Proceeding as before, a second-order approximation of the equation describing the temporal dynamics of ν_y is given by (see Eq. (17) of the methods section):

$$\frac{d\nu_y}{dt} = \nu_y^2 \frac{\partial^2 F}{\partial y^2}\bigg|_{y=\bar{y}} \tag{D10}$$

Applying first the chain rule and then using Eqs. (D9) and (D10), the equation describing the temporal dynamics of ν_x is given by:

$$\frac{d\nu_x}{dt} = \frac{\partial \nu_x}{\partial \nu_y} \frac{d\nu_y}{dt} + \frac{\partial \nu_x}{\partial \bar{x}} \frac{d\bar{x}}{dt} = \frac{\partial \nu_x}{\partial \nu_y} \nu_y^2 \frac{\partial^2 F}{\partial y^2}\bigg|_{y=\bar{y}} + \frac{\partial \nu_x}{\partial \bar{x}} \nu_x \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \tag{D11}$$

Given that $\nu_y = \nu_x (\bar{x}(1-\bar{x}))^{-2}$we can express $\partial \nu_x / \partial \bar{x}$ as:

$$\frac{\partial \nu_x}{\partial \bar{x}} = 2\nu_y (\bar{x}(1-\bar{x}))(1-2\bar{x}) = \nu_x \frac{2(1-2\bar{x})}{\bar{x}(1-\bar{x})} \tag{D12}$$

Inserting Eq. (D12), $\partial \nu_x / \partial \nu_y = (\bar{x}(1-\bar{x}))^2$, and $\nu_y^2 = \nu_x^2 (\bar{x}(1-\bar{x}))^{-4}$ into Eq. (D11) yields:

$$\frac{d\nu_x}{dt} = \nu_x^2 (\bar{x}(1-\bar{x}))^{-2} \frac{\partial^2 F}{\partial y^2}\bigg|_{y=\bar{y}} + \nu_x^2 \frac{2(1-2\bar{x})}{\bar{x}(1-\bar{x})} \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \tag{D13}$$

Rewriting $\frac{\partial^2 F}{\partial y^2}\Big|_{y=\bar{y}}$ leads to:

$$\frac{\partial^2 F}{\partial y^2}\bigg|_{y=\bar{y}} = \frac{\partial \left(\frac{\partial F}{\partial y} \right)}{\partial y}\bigg|_{y=\bar{y}} = \frac{\partial \left(\left(\frac{dx}{dy} \frac{\partial F}{\partial x} \right)\Big|_{x=\bar{x}} \right)}{\partial y}\bigg|_{y=\bar{y}} = \frac{\partial^2 F}{\partial x \partial y}\bigg|_{x=\bar{x},\, y=\bar{y}} \frac{d\bar{x}}{d\bar{y}} + \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \frac{d^2\bar{x}}{d\bar{y}^2} = \frac{\partial^2 F}{\partial x^2}\bigg|_{x=\bar{x}} \frac{d\bar{x}}{d\bar{y}} \frac{d\bar{x}}{d\bar{y}} + \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \frac{d^2\bar{x}}{d\bar{y}^2} = \frac{\partial^2 F}{\partial x^2}\bigg|_{x=\bar{x}} \left(\frac{d\bar{x}}{d\bar{y}} \right)^2 + \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \frac{d^2\bar{x}}{d\bar{y}^2} \tag{D14}$$

Using Eq. (D7), the second derivative of $\bar{x}$ with respect to $\bar{y}$ is given by:

$$\frac{d^2\bar{x}}{d\bar{y}^2} = \frac{d\left(\frac{d\bar{x}}{d\bar{y}} \right)}{d\bar{y}} = \frac{d(\bar{x}(1-\bar{x}))}{d\bar{y}} = \left(\frac{d\bar{x}}{d\bar{y}}(1-\bar{x}) - \bar{x} \frac{d\bar{x}}{d\bar{y}} \right) = \frac{d\bar{x}}{d\bar{y}}(1-2\bar{x}) = \bar{x}(1-\bar{x})(1-2\bar{x}) \tag{D15}$$

Inserting Eq. (D7) and Eq. (D15) into Eq. (D14) yields:

$$\frac{\partial^2 F}{\partial y^2}\bigg|_{y=\bar{y}} = \left(\frac{\partial^2 F}{\partial x^2}\bigg|_{x=\bar{x}} (\bar{x}(1-\bar{x}))^2 + \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} \bar{x}(1-\bar{x})(1-2\bar{x}) \right) \tag{D16}$$

Now, substituting Eq. (D16) into Eq. (D13) and simplifying the resulting expression gives:

$$\frac{d\nu_x}{dt} = \nu_x^2\left(\frac{\partial^2 F}{\partial x^2}\bigg|_{x=\bar{x}} + \frac{3(1-2\bar{x})}{\bar{x}(1-\bar{x})}\frac{\partial F}{\partial x}\bigg|_{x=\bar{x}}\right) \tag{D17}$$

In line with Eq. (14) of the methods section, we can rewrite Eq. (D17) as:

$$\frac{d\nu_x}{dt} = S\nu_x^{\frac{3}{2}}\frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} + \nu_x^2\frac{\partial^2 F}{\partial x^2}\bigg|_{x=\bar{x}} \tag{D18}$$

with $S := \frac{3(1-2\bar{x})\sqrt{\nu_x}}{\bar{x}(1-\bar{x})}$ representing the skewness of the trait distribution. Notice that S is symmetrical around 0.5 and tends to (positive or negative) infinity when $\bar{x}$ approaches the extremes of the trait range, 0 and 1. Finally, we add the immigration terms of Eqs. (13) and (14) to Eqs. (D9) and (D18) leading to the following approximated equations of an aggregate model based on the assumption that x follows a unimodal but skewed trait distribution:

$$\frac{d\bar{x}}{dt} = \nu_x\frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} + \frac{I}{R_T}\left(\frac{1}{2}-\bar{x}\right) \tag{D19}$$

$$\frac{d\nu_x}{dt} = S\nu_x^{\frac{3}{2}}\frac{\partial F}{\partial x}\bigg|_{x=\bar{x}} + \nu_x^2\frac{\partial^2 F}{\partial x^2}\bigg|_{x=\bar{x}} + \frac{I}{R_T}\left(\frac{1}{12}-\nu_x+\left(\frac{1}{2}-\bar{x}\right)^2\right) \tag{D20}$$

This *Johnson* SB distribution-based aggregate model (*JA* model) underlies the results presented in Fig. D1.

Appendix E. Skewness and kurtosis jointly promote the maintenance of functional diversity

We found strong negative correlations between the skewness (S) and the mean ($\bar{x}$), and between the kurtosis (K) and the variance (ν) of the observed phytoplankton size distributions that are well captured by the beta distribution (Fig. 2). Here, we show that these relationships may provide a general and intrinsic mechanism to promote the maintenance of functional diversity, i.e. strictly positive values of the trait variance, in trait-based aggregate models, even in the absence of frequency-dependent selection. In order to do so, we analyzed the stability of potential fixed points of the approximated aggregate model equations under simplifying assumptions for two different moment closure approaches, and run additional simulations of an extended version of the *BA* model.

E.1. Convergent stable fitness minima under frequency-independent selection

According to Appendix A, the equations describing the temporal dynamics of $\bar{x}$ and ν of a full trait distribution (*FD*) model, which does not contain immigration, are approximated up to the second order as:

$$\frac{d\bar{x}}{dt} = \nu\frac{\partial F(x)}{\partial x}\bigg|_{x=\bar{x}} + \frac{S\nu^{\frac{3}{2}}}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\bar{x}} \tag{E1}$$

$$\frac{d\nu}{dt} = S\nu^{\frac{3}{2}}\frac{\partial F(x)}{\partial x}\bigg|_{x=\bar{x}} + \frac{(K\nu^2-\nu^2)}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\bar{x}} \tag{E2}$$

For convenience, we will first neglect frequency-dependent natural selection, which allows us to explore a novel mechanism that can stabilize a fixed point of the mean trait value at a fitness minimum. Hence, we suppose the species' fitness function F to depend only on the total biomass R_T and its individual trait value x but not on the mean and variance of the entire (frequency) trait distribution (cf. Abrams et al., 1993b). However, our results may also apply to some (special) cases of frequency-dependent selection as discussed in more detail below. Assuming a normal trait distribution enables a moment closure of the aggregate model based on $S = M_3\nu^{-\frac{3}{2}} = 0$ and $K = M_4\nu^{-2} = 3$ and thus allows a simplification of Eqs. (E1) and (E2) to:

$$\frac{d\bar{x}}{dt} = \nu\frac{\partial F(x)}{\partial x}\bigg|_{x=\bar{x}} \tag{E3}$$

$$\frac{d\nu}{dt} = \nu^2\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\bar{x}} \tag{E4}$$

Hence, the mean trait $\bar{x}$ changes in the direction of the local fitness gradient and thus moves towards a local maximum of the fitness landscape $F(x)$, where it remains constant once reaches it, because the local fitness gradient $\partial F/\partial x|_{x=\bar{x}}$ is zero at this point (Fig. E1). At or near a local fitness maximum, the second derivative of the fitness landscape evaluated at the mean trait value $\partial^2 F/\partial x^2|_{x=\bar{x}}$ attains a negative value, so that the trait variance ν vanishes, and the system loses its adaptability, since $\bar{x}$ can no longer change. This demonstrates that the equations of a *NA* model inherently promote the loss of functional diversity, also under scenarios of disruptive selection (Fig. E1A, D, G), so that it has to be maintained by other ecological processes such as immigration, mutation or more complex trade-off structures (Tirok et al., 2011; Merico et al., 2014; Coutinho et al., 2016). In contrast, as shown below, the relationships between lower and higher central moments entailed in the beta distribution-based moment closure equations offer an alternative explanation and mechanism for the maintenance of functional diversity and thus for keeping ν above zero.

In order to achieve a basic understanding of how the different functional terms of Eqs. (E1) and (E2) may interact to shape the overall temporal dynamics of $\bar{x}$ and ν, we first analyze the local stability of a potential fixed point $\bar{x}^*$ of Eq. (E1), i.e. we consider a situation where $d\bar{x}/dt|_{\bar{x}=\bar{x}^*} = 0$. For simplicity, we assume that the skewness depends only on the mean trait value, $S(\bar{x})$, while the kurtosis depends only on the trait variance, $K(\nu)$. This seems to be reasonable given that $\bar{x}$ explains much of the variation in S and ν explains much of the variation in K both for natural and simulated trait distributions (Fig. 2). We also assume that the trait distribution is symmetric at the equilibrium, i.e. $S(\bar{x}^*) = 0$, which seems to be valid under rather balanced selection regimes where the fitness landscape is also symmetric around $\bar{x}^*$ (Fig. E1). Under these simplifying conditions, the full

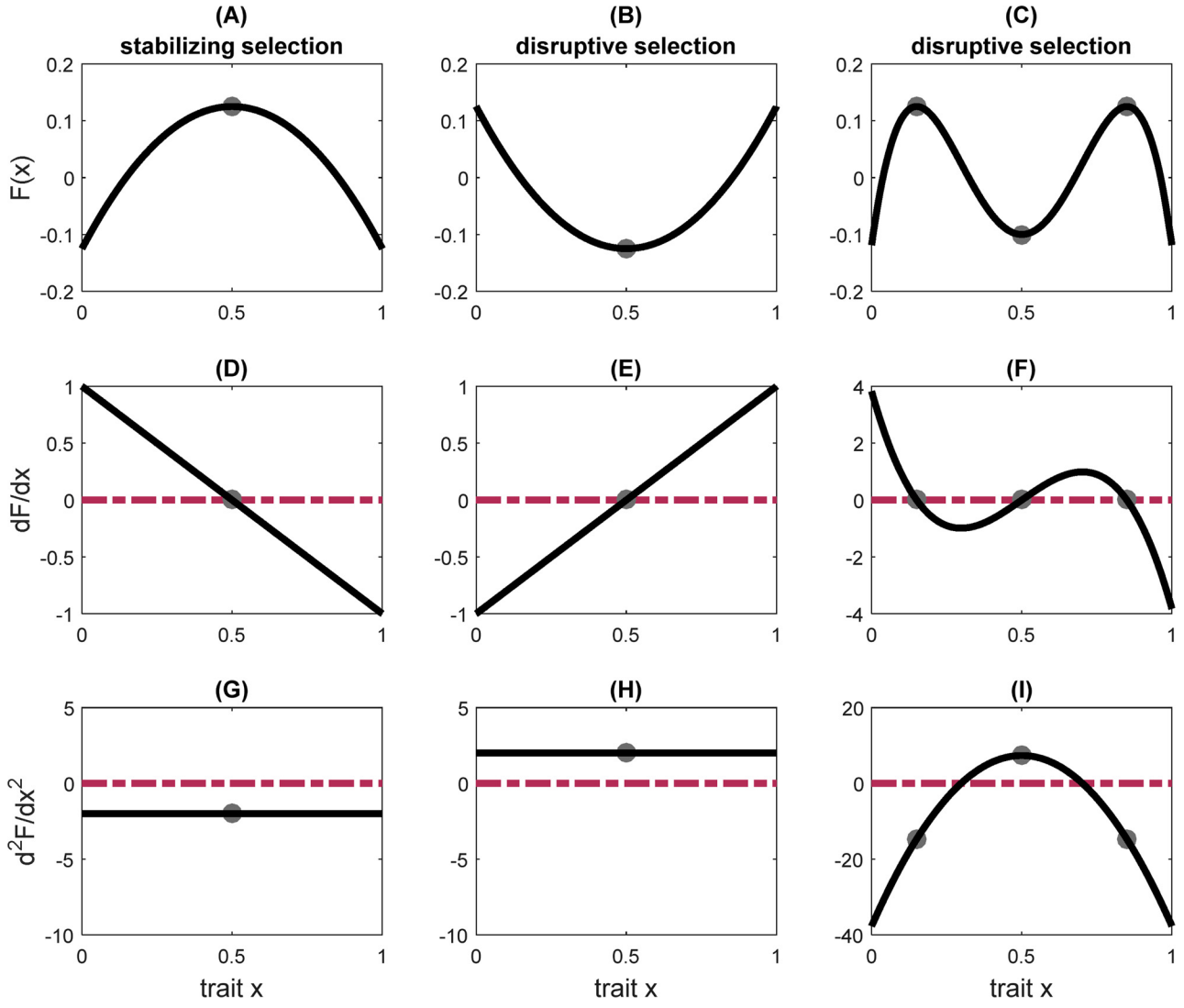

Fig. E1. Sketch of three different fitness landscapes (A–C) imposing stabilizing (A) or disruptive selection (B and C) and their corresponding first (D–F) and second (G–I) derivatives. The red dotted-dashed lines separate positive and negative values from each other. The grey dots indicate potential fixed points of the mean trait value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Jacobian of the two-dimensional system of Eqs. (E1) and (E2) at the fixed point with positive ν^* becomes diagonal, allowing us to perform the analysis of each equation separately. Hence, in this case, a potential fixed point $\overline{x}^*$ of Eq. (E1) satisfies the following condition: $\partial F/\partial x|_{\overline{x}=\overline{x}^*} = 0$. Therefore, $\overline{x}^*$ is situated either at a local fitness maximum or at a local fitness minimum of the corresponding fitness landscape, considering that $\partial^2 F/\partial x^2|_{x=\overline{x}^*} \neq 0$. We disregard the possibility of an inflection point, which would require very special assumptions about the fitness function F. A negative value of the derivative of Eq. (E1) with respect to $\overline{x}$ evaluated at $\overline{x}^*$ implies its local stability:

$$\frac{d(d\overline{x}/dt)}{d\overline{x}}\bigg|_{\overline{x}=\overline{x}^*} = \nu\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}^*} + \frac{\partial S}{\partial \overline{x}}\bigg|_{\overline{x}=\overline{x}^*}\frac{\nu^{\frac{3}{2}}}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}^*} + S\frac{\nu^{\frac{3}{2}}}{2}\frac{\partial^3 F(x)}{\partial x^3}\bigg|_{x=\overline{x}^*} \tag{E5}$$

Since the trait distribution is assumed to be symmetric at the equilibrium, i.e. $S(\overline{x}^*) = 0$, this simplifies to:

$$\frac{d(d\overline{x}/dt)}{d\overline{x}}\bigg|_{\overline{x}=\overline{x}^*} = \frac{\nu}{2}\left(2 + \frac{\partial S}{\partial \overline{x}}\bigg|_{\overline{x}=\overline{x}^*}\nu^{\frac{1}{2}}\right)\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}^*} \tag{E6}$$

In line with the general properties of Eq. (18) and the observed negative correlation between S and $\overline{x}$ of the phytoplankton size distributions (Fig. 2), we expect that $\partial S/\partial \overline{x}|_{\overline{x}=\overline{x}^*} < 0$. Given that $\partial^2 F/\partial x^2|_{x=\overline{x}^*}$ is negative at a fitness maximum (cf. Fig. E1), $\overline{x}^*$ will be stable when $|\partial S/\partial \overline{x}|_{\overline{x}=\overline{x}^*}\cdot\nu^{\frac{1}{2}}| < 2$ and unstable otherwise. In contrast, $\partial^2 F/\partial x^2|_{x=\overline{x}^*}$ is positive at a fitness minimum (cf. Fig. E1) so $\overline{x}^*$ will be stable if $2 < |\partial S/\partial \overline{x}|_{\overline{x}=\overline{x}^*}\cdot\nu^{\frac{1}{2}}|$ and unstable otherwise. Remarkably, this means that if a small perturbation of $\overline{x}$ from its equilibrium $\overline{x}^*$ leads to a substantial increase of S of the trait distribution, and ν is large enough, a fixed point at a fitness minimum will be stable. The local fitness landscape around the fixed point $\overline{x}^*$ can be approximated by a parabolic function, so a perturbation of $\overline{x}$ away from $\overline{x}^*$ leads to asymmetric growth, skewing the trait distribution. Moreover, since the fitness function decreases towards $\overline{x}^*$ (Fig. E1E), $S(\overline{x})$ will have the opposite sign of $\overline{x}-\overline{x}^*$, so $\partial S/\partial \overline{x}|_{x=\overline{x}^*}$ will be negative. This change in S increases with the degree of curvature of the local fitness landscape $\partial^2 F/\partial x^2|_{x=\overline{x}}$. Hence, a fixed point at a fitness minimum is likely to be stable when the local fitness landscape exhibits a pronounced valley (cf. Fig. E1). This contrasts with the results of a *NA* model (cf. Eqs.

(E3) and (E4)), where the fixed point $\overline{x}^*$ of $d\overline{x}/dt$ at a fitness maximum is always stable and at a fitness minimum always unstable.

With this insight at hand, we will now consider Eq. (E2) describing the temporal development of the trait variance ν, i.e. functional diversity, in response to selection. Since the mean trait value $\overline{x} = \overline{x}^*$ is at a fixed point at a fitness minimum or maximum, the local fitness gradient $\partial F(x)/\partial x|_{x=\overline{x}^*}$ equals zero. Hence, Eq. (E2) simplifies to:

$$\frac{d\nu}{dt} = \frac{(K\nu^2-\nu^2)}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}} \tag{E7}$$

This equation has (at most) two fixed points (once again disregarding an inflection point of the fitness function), an ecologically trivial one, $\nu^*=0$, and a nontrivial one that is implicitly defined by the equation $K(\nu^*) = 1$. The local stability of the latter can be determined from the sign of the following equation:

$$\frac{d\left(\frac{d\nu}{dt}\right)}{d\nu}\bigg|_{\nu=\nu^*} = \frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}}\left(\frac{\partial K}{\partial \nu}\frac{\nu^2}{2} + \frac{K2\nu}{2} - \frac{2\nu}{2}\right)\bigg|_{\nu=\nu^*} = \frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}}\left(\frac{\partial K}{\partial \nu}\frac{\nu^2}{2} + \nu(K-1)\right)\bigg|_{\nu=\nu^*} \tag{E8}$$

Hence, setting $K(\nu^*) = 1$ into Eq. (E8) leads to:

$$\frac{d\left(\frac{d\nu}{dt}\right)}{d\nu}\bigg|_{\nu=\nu^*} = \frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}}\left(\frac{\partial K}{\partial \nu}\frac{\nu^2}{2}\right)\bigg|_{\nu=\nu^*} \tag{E9}$$

In line with the general properties of Eq. (19) and the observed negative correlation between K and ν of the phytoplankton size distributions (Fig. 2), we expect that $\partial K/\partial \nu|_{\nu=\nu^*} < 0$. Hence, the ecologically non-trivial fixed point ν^* is stable when the fixed point $\overline{x}^*$ is at a fitness minimum and unstable when the fixed point $\overline{x}^*$ is at a fitness maximum.

Taking the two results of the preceding linear stability analyses of $\overline{x}$ and ν together, we found that the negative correlations between lower and higher central moments may jointly promote the maintenance of functional diversity, i.e. $\nu^* > 0$, by keeping $\overline{x}$ at a fitness minimum, where $\partial^2 F/\partial x^2|_{x=\overline{x}} > 0$. Again, this is in direct contradiction to the conclusions of a *NA* model, where the stability condition of a fixed point $\overline{x}^*$ of $d\overline{x}/dt$ at a fitness maximum (cf. Eqs. (E3) and (E4)) implies ν equals zero, losing all functional diversity. In addition, if the initial value of $\overline{x}$ is a fixed point $\overline{x}^*$ of $d\overline{x}/dt$ at a fitness minimum of the *NA* model, ν will increase without bound instead of vanishing, provided there is no feedback between the total biomass (influenced by $\frac{\nu}{2}\cdot\partial^2 F/\partial x^2|_{x=\overline{x}}$) and $F(\overline{x})$ itself. In that case, we would find ourselves with the opposite issue of exploding trait variances. This feature is known as runaway evolution of traits (cf. Abrams and Matsuda, 1997). Furthermore, this reasoning implies that any moment closure approach for trait-based aggregate models defined by Eqs. (E1) and (E2) based on the assumption of a trait distribution that exhibits constant values of S and K independent of $\overline{x}$ and ν will not be able to overcome the same pitfall of *NA* models: loss of functional diversity over time without frequency-dependent selection. However, as discussed below in more detail, frequency-dependent selection may also stabilize a fitness minimum within a *NA* model (Abrams et al., 1993b).

In line with the observed correlations between lower and higher central moments (Fig. 2), one of the simplest moment closures that can be conducted for an aggregate model defined by Eqs. (E1) and (E2), while adhering to the stability conditions outlined above, is given by assuming S and $ln(K)$ to be expressed as linear functions of $\overline{x}$ and $ln(\nu)$, respectively:

$$S = M_3\nu^{-\frac{3}{2}} = b_S\overline{x} + a_S \;\therefore\; M_3 = \nu^{\frac{3}{2}}(b_S\overline{x} + a_S) \tag{E10}$$

$$K = M_4\nu^{-2} = a_K\nu^{b_K} \;\therefore\; \ln(K) = b_K\ln(\nu) + \ln(a_K) \;\therefore\; M_4 = a_K\nu^{b_K+2} \tag{E11}$$

with a_K, $a_S > 0$ and b_K, $b_S < 0$ (cf. Gaedke and Klauschies, 2017). This enables a data-based moment closure of trait-based aggregate models that are defined by Eqs. (E1) and (E2), where the intercept (a) and slope (b) parameters of Eqs (E10) and (E11) are estimated from actual empirical data (see Fig. 2).

Finally, we support our analytical results by showing numerical simulations of an extended *BA* model that does not need immigration or any other mechanism to maintain ν, i.e. $\nu^* > 0$. When considering the central moments up to the fourth rather than the second order, the aggregate model with the highest accuracy with respect to the resolution of the entire fitness landscape is given by the following set of equations (cf. Appendix A):

$$\frac{dR_T}{dt} = F(\overline{x}) + \frac{\nu}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{S\nu^{\frac{3}{2}}}{6}\frac{\partial^3 F(x)}{\partial x^3}\bigg|_{x=\overline{x}} + \frac{K\nu^2}{24}\frac{\partial^4 F(x)}{\partial x^4}\bigg|_{x=\overline{x}} + I \tag{E12}$$

$$\frac{d\overline{x}}{dt} = \nu\frac{\partial F(x)}{\partial x}\bigg|_{x=\overline{x}} + \frac{S\nu^{\frac{3}{2}}}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{K\nu^2}{6}\frac{\partial^3 F(x)}{\partial x^3}\bigg|_{x=\overline{x}} + \frac{I}{R_T}\left(\frac{1}{2}-\overline{x}\right) \tag{E13}$$

$$\frac{d\nu}{dt} = S\nu^{\frac{3}{2}}\frac{\partial F(x)}{\partial x}\bigg|_{x=\overline{x}} + \frac{(K\nu^2-\nu^2)}{2}\frac{\partial^2 F(x)}{\partial x^2}\bigg|_{x=\overline{x}} + \frac{I}{R_T}\left(\frac{1}{12}-\nu + \left(\frac{1}{2}-\overline{x}\right)^2\right) \tag{E14}$$

In line with Eq. (20), the fourth-order approximation of the per-capita net-growth rate F is given by:

$$F(x) = \left(r\left(1-\frac{R_T}{\kappa}\right) - \sum_{k=1}^{3}\frac{a_k(x)C_k}{1 + h\left(a_k(\overline{x}) + \sum_{i=2}^{4}\frac{M_i}{i!}\frac{\partial^i a_k(x)}{\partial x^i}\bigg|_{x=\overline{x}}\right)R_T}\right) \tag{E15}$$

In accordance with the simulations presented in the results section, we run numerical simulations of this extended *BA* model by keeping all parameters the same except for the immigration rate I, which was set to zero. Depending on the initial conditions, our model predicts either oscillations or stasis in the predator and prey biomasses (Fig. E2). In the first case, the mean of the trait distribution approached a stable equilibrium

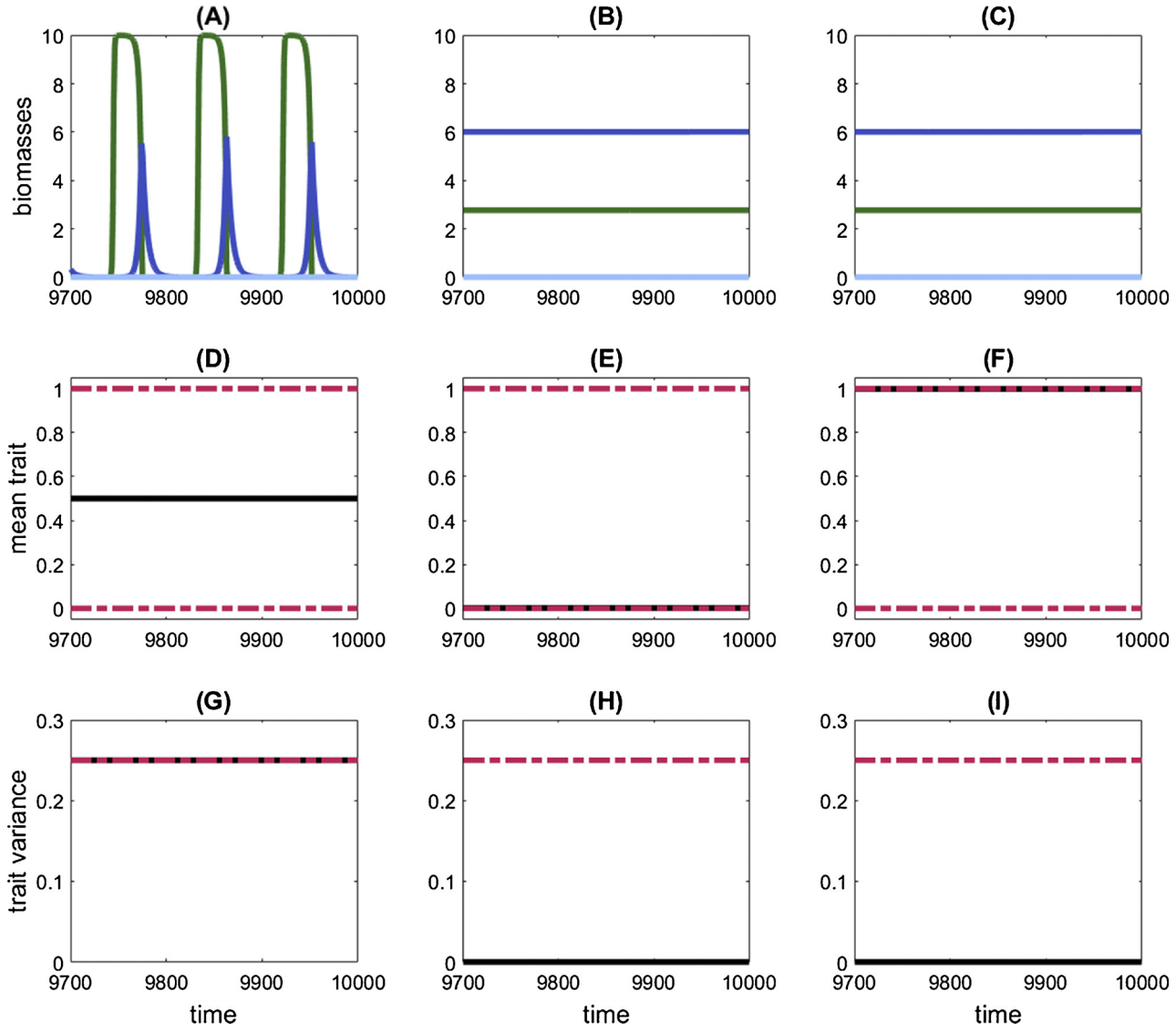

Fig. E2. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (green) and the 3 predator species that are specialized on low (black), intermediate (dark blue) and high (light blue) trait values of the prey community under disruptive selection for the extended beta distribution-based aggregate model without immigration, i.e. $I = 0$. The dashed-dotted red lines in (D)–(I) indicate the ecologically feasible minima and maxima of the mean and variance. Initial values of the mean trait and trait variance of the prey community were set to 0.45 and 0.01 for (A, D, G), 0.45 and 0.001 for (B, E, H) and 0.55 and 0.001 for (C, F, I). Note a value of $\nu \approx 0$ implies that the prey community is hardly adaptive anymore.

at an intermediate value that corresponded to a fitness minimum, whereas the value of the trait variance was at its maximum, indicating a bimodal distribution. In the second case, on the other hand, the mean trait value approached one of the two extremes that corresponded to a fitness maximum while the trait variance collapsed towards zero. These results illustrate how the beta distribution approach can keep high levels of standing trait variation in trait-based aggregate models without additional mechanisms. When adding immigration to the extended version of the *BA* model, the resulting biomass and trait dynamics no longer depended on the initial conditions and showed very high levels of standing trait variation (Fig. E3). In contrast to the results shown in Fig. E2, the value of ν was not constant and changed over time. This pattern arose from an interplay between the predator-prey oscillations and the density-dependent impact of immigration. Interestingly, the immigration term in $d\overline{x}/dt$ ensured global stability of the fixed point $\overline{x}^* = 0.5$ at a fitness minimum and thus overall maintenance of ν. However, at the same time, the immigration term in $d\nu/dt$ promoted temporal reductions of ν because the trait variance of the immigrating community is smaller than the trait variance of the resident community at equilibrium. Despite the good match between the trait dynamics, this model overestimated the grazing rate of the predators on the prey community, which promoted ongoing predator-prey cycles instead of stasis (see Fig. 6). Furthermore, this model was unable to reproduce the biomass and trait dynamics of the prey and predator community under fluctuating selection as displayed in Fig. 4. This indicates that trait-based aggregate models are quite sensitive to the choice of the moment closure approach and mechanism used to maintain ν. Hence, besides showing more thoroughly the actual potential of a beta distribution-based moment closure approach for trait-based aggregate models, these results also point towards potential limitations that may arise when accounting for higher derivatives of the fitness function, i.e. higher order Taylor expansions.

E.2. Convergent stable fitness minima under frequency-dependent selection

The preceding discussion assumed that natural selection is independent of $\overline{x}$ and ν of the trait distribution. Hence, we neglected the possibility for

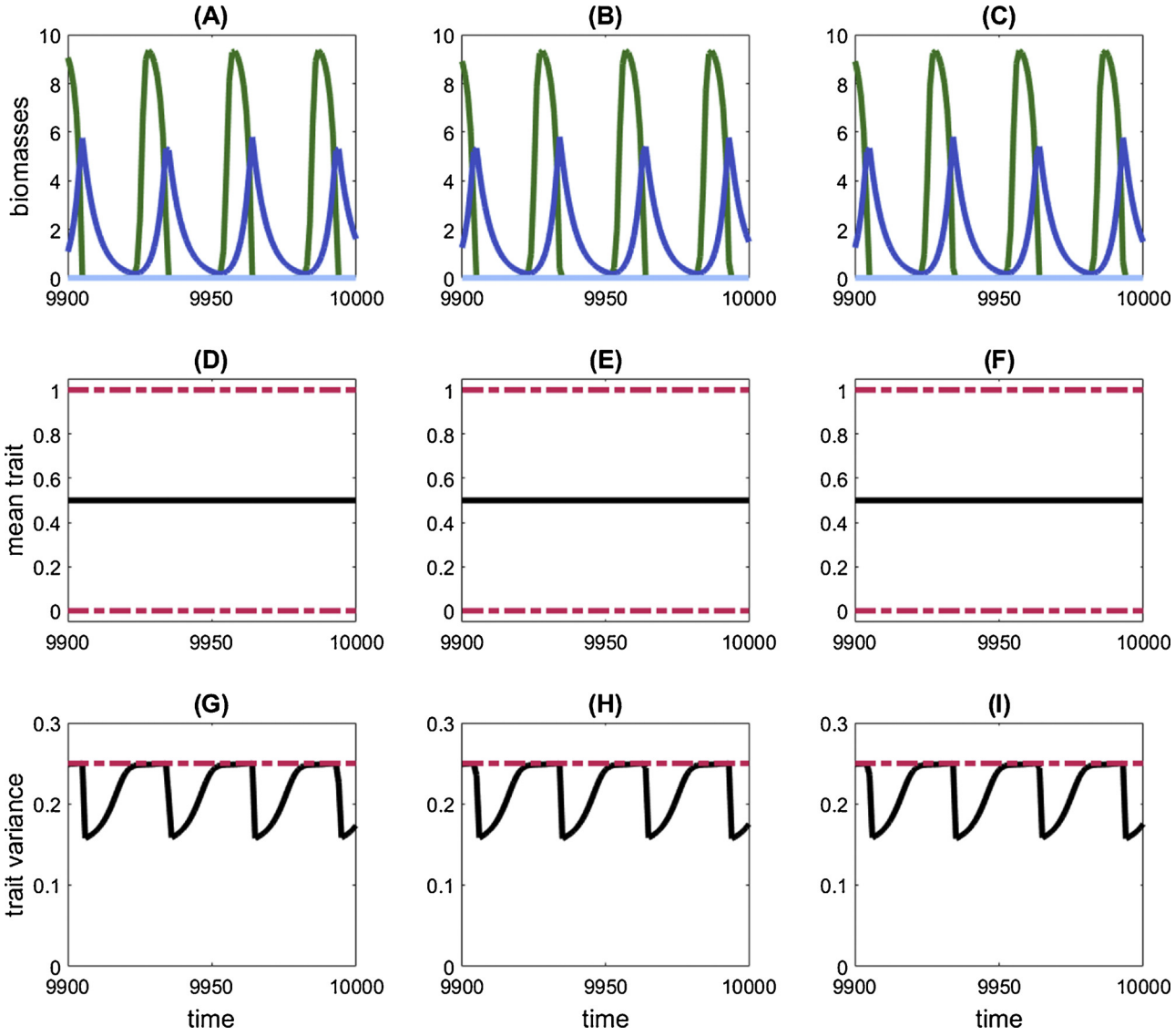

Fig. E3. Biomass (A–C), mean trait (D–F) and trait variance (G–I) dynamics of the prey community (green) and the 3 predator species that are specialized on low (black), intermediate (dark blue) and high (light blue) trait values of the prey community under disruptive selection for the extended beta distribution-based aggregate model with immigration, i.e. $I = 0.0001$. The dashed-dotted red lines in (D)–(I) indicate the ecologically feasible minima and maxima of the mean and variance. Initial values of the mean trait and trait variance of the prey community were set to 0.45 and 0.01 for (A, D, G), 0.45 and 0.001 for (B, E, H) and 0.55 and 0.001 for (C, F, I).

a convergent stable fitness minimum that may generally arise under frequency-dependent selection, in which the fitness and thus, the invasion success of a mutant strongly depend on the phenotype of the (monomorphic) resident community (Christiansen, 1991; Abrams et al., 1993b; Rueffler et al., 2006). In general, this frequency-dependent mechanism can stabilize a fitness minimum also in a *NA* model, which assumes that the trait values are normally distributed, and exhibit a small variance. However, we are going to show that such stable fitness minima do not exist in our *NA* model with a single predator. Hence, this mechanism cannot have stabilized the fitness minimum in our *FD* model under disruptive selection (Fig. 6). Instead, the fitness minimum in our *FD* model was likely stable due to the mechanism discussed in the previous section.

When selection is frequency-dependent, the species' fitness function F depends not only on the total biomass R_T and its individual trait value x but also on the mean $\bar{x}$ and variance ν of the entire trait distribution (cf. Eq. (20)). Hence, when considering a *NA* model under frequency-dependent selection, the linear stability condition for $\bar{x}$ at its equilibrium $\bar{x}^*$ is given more generally by the sign of the following equation:

$$\left.\frac{d(d\bar{x}/dt)}{d\bar{x}}\right|_{\bar{x}=\bar{x}^*} = \nu\left(\left.\frac{\partial^2 F(x,\bar{x},\nu)}{\partial x^2}\right|_{x=\bar{x}=\bar{x}^*} + \left.\frac{\partial^2 F(x,\bar{x},\nu)}{\partial x \partial \bar{x}}\right|_{x=\bar{x}=\bar{x}^*}\right) \tag{E16}$$

Hence, the fixed point $\bar{x}^*$ is locally stable if $\partial^2 F/\partial x^2|_{x=\bar{x}=\bar{x}^*} + \partial^2 F/\partial x \partial \bar{x}|_{x=\bar{x}=\bar{x}^*} < 0$ and $\bar{x}^*$ is unstable otherwise (Abrams et al., 1993b). In contrast to frequency-independent selection, this can also happen under frequency-dependent selection when $\bar{x}^*$ is located at a fitness minimum. In such a case, selection will drive $\bar{x}$ first towards a local fitness maximum by following the fitness gradient, which then turns into a local fitness minimum when $\bar{x}$ reaches the corresponding extreme trait value of the fitness landscape, at which point selection becomes disruptive (Rueffler et al., 2006). As a potential result, ν may exhibit stable growth in the absence of any other negative feedback. However, frequency-dependent selection does not ensure that $\bar{x}$ at a fitness minimum becomes convergent stable (Abrams et al., 1993b). In fact, the likelihood of fitness maxima to be unstable seems

to be much higher than the likelihood of fitness minima to be stable (Abrams et al., 1993b). Furthermore, in line with the stability conditions for $\bar{x}^*$ under frequency-independent selection, $\partial^2 F/\partial x \partial \bar{x}|_{x=\bar{x}=\bar{x}^*}$ equals zero also for some cases of frequency-dependent selection. For example, consider the predator-prey model used in the main text (Eqs. (1)–(2)) with only a single predator in place (cf. Fig. 6). In this case, we can express the first derivative of the fitness function of the prey with respect to x (Eq. (20)) as the product of a function $f_1(x)$ that only depends on x and a function $f_2(\bar{x}, \nu)$ that only depends on $\bar{x}$ and ν of the trait distribution, i.e. $\partial F(x, \bar{x}, \nu)/\partial x = f_1(x) \cdot f_2(\bar{x}, \nu)$. Given that the denominator of the functional response is positive, i.e. $f_2(\bar{x}, \nu) > 0$ for all $\bar{x}$ and sufficiently small ν, the second derivative $\partial^2 F/\partial x \partial \bar{x}|_{x=\bar{x}}$ equals zero when $\partial F/\partial x|_{x=\bar{x}} = f_1(x)|_{x=\bar{x}} = 0$ and, thus, when $\bar{x}$ is located at a fitness maximum or minimum. For the same reason, $\partial^2 F/\partial x \partial \nu|_{x=\bar{x}} = 0$ at such an equilibrium, and so the off-diagonal terms of the Jacobian matrix of the system given by Eqs. (E3)–(E4) cancel, allowing us to analyze each equation separately once again. Hence, under such circumstances, a convergent stable fitness minimum does not exist in our *NA* model even with frequency-dependent selection.

In contrast to frequency-independent selection, a potential equilibrium of ν in the *NA* model under frequency-dependent selection is given either by $\nu^*=0$ or by $\frac{\partial^2 F(x,\bar{x},\nu)}{\partial x^2}\Big|_{x=\bar{x},\nu=\nu^*} = 0$. The first case is the trivial one noted beforehand. However, also the realism of the second, non-trivial equilibrium of ν has to be questioned because it implies that $\bar{x}^*$ has to be located at an inflection point. Furthermore, the stability of the non-trivial ν^* is given by the sign of the following equation:

$$\frac{d\left(\frac{d\nu}{dt}\right)}{d\nu}\Bigg|_{\nu=\nu^*} = 2\nu^* \frac{\partial^2 F(x,\bar{x},\nu)}{\partial x^2}\Bigg|_{x=\bar{x},\nu=\nu^*} + \nu^{*2} \frac{\partial^3 F(x,\bar{x},\nu)}{\partial x^2 \partial \nu}\Bigg|_{x=\bar{x},\nu=\nu^*} = \nu^{*2} \frac{\partial^3 F(x,\bar{x},\nu)}{\partial x^2 \partial \nu}\Bigg|_{x=\bar{x},\nu=\nu^*} \tag{E17}$$

Since $\nu^* \geq 0$, the stability depends only on the sign of the third derivative. Although it is not immediately clear which ecological scenarios may correspond to this stability condition, the third derivative in Eq. (E17) is zero when considering once again our predator-prey model with a single predator in place, following the reasoning outlined above for Eq. (E16). This suggests that frequency-dependent selection is not able to promote non-zero variances at equilibrium in *NA*-models. However, in general, when considering *BA* models, the stability condition based on the correlations between lower and higher central moments (Eqs. (E6) and (E9)) may interact in several ways with the stability condition based on frequency-dependence (Eqs. (E16) and (E17)), jointly determining the overall stability of fitness minima and maxima.

References

Abrams, P.A., Harada, Y., Matsuda, H., 1993a. On the relationship between quantitative genetic and ESS models. Evolution 47, 982–985.

Abrams, P.A., Matsuda, H., Harada, Y., 1993b. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. Evol. Ecol. 7, 465–487.

Abrams, P.A., Matsuda, H., 1997. Fitness minimization and dynamic instability as a consequence of predator-prey coevolution. Evol. Ecol. 11, 1–20.

Abrams, P.A., 1999. The adaptive dynamics of consumer choice. Am. Nat. 153, 83–97.

Abrams, P.A., 2006. The prerequisites for and likelihood of generalist-specialist coexistence. Am. Nat. 167, 329–342. http://dx.doi.org/10.1086/303154.

Abrams, P.A., 2010. Quantitative descriptions of resource choice in ecological models. Popul. Ecol. 52, 47–58. http://dx.doi.org/10.1007/s10144-009-0175-z.

Acevedo-Trejos, E., Brandt, G., Bruggeman, J., Merico, A., 2015. Mechanisms shaping size structure and functional diversity of phytoplankton communities in the ocean. Sci. Rep. 5, 8918. http://dx.doi.org/10.1038/srep08918.

Acevedo-Trejos, E., Brandt, G., Smith, S.L., Merico, A., 2016. PhytoSFDM version 1.0.0: phytoplankton size and functional diversity model. Geosci. Model. Dev. 9, 4071–4085. http://dx.doi.org/10.5194/gmd-9-4071-2016.

Barton, N.H., Turelli, M., 1989. Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23, 337–370.

Bolker, B., Pacala, S.W., 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. Theor. Popul. Biol. 52, 179–197. http://dx.doi.org/10.1006/tpbi.1997.1331.

Bruggeman, J., Kooijman, S.A.L.M., 2007. A biodiversity-inspired approach to aquatic ecosystem modeling. Limnol. Oceanogr. 52, 1533–1544. http://dx.doi.org/10.4319/lo.2007.52.4.1533.

Christiansen, F.B., 1991. On conditions for evolutionary stability for a continuous varying character. Am. Nat. 138, 37–50.

Clancy, D., Mendy, S.T., 2011. Approximating the quasi-stationary distribution of the SIS model for endemic infection. Methodol. Comput. Appl. Probab. 13, 603–618. http://dx.doi.org/10.1007/s11009-010-9177-8.

Cortez, M.H., 2016. How the magnitude of prey genetic variation alters predator-prey eco-evolutionary dynamics. Am. Nat. 188, 329–341. http://dx.doi.org/10.1086/687393.

Cortez, M.H., Patel, S., 2017. The effects of predator evolution and genetic variation on predator—prey population-level dynamics. Bull. Math. Biol. 79, 1510–1538. http://dx.doi.org/10.1007/s11538-017-0297-y.

Coutinho, R.M., Klauschies, T., Gaedke, U., 2016. Bimodal trait distributions with large variances question the reliability of trait-based aggregate models. Theor. Ecol. 9, 389–408. http://dx.doi.org/10.1007/s12080-016-0297-9.

Day, T., 2005. Modelling the ecological context of evolutionary change: déjà vu or something new? In: Cuddington, K., Beisner, B.E. (Eds.), Ecological Paradigms Lost: Routes to Theory Change. Academic Press, New York, pp. 273–309.

Day, T., Proulx, S.R., 2004. A general theory for the evolutionary dynamics of virulence. Am. Nat. 163, E40–E63. http://dx.doi.org/10.1086/382548.

Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol. 34, 579–612.

Doebeli, M., Blok, H.J., Leimar, O., Dieckmann, U., 2007. Multimodal pattern formation in phenotype distributions of sexual populations. Proc. R. Soc. Lond. B Biol. Sci. 274, 347–357. http://dx.doi.org/10.1098/rspb.2006.3725.

Downing, A.S., Hajdu, S., Hjerne, O., Otto, S.A., Blenckner, T., Larsson, U., Winder, M., 2014. Zooming in on size distribution patterns underlying species coexistence in Baltic Sea phytoplankton. Ecol. Lett. 17, 1219–1227. http://dx.doi.org/10.1111/ele.12327.

Fox, J.W., Nelson, W.A., McCauley, E., 2010. Coexistence mechanisms and the paradox of the plankton: quantifying selection from noisy data. Ecology 91, 1774–1786. http://dx.doi.org/10.1890/09-0951.1.

Gaedke, U., 1992. The size distribution of plankton biomass in a large lake and its seasonal variability. Limnol. Oceanogr. 37, 1202–1220. http://dx.doi.org/10.4319/lo.1992.37.6.1202.

Gaedke, U., Klauschies, T., 2017. Analysing the shape of observed trait distributions enables a data-based moment closure of aggregate models. Limnol. Oceanogr. Methods 15, 979–994. http://dx.doi.org/10.1002/lom3.10218.

Geritz, S.A.H., Kisdi, E., Meszena, G., Metz, J.A.J., 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12, 35–57.

Gupta, A.K., Nadarajah, S., 2004. Mathematical properties of the beta distribution. In: Gupta, A.K., Nadarajah, S. (Eds.), Handbook of Beta Distribution and Its Applications. CRC Press, pp. 33–54.

Havlicek, T.D., Carpenter, S.R., 2001. Pelagic species size distributions in lakes: are they discontinuous? Limnol. Oceanogr. 46, 1021–1033. http://dx.doi.org/10.4319/lo.2001.46.5.1021.

Iwasa, Y., Andreasen, V., Levin, S., 1987. Aggregation in model ecosystems. I. Perfect aggregation. Ecol. Model. 37, 287–302.

Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. Biometrika 36, 149–176.

Kirkpatrick, M., Rousset, F., 2005. Wright meets AD: not all landscapes are adaptive. J. Evol. Biol. 18, 1166–1169. http://dx.doi.org/10.1111/j.1420-9101.2004.00847.x.

Klais, R., Norros, V., Lehtinen, S., Tamminen, T., Olli, K., 2016a. Community assembly and drivers of phytoplankton functional structure. Funct. Ecol. 31, 760–767. http://dx.doi.org/10.1111/1365-2435.12784.

Klais, R., Norros, V., Lehtinen, S., Tamminen, T., Olli, K., 2016b. Data from: community assembly and drivers of phytoplankton functional structure. Dryad Digital Repository doi: 10.5061/dryad.p811t.

Klauschies, T., Vasseur, D., Gaedke, U., 2016. Trait adaptation promotes species coexistence in diverse predator and prey communities. Ecol. Evol. 6, 4141–4159. http://dx.doi.org/10.1002/ece3.2172.

Kremer, C.T., Klausmeier, C.A., 2017. Species packing in eco-evolutionary models of seasonally fluctuating environments. Ecol. Lett. 20, 1158–1168. http://dx.doi.org/10.1111/ele.12813.

Krishnarajah, I., Cook, A., Marion, G., Gibson, G., 2005. Novel moment closure approximations in stochastic epidemics. Bull. Math. Biol. 67, 855–873. http://dx.doi.org/10.1016/j.bulm.2004.11.002.

Kuehn, C., 2015. Moment closure—a brief review. In: Schöll, E., Klapp, S.H.L., Hövel, P. (Eds.), Control of Self-Organizing Nonlinear Systems. Springer International Publishing, pp. 253–271. http://dx.doi.org/10.1007/978-3-319-28028-8_13.

Lande, R., 1976. Natural-selection and random genetic drift in phenotypic evolution. Evolution 30, 314–334.

Lande, R., 1982. A quantitative genetic theory of life-history evolution. Ecology 63, 607–615.

Lion, S., 2018. Theoretical approaches in evolutionary ecology: environmental feedback as a unifying perspective. Am. Nat. 191, 21–44. http://dx.doi.org/10.1086/694865.

Macarthur, R., Levins, R., 1967. The limiting similarity, convergence, and divergence of coexisting species. Am. Nat. 101, 377–385.
Maurer, B.A., Brown, J.H., Dayan, T., Enquist, B.J., Ernest, S.M., Hadly, E.A., Haskell, J.P., Jablonski, D., Jones, K.E., Kaufman, D.M., et al., 2004. Similarities in body size distributions of small-bodied flying vertebrates. Evol. Ecol. Res. 6, 783–797.
McGill, B.J., Brown, J.S., 2007. Evolutionary game theory and adaptive dynamics of continuous traits. Annu. Rev. Ecol. Evol. Syst. 38, 403–435. http://dx.doi.org/10.1146/annurev.ecolsys.36.091704.175517.
Merico, A., Bruggeman, J., Wirtz, K., 2009. A trait-based approach for downscaling complexity in plankton ecosystem models. Ecol. Model. 220, 3001–3010. http://dx.doi.org/10.1016/j.ecolmodel.2009.05.005.
Merico, A., Brandt, G., Smith, S.L., Oliver, M., 2014. Sustaining diversity in trait-based models of phytoplankton communities. Front. Ecol. Evol. 2, 1–8. http://dx.doi.org/10.3389/fevo.2014.00059.
Nelson, W.A., McCauley, E., Wrona, F.J., 2005. Stage-structured cycles promote genetic diversity in a predator–prey system of Daphnia and algae. Nature 433, 413–417. http://dx.doi.org/10.1038/nature03212.
Norberg, J., Swaney, D.P., Dushoff, J., Lin, J., Casagrandi, R., Levin, S.A., 2001. Phenotypic diversity and ecosystem functioning in changing environments: a theoretical framework. Proc. Natl. Acad. Sci. U. S. A. 98, 11376–11381. http://dx.doi.org/10.1073/pnas.171315998.
Norberg, J., 2004. Biodiversity and ecosystem functioning: a complex adaptive systems approach. Limnol. Oceanogr. 49, 1269–1277. http://dx.doi.org/10.4319/lo.2004.49.4_part_2.1269.
Norberg, J., Urban, M.C., Vellend, M., Klausmeier, C.A., Loeuille, N., 2012. Eco-evolutionary responses of biodiversity to climate change. Nat. Clim. Change 2, 747–751. http://dx.doi.org/10.1038/NCLIMATE1588.
Pahlow, M., Prowe, A.E.F., 2010. Model of optimal current feeding in zooplankton. Mar. Ecol. Prog. Ser. 403, 129–144. http://dx.doi.org/10.3354/meps08466.
Parker, G.A., Maynard Smith, J., 1990. Optimality theory in evolutionary biology. Nature 348, 27–33.
Rueffler, C., Van Dooren, T.J.M., Leimar, O., Abrams, P.A., 2006. Disruptive selection and then what? Trends Ecol. Evol. 21, 238–245. http://dx.doi.org/10.1016/j.tree.2006.03.003.
Sasaki, A., Dieckmann, U., 2011. Oligomorphic dynamics for analyzing the quantitative genetics of adaptive speciation. J. Math. Biol. 63, 601–635. http://dx.doi.org/10.1007/s00285-010-0380-6.
Savage, V.M., Webb, C.T., Norberg, J., 2007. A general multi-trait-based framework for studying the effects of biodiversity on ecosystem functioning. J. Theor. Biol. 247, 213–229. http://dx.doi.org/10.1016/j.jtbi.2007.03.007.
Schreiber, S.J., Burger, R., Bolnick, D.I., 2011. The community effects of phenotypic and genetic variation within a predator population. Ecology 92, 1582–1593. http://dx.doi.org/10.1890/10-2071.1.
Siepielski, A.M., DiBattista, J.D., Carlson, S.M., 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. Ecol. Lett. 12, 1261–1276. http://dx.doi.org/10.1111/j.1461-0248.2009.01381.x.
Slatkin, M., 1979. Frequency- and density-dependent selection on a quantitative character. Genetics 93, 755–771.
Smith, S.L., Yamanaka, Y., Pahlow, M., Oschlies, A., 2009. Optimal uptake kinetics: physiological acclimation explains the pattern of nitrate uptake by phytoplankton in the ocean. Mar. Ecol. Prog. Ser. 384, 1–12. http://dx.doi.org/10.3354/meps08022.
Smith, S.L., Vallina, S.M., Merico, A., 2016. Phytoplankton size-diversity mediates an emergent trade-off in ecosystem functioning for rare versus frequent disturbances. Sci. Rep. 6, 34170. http://dx.doi.org/10.1038/srep34170.
Taper, M.L., Case, T.J., 1985. Quantitative genetic models for the coevolution of character displacement. Ecology 66, 355–371.
Taper, M.L., Case, T.J., 1992. Models of character displacement and the theoretical robustness of taxon cycles. Evolution 46, 317–333.
Taylor, P.D., 1996. The selection differential in quantitative genetics and ESS models. Evolution 50, 2106–2110.
Taylor, P., Day, T., 1997. Evolutionary stability under the replicator and the gradient dynamics. Evol. Ecol. 11, 579–590.
Terseleer, N., Bruggeman, J., Lancelot, C., Gypens, N., 2014. Trait-based representation of diatom functional diversity in a plankton functional type model of the eutrophied southern North Sea. Limnol. Oceanogr. 59, 1958–1972. http://dx.doi.org/10.4319/lo.2014.59.6.1958.
Tackx, M.L.M., Herman, P.M.J., van Rijswijk, P., Vink, M., Bakker, C., 1994. Plankton size distributions and trophic relations before and after the construction of the storm-surge barrier in the oosterschelde estuary. Hydrobiologia 282 (283), 145–152.
Thibault, K.M., White, E.P., Hurlbert, A.H., Ernest, S., 2011. Multimodality in the individual size distributions of bird communities. Global Ecol. Biogeogr. 20, 145–153. http://dx.doi.org/10.1111/j.1466-8238.2010.00576.x.
Tirok, K., Bauer, B., Wirtz, K., Gaedke, U., 2011. Predator-prey dynamics driven by feedback between functionally diverse trophic levels. Plos One 6, e27357. http://dx.doi.org/10.1371/journal.pone.0027357.
Tirok, K., Gaedke, U., 2010. Internally driven alternation of functional traits in a multi-species predator-prey system. Ecology 91, 1748–1762. http://dx.doi.org/10.1890/09-1052.1.
Tirok, K., Gaedke, U., 2007. Regulation of planktonic ciliate dynamics and functional composition during spring in Lake Constance. Aquat. Microb. Ecol. 49, 87–100. http://dx.doi.org/10.3354/ame01127.
Turelli, M., Barton, N.H., 1994. Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? Genetics 138, 913–941.
Weithoff, G., 2003. The concepts of 'plant functional types' and 'functional diversity' in lake phytoplankton—a new understanding of phytoplankton ecology? Freshw. Biol. 48, 1669–1675. http://dx.doi.org/10.1046/j.1365-2427.2003.01116.x.
Wirtz, K.W., Eckhardt, B., 1996. Effective variables in ecosystem models with an application to phytoplankton succession. Ecol. Model. 92, 33–53.