Supplementary Materials for

Environment-dependent diversity-stability relationship in microbial communities

Zifeng Hu *et al.*

*Corresponding author: jilianghu@szbl.ac.cn

The PDF file includes:

Materials and Methods

Bacterial isolates, media and culturing conditions

Two libraries of bacterial isolates were used: a 25-isolate library for the temperature experiment and an independent 36-isolate library for the dilution-factor experiment. The isolates were derived from bacteria isolated from maize seeds and from soil collected in Shenzhen, China. The taxonomic composition and phylogenetic relationships of both libraries are shown in fig. S1. Isolates were purified by dilution plating on LB broth agar and picking single colonies, identified from 16S rRNA gene sequences, and stored as glycerol stocks (25% glycerol) at -80 °C.

Experimental communities were cultured in Base Medium (BM): 1 g L^{-1} yeast extract and 1 g L^{-1} soytone from Becton Dickinson, 10 mM sodium phosphate, 0.1 mM $CaCl_2$, 2 mM $MgCl_2$, 4 mg L^{-1} $NiSO_4$ and 50 mg L^{-1} $MnCl_2$, with the initial pH adjusted to 6.5. For the temperature experiment, BM was supplemented with 5 g L^{-1} glucose and 4 g L^{-1} urea. For the dilution-factor experiment, BM was supplemented with 10 g L^{-1} glucose and 8 g L^{-1} urea. All media were filter sterilized using Bottle Top Filtration Units.

Both monocultures and communities were grown in sealed 96-deep-well plates. Temperature-experiment plates were incubated at 10, 20, 30, 40 or 50 °C and shaken at 600 r.p.m. Dilution-factor-experiment plates were incubated at 30 °C and shaken at 1,000 r.p.m.

Pre-cultures, community assembly, daily dilutions, dispersal, and biomass measurements

Before all community-assembly and monoculture experiments, each isolate was inoculated individually into BM and cultured for 24 hours. Aliquots from these 24-hour pre-cultures were then inoculated into the corresponding target medium. For the temperature experiment, 12 synthetic communities were assembled from defined subsets of the 25-isolate library. The pre-grown monocultures were mixed to establish equal initial representation of the community members, and each community-temperature combination was established in three biological replicates.

The temperature experiment was extended for 10 daily growth and transfer cycles. At each passage, the cultures were thoroughly mixed and 20 μL of community culture was transferred into 600 μL of fresh medium. A 6-μL dispersal inoculum was added to each well. The dispersal pool for each community was prepared only from the isolates belonging to that community's species pool and was diluted to give an effective dispersal rate of approximately 10^{-6}.

For the dilution-factor experiment, 12 synthetic communities were assembled from the 36-isolate library and propagated at 30 °C under daily dilution factors of 10^1, 10^2, 10^3, 10^4 or 10^5. Each community-dilution combination contained three biological replicates and was propagated for six 24-hour cycles. Each passage contained 540 μL of fresh medium, 60 μL of community inoculum prepared at the designated dilution factor, and 6 μL of community-specific dispersal inoculum at an effective rate of approximately 10^{-6}. Community and

dispersal inoculum were diluted in tenfold steps by transferring 20 µL into 180 µL of sterile phosphate-buffered saline (PBS) until the designated dilution was reached.

At the end of each daily cycle, cultures were homogenized and transferred to flat-bottom 96-well plates for optical-density measurements at 600 nm. Community OD was measured using 100 µl per well for the temperature experiment and 180 µl per well for the dilution-factor experiment; monoculture OD was measured using 150 µl per well. Sterile medium was used for background correction. The remaining culture was stored at -80 °C for subsequent DNA extraction where applicable.

Community pH was measured at the end of each growth cycle. Raw OD values were corrected by subtracting experiment-specific sterile-medium backgrounds (0.035 for the temperature experiment and 0.040 for the dilution-factor experiment), with negative corrected values set to zero. Species-level absolute abundance was reconstructed from the corrected OD and the sequencing-derived relative abundance:

$$N_i(t) = OD(t) \times q_i(t)$$

where $q_i(t)$ is the relative abundance of species i at passage t.

Monoculture physiology across temperatures

The 16 isolates with the highest abundance across the temperature experiment were selected for within-cycle physiological characterization at 20, 30 and 40 °C. For each isolate and temperature, a common inoculum was distributed among 12 parallel wells so that the culture volume remained identical at each sampling time. Each well contained 600 µl of fresh medium inoculated with 20 µl of a 24-hour BM pre-culture. A common baseline was measured at 0 hours, after which one well was destructively sampled every 2 hours from 2 to 24 hours for OD_{600}, viable-cell density and culture pH. OD and pH were available for all 16 isolates. CFU measurements were absent for Staphylococcus sp.1 (species 10); analyses that required viable-cell counts, including the cross-temperature summary panels, therefore used the 15 isolates with complete CFU records. The parallel wells represent destructive time points, not biological replicates.

Viable-cell density was quantified by tenfold serial dilution in sterile PBS, transferring 20 µl of culture into 180 µl of PBS at each step. Five-microliter aliquots were spotted onto TSB-agar plates and incubated at room temperature for 24 hours. Counts were converted to CFU per milliliter by multiplying by the reciprocal dilution and by 200, the conversion from a 5-microliter spot to 1 ml. Colonies were counted at dilutions with clearly separated colonies; values below the countable range were recorded using the lower-limit convention in the source worksheet. Clear within-cycle death was defined operationally as a decline of at least 1 $\log_{10}$ CFU ml^{-1} from the within-cycle maximum to a subsequent minimum. This is a descriptive threshold; no within-isolate significance test was applied to a single destructively sampled trajectory.

Maximum growth rate was calculated from background-corrected OD over 0-24 hours. A linear model was fitted to ln(OD - 0.035) in every three-consecutive-time-point window, and the largest slope was retained as r_{max}. End-of-cycle biomass and viable-cell density were the

24-hour OD and CFU measurements. Viability loss was summarized as $\log_{10}$[CFU(24 h)/CFU_{max}]. For analyses of death kinetics, the maximum death rate was the largest decline in $\log_{10}$ CFU between consecutive 2-hour samples after the within-cycle maximum, provided that the total post-maximum decline reached 1 $\log_{10}$. The pH deviation associated with a trajectory was calculated relative to pH 6.5 at the onset of the detected decline, or at 24 hours when no clear decline was detected.

Physiological responses to initial pH

To characterize species-specific responses to initial pH, the assayed isolates were inoculated into BM containing 5 g l^{-1} glucose and 4 g l^{-1} urea and adjusted to initial pH 3, 5, 7, 9 or 11. Growth curves were recorded for 24 hours using an automated growth-curve reader. The area under the background-corrected OD curve was calculated as an integrated measure of growth. This automated-reader assay was used only for the initial-pH experiment. One trajectory is shown for each plotted isolate-pH combination; time-resolved measurements within a trajectory were not treated as independent replicates.

DNA extraction, 16S rRNA sequencing and data analysis

Community composition was measured by 16S rRNA amplicon sequencing. In the temperature experiment, biological replicate 1 was sequenced across all 10 passages. Replicates 2 and 3 were sequenced primarily over the terminal sampling window. The final temperature feature table contained 924 samples: 600 from replicate 1 and 162 from each of replicates 2 and 3. Coverage of replicates 2 and 3 at 50 °C was incomplete because many cultures had collapsed; only measured samples with matched OD were used as observations. In the dilution-factor experiment, one biological replicate for each of 12 communities at five dilution factors was sequenced on all six passages, giving 360 sequenced samples.

For the temperature experiment, DNA was extracted with the Zymo Research Quick-DNA Fungal/Bacterial 96 Kit. The V4-V5 region was amplified with primers 515F (GTGCCAGCMGCCGCGGTAA) and 907R (CCGTCAATTCMTTTRAGTTT) and sequenced as paired-end reads. Two sequencing batches were processed separately in QIIME 2 (release 2023.9) (32). Primers were removed with cutadapt using --discard-untrimmed, after which DADA2 paired-end denoising was run with trim-left 0/0, truncation lengths 228/205 bp and maximum expected errors 2/2 (33). Feature tables and representative sequences were merged only after batch-specific denoising. The resulting table contained 5,628 ASVs across 924 samples. Temperature-experiment ASVs were annotated by local BLAST against the NCBI 16S ribosomal RNA database, retaining hits with at least 90% identity and 70% query coverage (35).

For the dilution-factor experiment, the sequencing company supplied demultiplexed, primer-trimmed clean FASTQ files for the V4 515F-806R amplicon. These files were imported directly as single-end Phred33 reads; no paired-read merging or additional primer removal was performed in our pipeline. DADA2 denoise-single was run in QIIME 2 2023.9.2 with trim-left 0, truncation length 250 bp, maximum expected errors 2, quality truncation 2, independent pooling, consensus chimera removal and 1,000,000 reads for error-model

learning (33). Taxonomy was assigned with the SILVA 138 NR99 515F-806R classifier at a confidence threshold of 0.7 (34). The final table contained 2,176 ASVs across 360 samples and 35,614,246 non-chimeric reads.

ASV consolidation, isolate matching and abundance reconstruction

The same deterministic post-processing pipeline was applied to both experiments. ASVs were ranked by total reads in the analysis samples, and the 45 most abundant ASVs were considered for species reconstruction. Relative-abundance profiles were normalized within these 45 ASVs. Candidate ASV pairs were joined only when they belonged to the same informative genus and met one of two rules: (i) Pearson $r > 0.95$ together with sequence identity at least 0.99 or a minor-to-major total-read ratio no greater than 0.30; or (ii) sequence identity at least 0.995, Pearson r at least 0.90, presence Jaccard similarity at least 0.85 and a minor-to-major read ratio no greater than 0.30. Connected components of accepted pairwise links defined species-level units, within which raw read counts were summed; the highest-read ASV was retained as the representative sequence. A unit was retained for a community when it exceeded 1% relative abundance in at least three analysis samples or reached 5% in any sample. This produced 25 retained species-level units for the temperature experiment (three merged ASV pairs) and 38 for the dilution-factor experiment (four merged pairs and one three-ASV group). Unmeasured worksheet slots filled with a value of 1 for spreadsheet compatibility were not treated as experimental sampling events.

For downstream analyses, only retained units assigned to the defined species pool of a community were used. Their counts were renormalized to sum to 100% within each measured sample, then multiplied by matched background-corrected OD to reconstruct species-level absolute abundance. Temperature ASV ranking, grouping and presence calls used all W1-W4 samples and OD-supported W5 samples; noise-dominated W5 samples from collapsed cultures were excluded from those steps. Amplicon-derived relative abundance remains subject to extraction, amplification and 16S copy-number biases and should not be interpreted as an exact cell-number fraction. The defined isolate pools, expected-community filter and agreement of retained trajectories across repeated samples limit contamination effects, but uncertainty from both sequencing proportions and OD propagates into reconstructed absolute abundance.

Definition and quantification of experimental population dynamics

Shannon diversity was calculated from all retained taxa in a measured sample. The 1% relative-abundance threshold was used only for richness and survival-fraction calculations and was not used to remove rare taxa from the Shannon probability distribution. For the primary temperature analysis, endpoint Shannon diversity was calculated from day 10; the corresponding endpoint for the dilution-factor experiment was day 6:

$$H(t) = -\Sigma_i \, q_i(t) \ln[q_i(t)]$$

where $q_i(t)$ is the renormalized relative abundance of species i. Accumulated Shannon diversity was calculated by pooling reconstructed absolute abundance over the analysis window (days 8–10 for temperature and days 4–6 for dilution). Days with total reconstructed

abundance below 0.05 were excluded from pooling so that sequencing noise from a collapsed sample could not create artificial diversity. The pooled fraction was

$$q_i,acc = \Sigma_t N_i(t) / [\Sigma_j \Sigma_t N_j(t)],$$

and $Hacc = -\Sigma_i q_i,acc \ln(q_i,acc)$. Hacc therefore captures the number and evenness of populations that contributed over the sampled window, including temporal turnover that can be missed by an endpoint value.

Population instability was calculated from species-resolved absolute abundance after daily renormalization of the retained species to 100%. For species i, $\sigma_i = SD_t[N_i(t)]$ was the sample standard deviation across the analysis window. Population instability (CV) and population instability (std) were defined as

$$\text{population instability (CV)} = \Sigma_i \sigma_i / \text{Mean}_t[\Sigma_i N_i(t)]$$

$$\text{population instability (std)} = \Sigma_i \sigma_i.$$

Both metrics were evaluated over days 8–10 for the temperature experiment (replicates 1-3) and days 4–6 for the dilution-factor experiment (the sequenced replicate). They are sums of population-level temporal variation and are not the CV or standard deviation of total OD. Aggregate-biomass CV and std were calculated separately from the OD time series for robustness analyses. A trajectory was classified as collapsed when mean total reconstructed abundance over the window was below 0.05; collapsed W5 temperature trajectories were assigned zero diversity and instability for condition-level summaries but excluded from within-temperature rank correlations. Figure 1F ranked all non-collapsed temperature observations by population instability (CV) and divided the global ranking into equal high- and low-instability halves. Temperature-specific associations were evaluated continuously by Spearman rank correlation. Binary trajectory displays used an operational population-instability threshold of 0.25; continuous correlations and the global-median analysis do not depend on this threshold.

For the dilution-factor experiment, the accumulated-Shannon-diversity versus population-instability plane was divided descriptively at population instability (CV) = 0.25 and Hacc = 0.8. Low-instability trajectories were labeled stable. High-instability trajectories with high Hacc generally showed turnover among multiple populations and were described as collective fluctuations, whereas high-instability trajectories with low Hacc were repeatedly dominated by one or a few populations and were described as dominant-species fluctuations. These shaded regions organize the observed trajectories; they were not obtained by unsupervised clustering and do not imply discrete dynamical classes.

Replication and statistical units in experimental analyses

The temperature experiment comprised 180 community-temperature replicate time series (12 communities × 5 temperatures × 3 biological replicates). Species-resolved final-window analyses used each replicate as a statistical unit when matched sequencing and OD were available. Replicate 1 was sequenced throughout the 10-passage experiment; replicates 2 and 3 contributed terminal-window observations. The processed feature table contained 924 samples rather than the nominal 960 because sequencing coverage of replicates 2 and 3 was incomplete at 50 °C. Long-trajectory analyses used replicate 1 only.

The dilution-factor experiment comprised 180 biomass time series (12 communities × 5 dilution factors × 3 biological replicates). Only one biological replicate per community-dilution combination was sequenced through all six passages, producing 60 complete composition time series and 360 sequenced samples. All species-resolved quantities in Fig. 4, including population instability, Shannon diversity and accumulated diversity, use this sequenced replicate and its matched OD trajectory. The other two replicates contribute only to analyses based solely on OD or pH and are not independent compositional replicates.

For monoculture physiology, one destructively sampled time course was collected for each isolate-temperature combination; parallel wells were time points rather than replicate cultures. OD and pH were measured for 48 isolate-temperature trajectories (16 isolates × 3 temperatures), whereas CFU-dependent analyses used 45 trajectories (15 isolates × 3 temperatures) because species 10 lacked CFU measurements at all temperatures. Cross-temperature tests treated isolate identity as the paired biological unit. For the initial-pH experiment, one growth trajectory is shown for each plotted isolate-pH combination; time points within a trajectory were not treated as independent replicates.

pH-coupled consumer-resource model

We modeled S microbial species competing for M resources during repeated 24-hour growth and transfer cycles. Species abundance, resource concentration, the shared environmental pH and cumulative biomass production were denoted N_i, $R\alpha$, p and B_i. Their within-cycle dynamics followed Eqs. 1–4:

$$dN_i/dt = N_i[r_i(T)\ \Sigma\alpha\ I_i\alpha(R\alpha,p) - \mu_i(T,p)] \quad (1)$$

$$dR\alpha/dt = -\Sigma_i\ N_i\ r_i(T)\ I_i\alpha(R\alpha,p) \quad (2)$$

$$dp/dt = \Sigma_i\ N_i\ \Sigma\alpha\ r_i(T)\ \gamma\alpha\ I_i\alpha(R\alpha,p) \quad (3)$$

$$dB_i/dt = N_i\ r_i(T)\ \Sigma\alpha\ I_i\alpha(R\alpha,p) \quad (4)$$

The environment-dependent death rate was defined by Eq. 5:

$$\mu_i(T,p) = h(T)[p - p_{opt,i}]^2 \quad (5)$$

Here $p_{opt,i}$ is the pH optimum of species i and h controls the strength of pH-associated death. Resource uptake followed pH-gated Monod kinetics, $I_i\alpha(R\alpha,p) = f_i\alpha R\alpha/(K_m + R\alpha)\ \varphi_i(p)$, with $K_m = 0.3$. The triangular tolerance function $\varphi_i(p) = \max[0, 1 - |p - p_{opt,i}|/p_{tol}]$ had $p_{tol} = 2$ pH units. Each species consumed six randomly selected resources with Dirichlet(1,...,1) preference weights; its growth coefficient r_i scaled this preference vector. Each resource had a coefficient $\gamma\alpha$ specifying how its consumption changed the shared pH. For each simulated community, $\gamma\alpha$ values were drawn independently from Normal(0, 3^2), held fixed over all passages of that community, and redrawn for the next community. Species could therefore affect one another through both resource depletion and the shared pH variable.

N_i represents viable abundance carried through the within-cycle dynamics, whereas B_i records cumulative gross biomass production. Growth contributes to both variables, but death is subtracted only from N_i. B_i is reset to zero at the beginning of each passage and is not transferred. The primary model-ensemble analyses use B_i as the readout compared with the experimentally reconstructed OD-based abundance; N_i determines viable carry-over.

At the end of each cycle, abundances below 10^{-7} were set to zero and populations were transferred as $N_{i,next} = f(N_{i,end} + D)$, with $f = 1/30$ and $D = 10^{-6}$. Resources were reset to $R_0 = 1$ for each of the 24 resources and pH was reset to $p_{env} = 7$. Abundances below 10^{-12} after transfer were treated as zero. The $h = 0$ reference removes pH-associated death while retaining resource competition, pH modification and pH gating of uptake.

Model parameterization, numerical integration and ensemble simulations

The main ensemble used $S = 12$ species and $M = 24$ resources. Species growth rates were drawn as $r_i = \langle r_i \rangle \max(1 + 0.3\, Z_i, 0)$, with Z_i standard normal, and pH optima were drawn independently from Uniform(5.5, 8.5). The parameter grid contained 100 mean-growth values from 0 to 1 and 100 h values from 0 to 0.15. Figure 3 focuses on $h = 0.03$–0.15, with $h = 0$ shown as the death-free reference. At every grid point, 5,000 independently assembled communities were simulated. The resource preferences, growth-rate deviations, pH optima and $\gamma\alpha$ values constituted quenched disorder: they were fixed during a community trajectory and resampled between ensemble members.

Within-cycle ordinary differential equations were integrated for 24 model hours with scipy.integrate.solve_ivp using LSODA (relative tolerance 10^{-6}, absolute tolerance 10^{-9} and maximum step 0.48 h), with BDF as a fallback when LSODA did not converge (36). The right-hand side was compiled with Numba. Each community was propagated for 30 passages; the final 10 endpoints were saved, and all primary classification, participation and diversity-ordering statistics were calculated from the final three passages. Independent deterministic seeds were generated as seed_base + (i_r + 1) × 1,000,000 + (i_h + 1) × 10,000 + (i_ensemble + 1), permitting exact regeneration of each member.

Definition of stable, fluctuating and collapsed model trajectories

Classification used B_i over the final three passages. Model population instability (CV) was defined exactly as in the experimental analysis, $\Sigma_i SD_t[B_i(t)]/Mean_t[\Sigma_i B_i(t)]$. A simulation was classified as collapsed when its mean total B over the window was below 10^{-3}. Among non-collapsed simulations, population instability (CV) greater than 0.1 defined a fluctuating trajectory and lower values defined a stable trajectory. The empirical distribution of model population instability was bimodal across the analyzed ensemble, supporting this operational separation (fig. S16). Shannon diversity was calculated from the final-passage B_i distribution after setting B_i values at or below 10^{-3} to zero. Parameter points with fewer than 10 stable or 10 fluctuating communities were left blank in between-class maps.

Fluctuation participation

For each fluctuating simulation, σ_i was the temporal standard deviation of species i's B_i over the final three passages. The normalized fluctuation-participation ratio was

$$PR = (\Sigma_i \sigma_i)^2 / (S \Sigma_i \sigma_i^2).$$

PR ranges from 1/S when temporal variation is concentrated in one species to 1 when all S species contribute equally. PR measures how broadly variation is distributed; it does not measure synchrony, phase relationships or compensation.

Pairwise probability difference in diversity

At each parameter point with at least 10 stable and 10 fluctuating non-collapsed simulations, one member was drawn conceptually from each subset and their final-passage Shannon diversities were compared. The reported effect size was

$$\Delta P = P(H\_F > H\_S) - P(H\_F < H\_S).$$

Here H_F and H_S denote the final-passage Shannon diversities of a fluctuating and a stable simulation, respectively. ΔP ranges from −1 to 1. Positive values mean that a randomly selected fluctuating simulation is more likely to be more diverse than a randomly selected stable simulation; values near zero indicate no consistent ordering. All fluctuating-stable pairs in the 5,000-member ensemble were represented through the equivalent rank-biserial statistic; ties contribute zero. ΔP is an ensemble-level effect size, not a correlation coefficient and not an assignment of dynamical origin to an individual trajectory.

Schematic dynamical summary (Fig. 3F)

Figure 3F summarizes three continuous ensemble quantities: the fluctuating fraction, normalized PR among fluctuating communities and ΔP. At moderate h, fluctuations were relatively uncommon but broadly distributed among species and ΔP was positive. At larger h, fluctuations were more frequent, participation declined and ΔP approached zero. Separate S = 1 simulations and explicit constituent-monoculture trajectories were used to determine whether instability could occur without community assembly, whereas S = 24 and alternative resource-supply and γα distributions tested structural robustness (fig. S18).

The regions and boundaries in Fig. 3F are schematic guides to these continuous trends, not fitted transitions or a classification imposed on every simulation. 'Collective fluctuation' denotes the part of the parameter range with broader participation and positive diversity ordering; 'dominant-species fluctuation' denotes ranges in which variation is more concentrated and diversity no longer orders stable and fluctuating ensembles. Both patterns can occur near the same parameter values.

Relationship between experimental conditions and model parameters

The notation $r_i(T)$ and h(T) permits temperature dependence in principle, but no function mapping temperature to either parameter was fitted. Experimental temperatures were placed only ordinally in the schematic summary using measured growth and end-of-cycle viability patterns. Temperature changes growth, pH modification, stress tolerance and other traits simultaneously, so the placement is not a mechanistic assignment of an experimental trajectory to a unique model coordinate or fluctuation form.

Model controls and robustness analyses

The h = 0 sweep served as the death-free resource-competition reference. S = 1 simulations quantified monoculture instability, and S = 12 and S = 24 ensembles tested the dependence of the maps on species-pool size. The main S = 12 ensemble used $\gamma\alpha \sim \text{Normal}(0, 3^2)$, $R_0 = 1$

and six-resource preference vectors. Additional ensembles varied S, the γα distribution and variance, and resource supply. These controls assess whether the qualitative progression from stable dynamics to broadly shared and then more concentrated fluctuations persists, rather than whether boundary coordinates are invariant.

The primary metrics use the final three passages. The final 10 saved passages were retained for trajectory inspection and window sensitivity, but were not substituted for the three-passage PR reported in the main analysis. Classification robustness was checked against the continuous bimodal distribution of model population instability and the independent collapse criterion (fig. S16). Fixed-growth slices expose the same h-dependent trends without relying on interpolation across the two-dimensional parameter grid (fig. S17).

The dilution-factor experiment was not fitted by changing a single model parameter. Dilution changes carry-over, initial abundance and time spent at high density simultaneously, whereas h changes only the strength of the model death term. Agreement between their qualitative trends is therefore interpreted as consistency between two serial-passage systems, not as a calibrated correspondence between dilution factor and h.

Statistical analysis

Analyses were implemented in Python and R from the scripts archived with the source data. Spearman rank correlations were calculated separately within each temperature using non-collapsed observations. The global high- versus low-instability comparison in the analysis code used an unpaired one-sided Wilcoxon rank-sum test for the directional hypothesis that the high-instability half had greater diversity; Benjamini-Hochberg-adjusted values were also written by the script. Continuous between-condition analyses report both t tests and Wilcoxon tests, with Benjamini-Hochberg correction within each metric, experiment and analysis window. Differences in fluctuating proportions were evaluated by Fisher's exact test and an uncorrected two-sample proportion test; dilution trends were evaluated with a binomial generalized linear model for binary outcomes and a linear model for continuous outcomes. Monoculture temperature comparisons were paired by isolate. Paired t tests were used when Shapiro-Wilk tests supported normality of paired differences and paired Wilcoxon signed-rank tests otherwise. OD and growth-rate comparisons were two-sided. Directional CFU comparisons were one-sided on $\log_{10}$ CFU; changes smaller than 1 $\log_{10}$ were treated as plate-counting ties in that analysis. No independent replicate was inferred from destructive time points. Exact n and the statistical unit are reported in the corresponding methods and legends.

Data and code availability

Raw clean FASTQ files, exported ASV tables and sequences, species-level Excel tables, community OD and pH measurements, monoculture OD-CFU-pH trajectories, processed source data and the analysis and simulation scripts will be deposited in a public repository before publication. The archived model package will include the configuration files, deterministic seed rule and scripts used to reproduce the ensemble summaries.

Supplementary Text

Interpretation of collective and dominant-species fluctuations

Collective and dominant-species fluctuations are phenomenological descriptions of how temporal variation is distributed among populations. In a collective fluctuation, several coexisting populations contribute appreciably to the trajectory. In a dominant-species fluctuation, most temporal variation is concentrated in one or a few abundant populations. The two forms are endpoints of a continuum and can coexist within the same environmental condition; neither label identifies a mechanism from an experimental trajectory alone.

PR and ΔP provide complementary ensemble summaries. PR quantifies the number of populations contributing variation within a model trajectory, whereas ΔP asks whether diversity tends to be ordered between fluctuating and stable ensembles at the same parameter values. High PR with positive ΔP is consistent with collective fluctuations being more prevalent; low PR with ΔP near zero is consistent with variation being concentrated in fewer populations. These statistics do not prove the origin of an individual fluctuation, and PR does not distinguish synchronous from compensatory variation.

Interpretation of the model growth-death parameter space

Within the model, mean growth controls renewal during a passage and h controls how strongly deviation from a species' pH optimum reduces viable abundance. At $h = 0$, surviving communities converged under the simulated serial-passage protocol. At moderate h, environmental modification supplied an additional channel of coupling among species; fluctuations were rare but often involved several populations, and fluctuating communities were more likely to be diverse. At larger h, single-species oscillation or collapse became more frequent, and community variation could be dominated by a population that became unstable after reaching high abundance. These are behaviors generated by the equations, not demonstrations of the mechanism in the experiment.

The parameter space should not be read as a claim that temperature changes only growth and death. Temperature can alter resource uptake, metabolic products, pH tolerance and interactions simultaneously. The model also treats acidification and alkalinization symmetrically: pH is reset to 7, pH optima are sampled uniformly from 5.5 to 8.5, and death is quadratic in the deviation from each optimum. This symmetry is an ensemble-design simplification. The experimental medium began at pH 6.5, and acidifying and alkalinizing trajectories need not be physiologically equivalent.

Interpretation of the dilution-factor experiment

The dilution-factor experiment tests whether the two fluctuation forms recur at a fixed temperature along an independent propagation axis. Figure 4C is not an unsupervised classification. The shaded areas denote low-instability trajectories, high-instability trajectories with high accumulated diversity and multispecies turnover, and high-instability trajectories with low accumulated diversity and repeated dominance by one or a few

populations. Representative trajectories connect positions in this plane to the underlying composition time series.

Dilution factor changes carry-over abundance, bottleneck strength, demographic noise, the time required to reach high density and the duration of late-cycle environmental feedback. Within the tested range, stronger dilution gives populations more time to grow from low density before the next high-density phase and was associated with lower inter-passage population instability. Because these effects are coupled, the experiment does not isolate within-cycle death or establish that any one process caused the observed trend. The enrichment of dominant-species fluctuations at lower dilution is therefore a descriptive result that is qualitatively consistent with, but not a direct test of, the model interpretation.

Robustness and scope of the central conclusions

The experimental diversity-instability result was examined with endpoint, mean-daily and accumulated Shannon diversity, richness and Simpson-type measures, and with normalized and unnormalized population-instability measures. Sliding windows that all ended on the final experimental day tested sensitivity to the starting passage, while the final-three-passage window was retained as the primary definition. Aggregate-biomass variation and temporal Bray-Curtis dissimilarity were analyzed separately and were not substituted for population instability.

Model robustness focused on the continuity of the ensemble trends across species-pool size, resource supply and pH-modification distributions, the operational stability and collapse criteria, and fixed-growth slices through h. The most specific mechanistic example compared each constituent monoculture with the assembled community under identical quenched traits: all monocultures were stable, whereas the assembled high-diversity community showed persistent irregular fluctuations (fig. S18). This example establishes emergence within the model realization; it does not identify the cause of the experimental fluctuations.

References

32. E. Bolyen et al., Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852-857 (2019).

33. B. J. Callahan et al., DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581-583 (2016).

34. C. Quast et al., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590-D596 (2013).

35. C. Camacho et al., BLAST+: architecture and applications. BMC Bioinformatics 10, 421 (2009).

36. P. Virtanen et al., SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261-272 (2020).

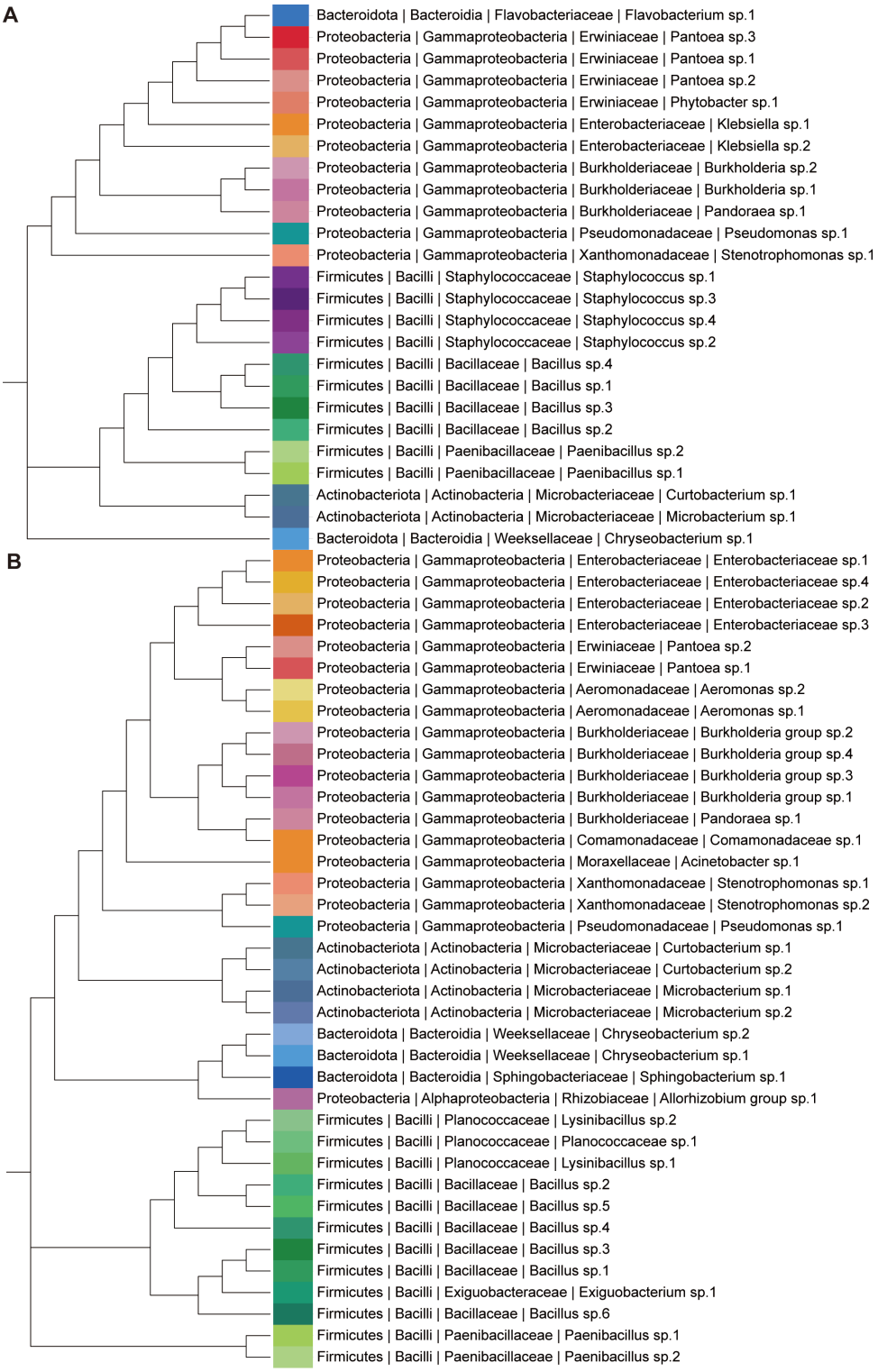

Fig. S1. Taxonomic composition and phylogeny of the two isolate libraries. (A) The 25-isolate library used in the temperature experiment. (B) The 36-isolate library used in the dilution-factor experiment. Tip labels report taxonomic assignments; branch lengths indicate relative phylogenetic distance, and colors distinguish higher taxonomic groups.

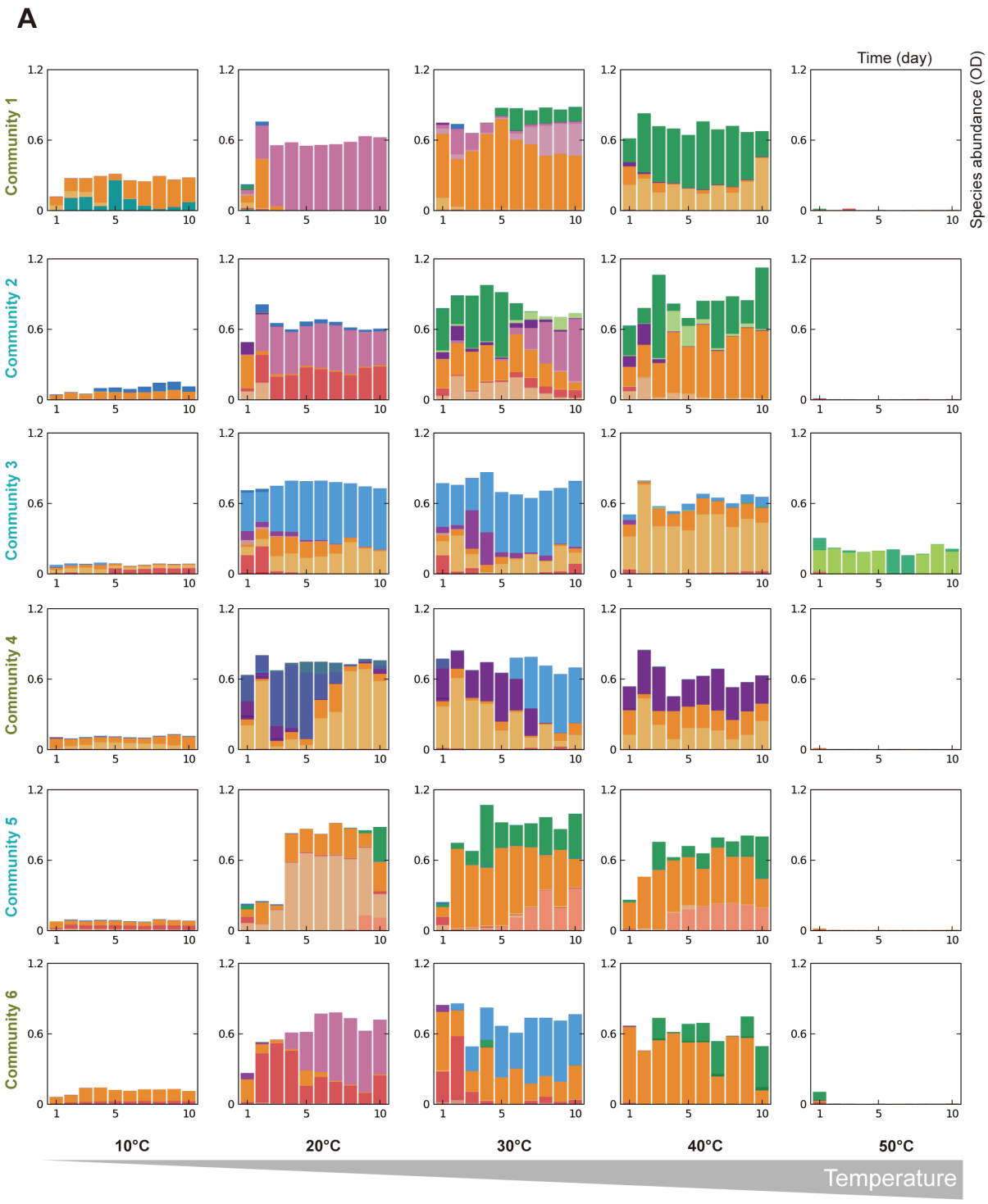

Fig. S2. Species-resolved absolute-abundance trajectories for Communities 1-6 across temperature. Bars show reconstructed absolute abundance (background-corrected OD × sequencing relative abundance) for replicate 1 over 10 passages; columns are 10, 20, 30, 40 and 50 °C. Colors identify retained species-level units.

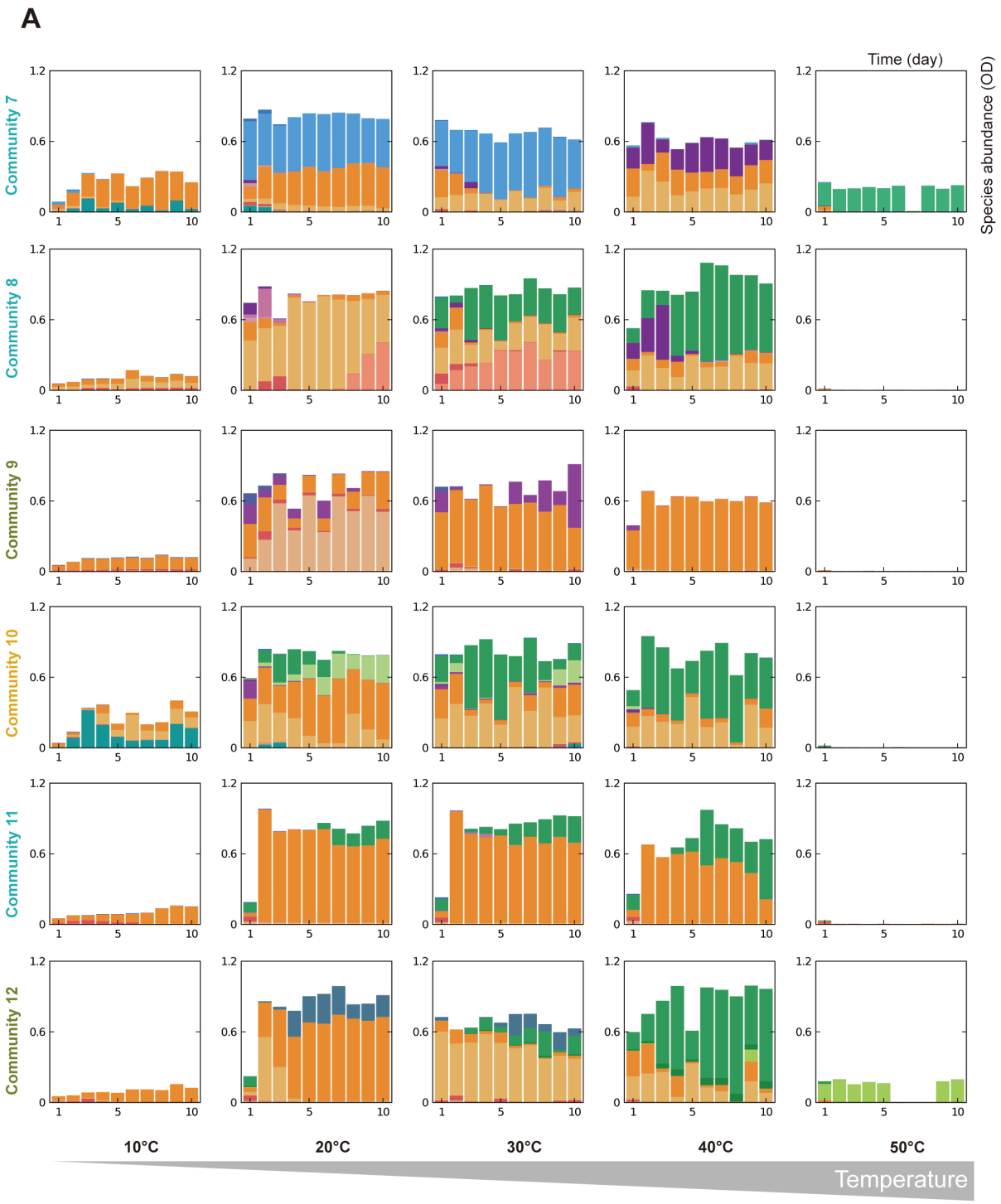

Fig. S3. Species-resolved absolute-abundance trajectories for Communities 7-12 across temperature. Plotting conventions are as in fig. S2.

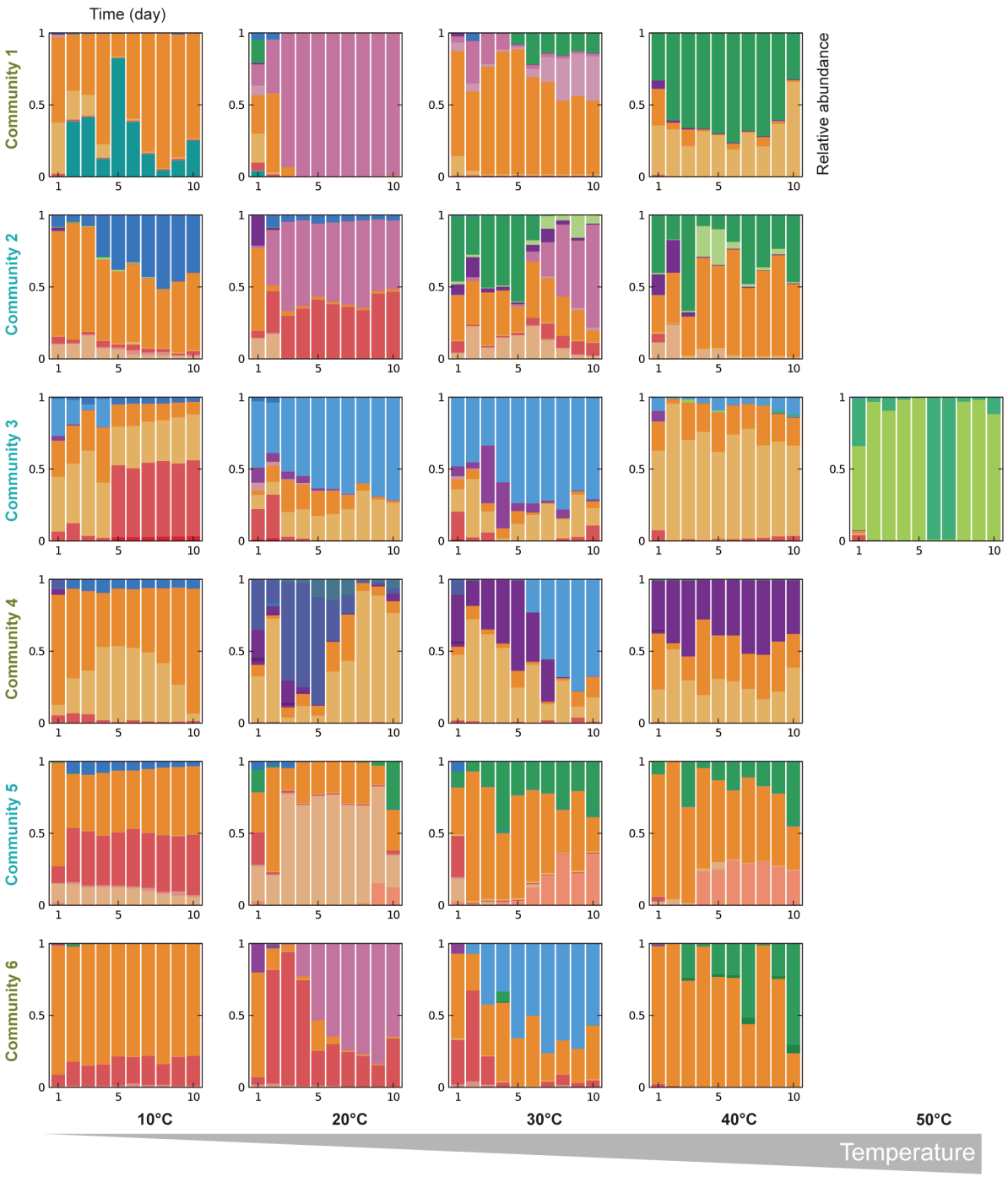

Fig. S4. Species relative-abundance trajectories for Communities 1-6 across temperature. Bars are normalized within each measured sample.

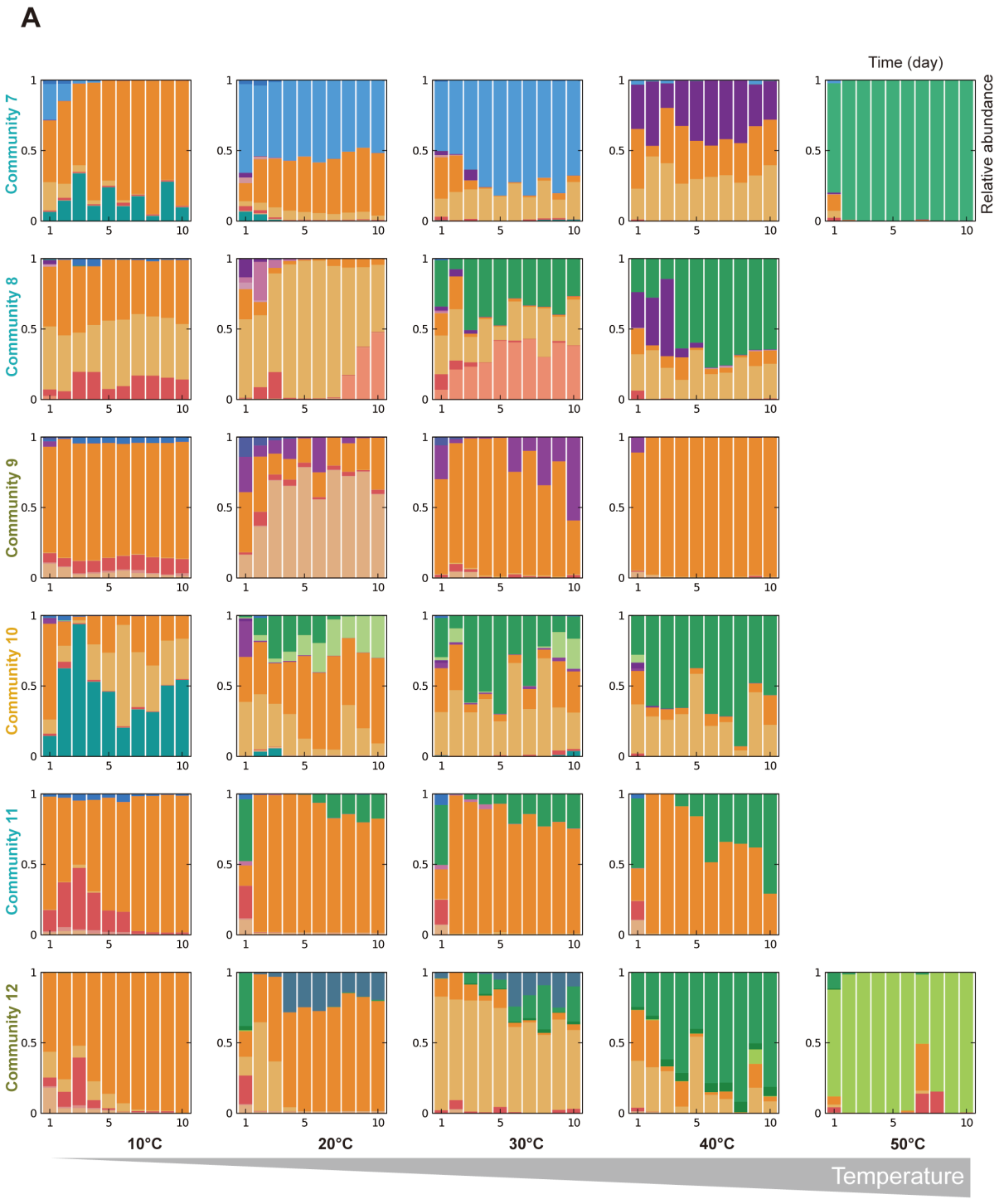

Fig. S5. Species relative-abundance trajectories for Communities 7-12 across temperature. Plotting conventions are as in fig. S4.

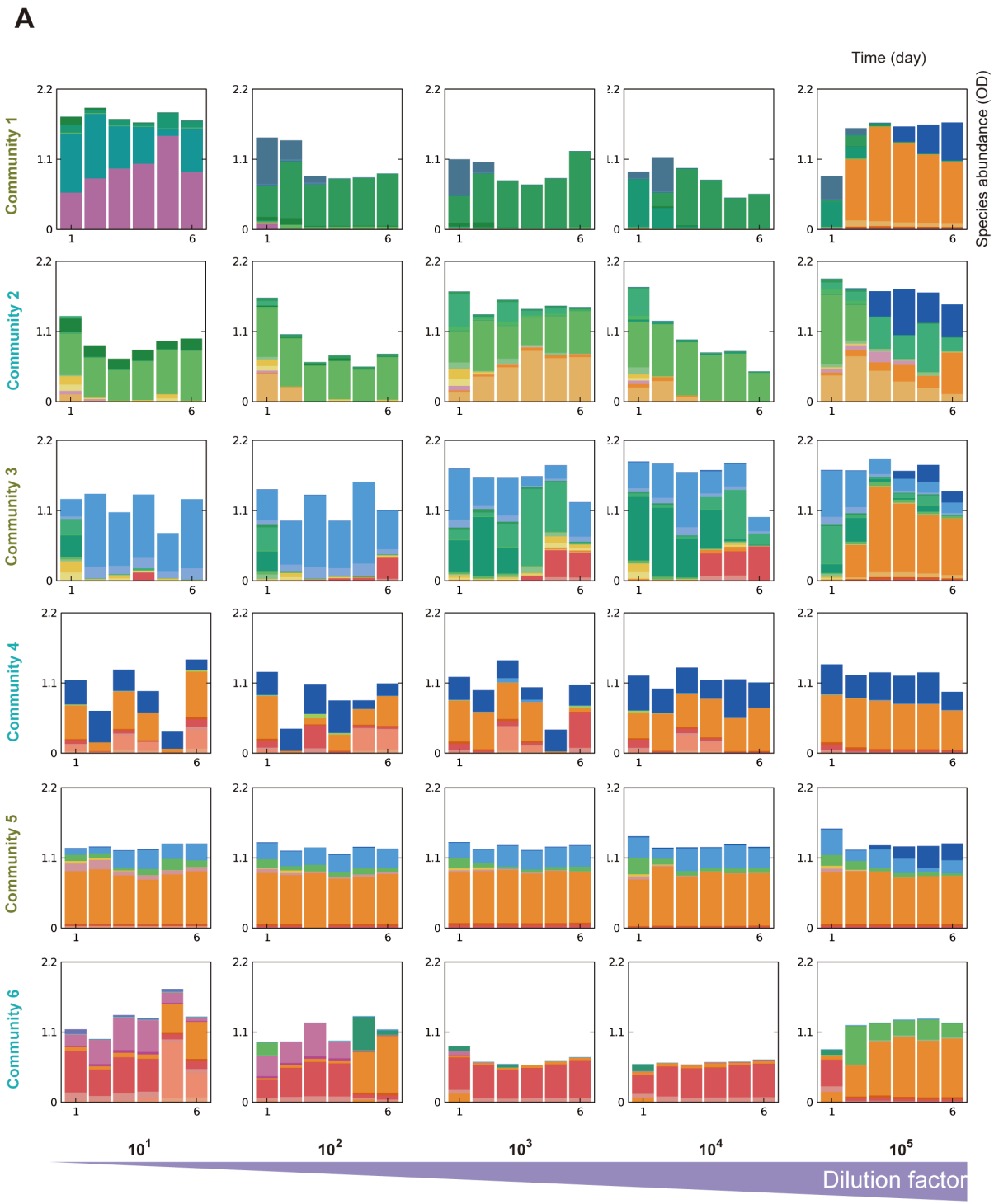

Fig. S6. Species-resolved absolute-abundance trajectories for Communities 1-6 across the dilution-factor experiment. Columns show daily dilution factors from 10^1 to 10^5 at 30 °C; bars show the sequenced replicate over six passages.

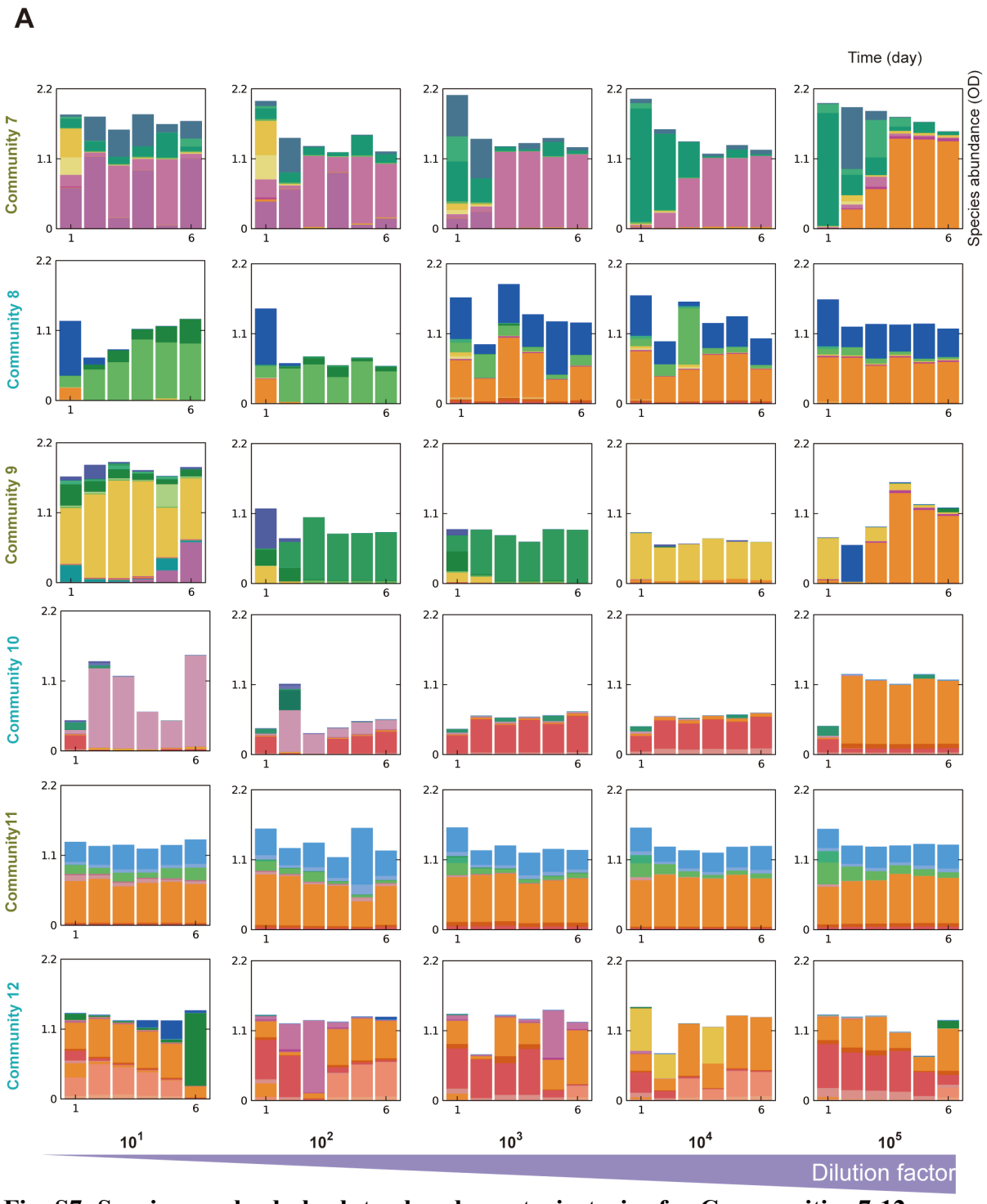

Fig. S7. Species-resolved absolute-abundance trajectories for Communities 7-12 across the dilution-factor experiment. Plotting conventions are as in fig. S6.

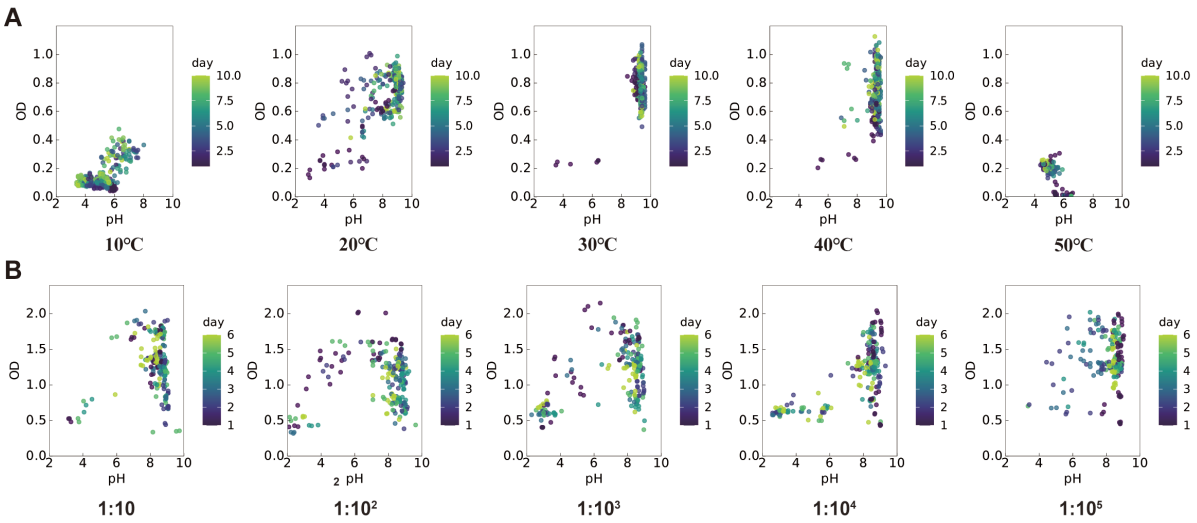

Fig. S8. Relationship between pH and biomass across propagation conditions. (A) All community measurements across the five temperatures; color denotes passage day 1-10. (B) All measurements across dilution factors 10^1–10^5 at 30 °C; color denotes passage day 1-6.

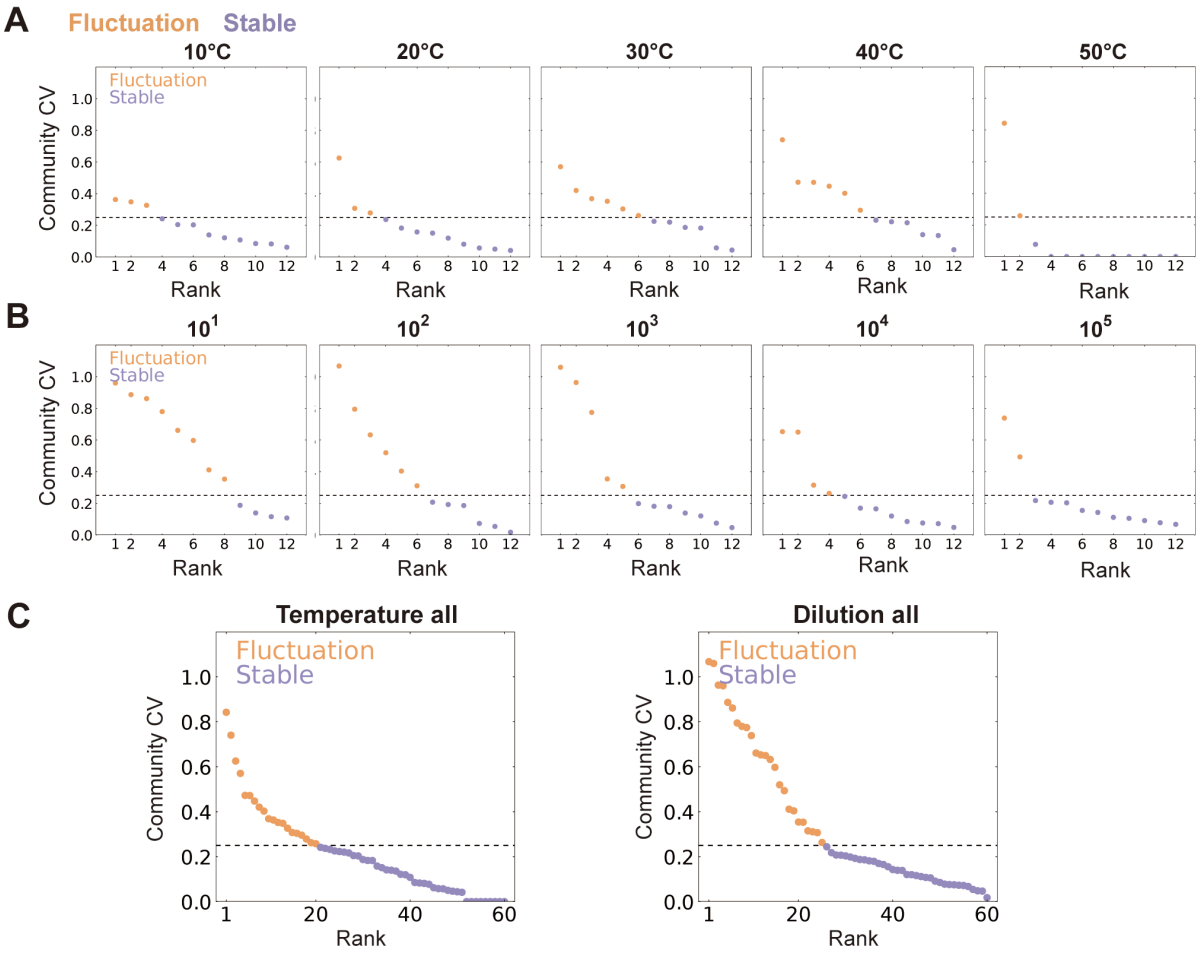

Fig. S9. Ranked distributions of population instability (CV). (A) Temperature conditions. (B) Dilution factors. (C) Pooled rankings within each experiment. The dashed line marks the operational value 0.25 used for binary trajectory displays; orange and purple denote values above and below this guide, respectively.

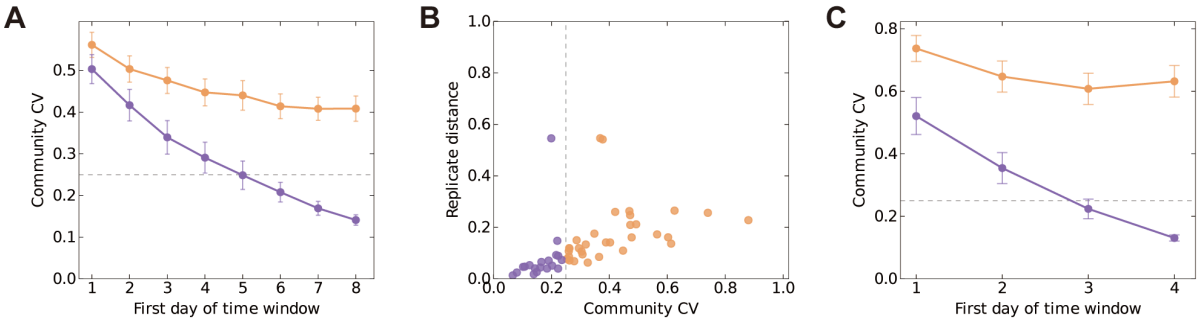

Fig. S10. Sensitivity of population instability to the analysis window. (A) Temperature experiment: mean population instability for windows ending on day 10 as the first included day was shifted from day 1 to day 8. (B) Between-replicate compositional distance over days 8–10 versus population instability in the temperature experiment. (C) Dilution-factor experiment: analogous windows ending on day 6 with starting days 1-4. Orange and purple show trajectories classified above and below 0.25 from the final-three-passage window; points and error bars are mean ± s.e.m.

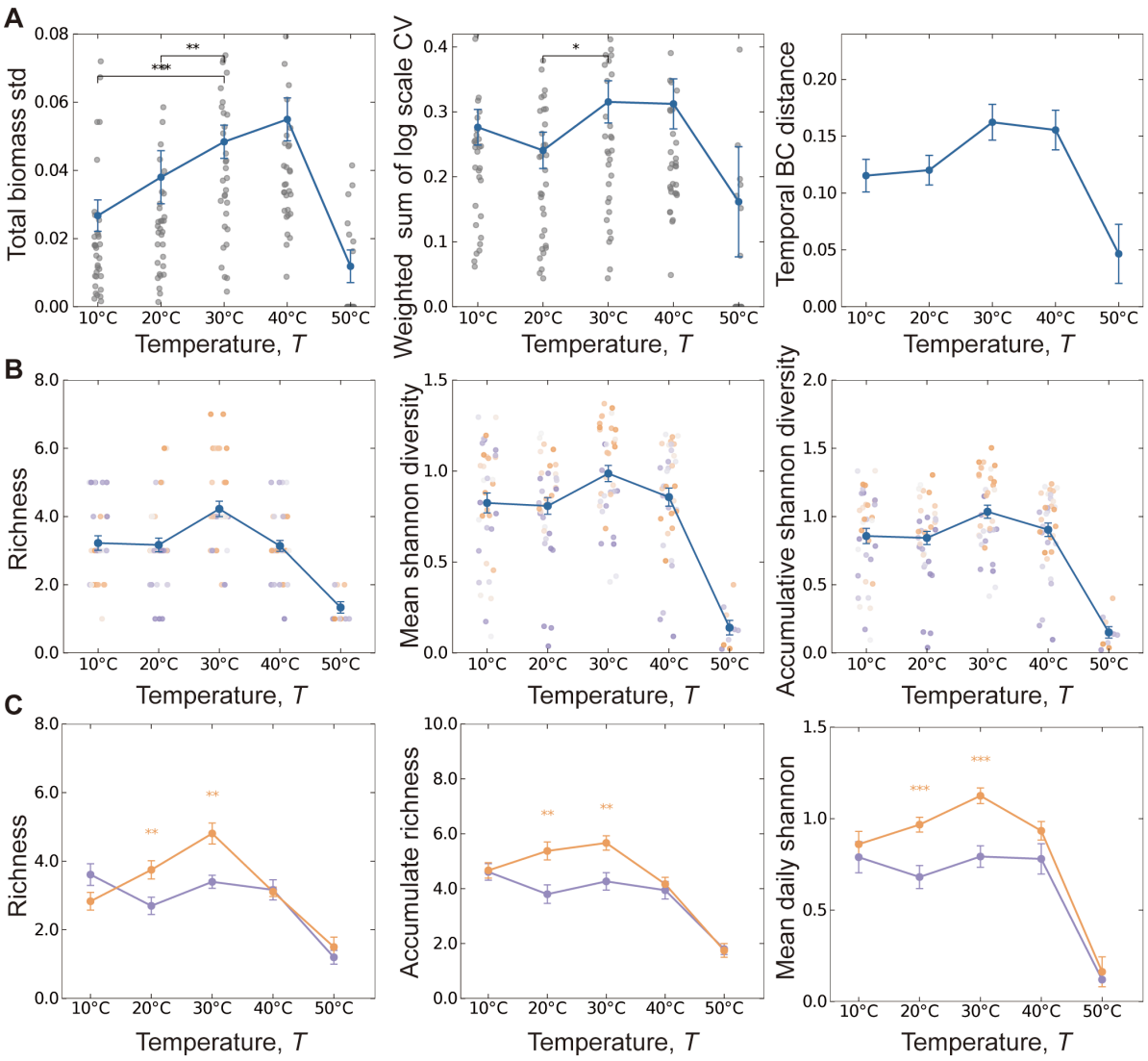

Fig. S11. Robustness of the temperature-dependent dynamics to alternative summary measures. (A) Absolute variation of aggregate biomass, abundance-weighted log-scale population variation and mean temporal Bray-Curtis dissimilarity. (B) Richness, mean-daily Shannon diversity and accumulated Shannon diversity. (C) Diversity summaries after division into high- and low-instability groups. Points show individual community-replicate observations where applicable; lines and error bars show mean ± s.e.m.

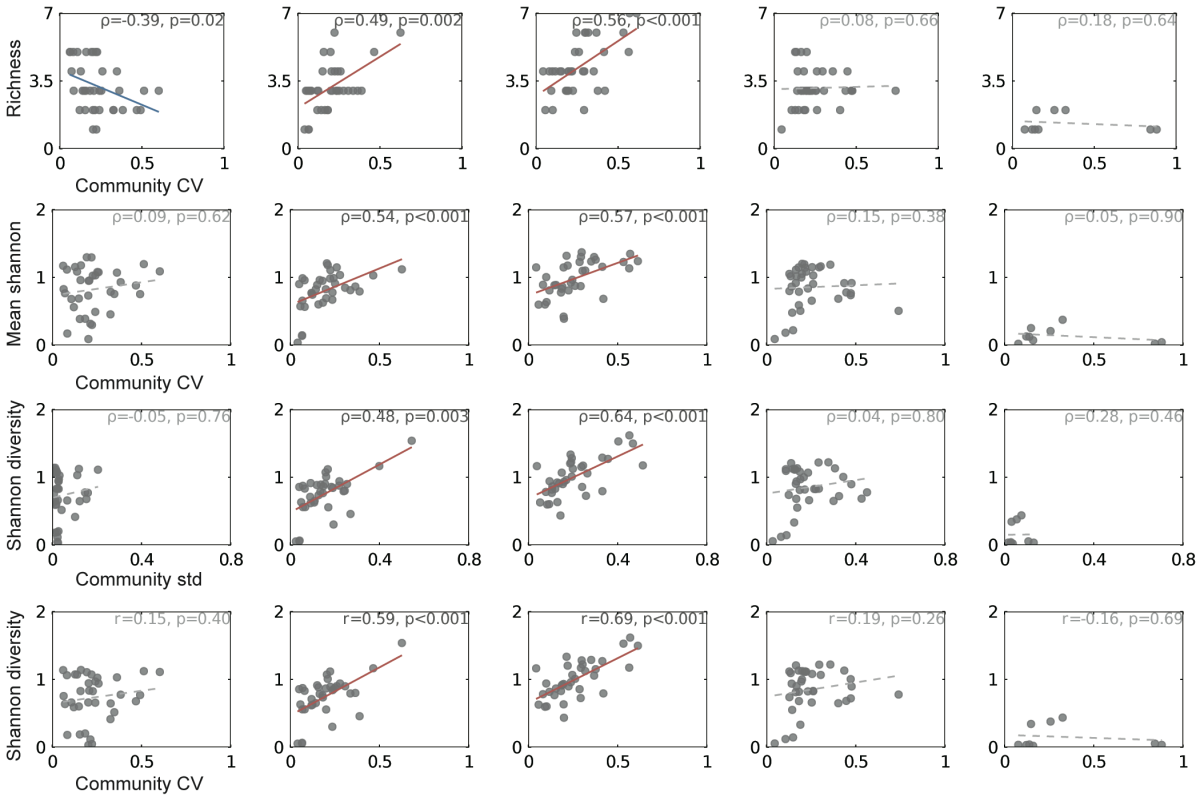

Fig. S12. Diversity-instability associations across alternative diversity and instability metrics. Columns correspond to 10, 20, 30, 40 and 50 °C. Rows show richness versus population instability (CV), mean Shannon diversity versus population instability (CV), Shannon diversity versus population instability (std), and Shannon diversity versus population instability (CV). Correlation coefficients and P values are shown in each panel; solid colored fits mark significant associations and dashed fits non-significant associations.

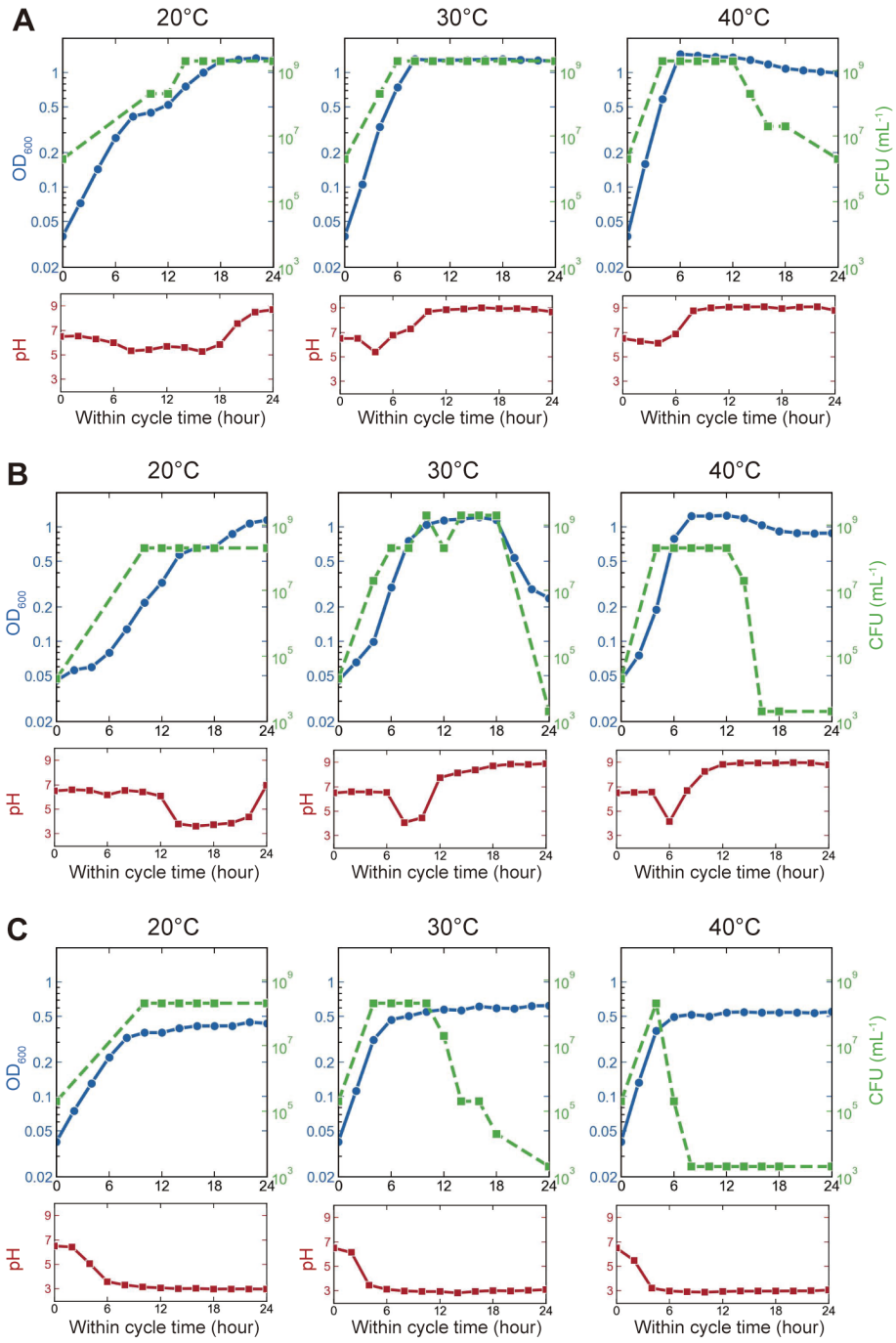

Fig. S13. Representative within-cycle monoculture trajectories. OD_{600}, pH and CFU measurements are shown for three selected isolates at 20, 30 and 40 °C. Each time point was obtained from a separate destructively sampled well; lines connect measurements from the same isolate-temperature trajectory.

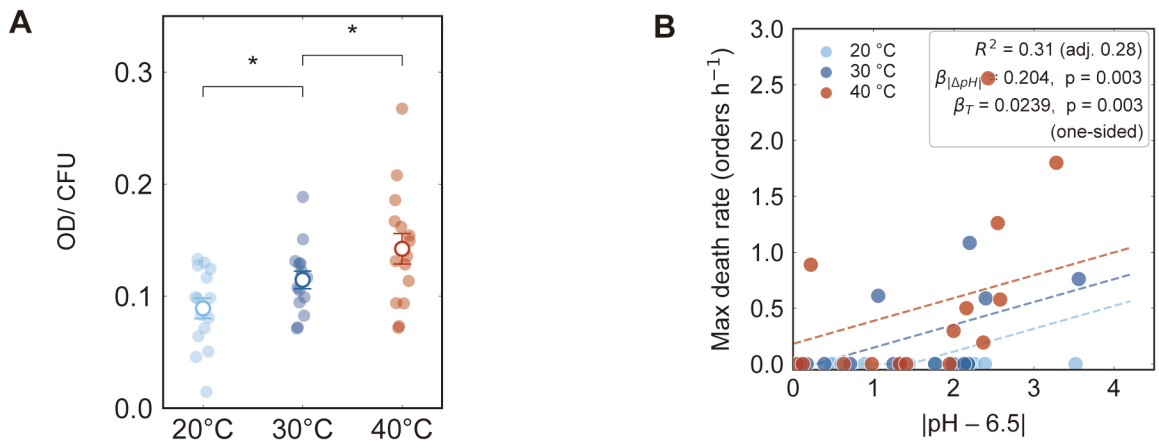

Fig. S14. Temperature and pH associations of within-cycle viability loss. (A) End-of-cycle OD_{600} divided by $\log_{10}$-transformed CFU density for the 15 isolates with CFU measurements at 20, 30 and 40 °C. Points show isolates; open circles and error bars show mean ± s.e.m.; brackets denote the directional paired comparisons specified in the statistical methods. (B) Maximum within-cycle death rate versus the absolute pH deviation from 6.5. Maximum death rate was the largest decline in $\log_{10}$ CFU between consecutive 2-hour samples after the within-cycle maximum and was set to zero unless the total post-maximum decline reached 1 $\log_{10}$. Dashed lines are predictions from the linear model containing pH deviation and temperature; the inset reports the adjusted R2 and one-sided coefficient tests. Colors denote temperature. Staphylococcus sp.1 was omitted because no CFU series was available.

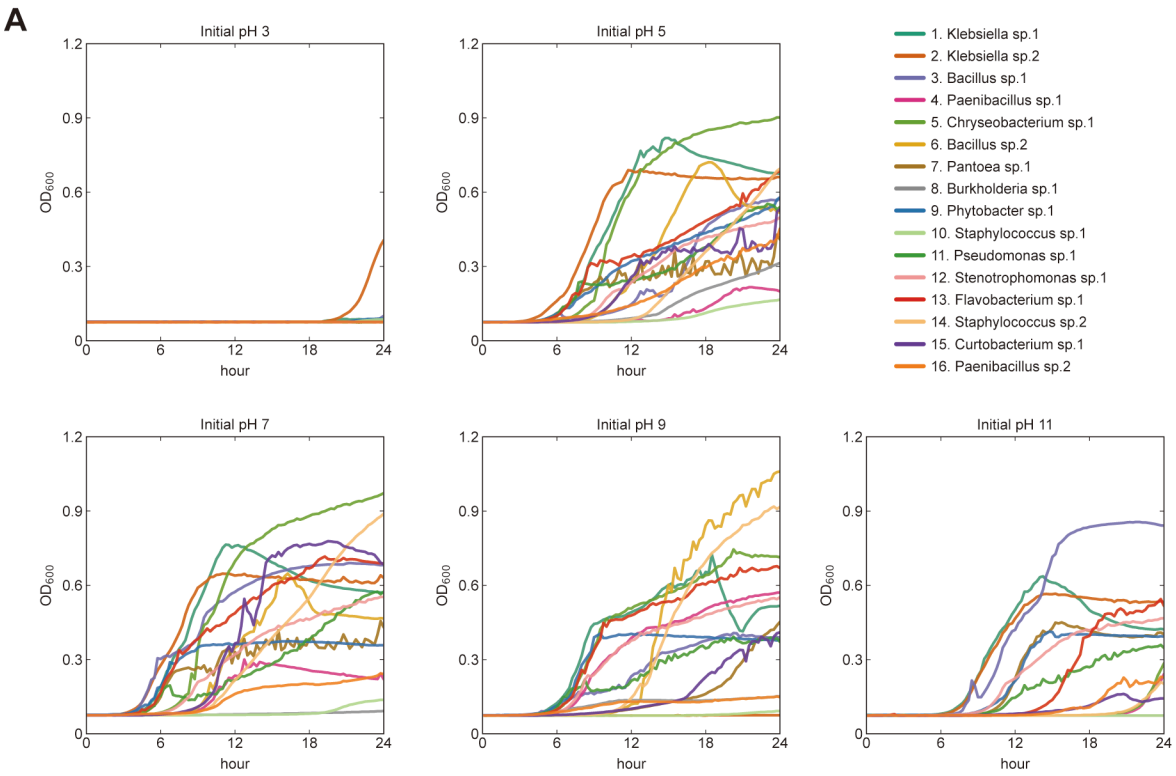

Fig. S15. Isolate growth responses to initial pH. OD_{600} trajectories were recorded for 24 hours in BM adjusted to initial pH 3, 5, 7, 9 or 11. Colors identify the isolate trajectories shown in the legend.

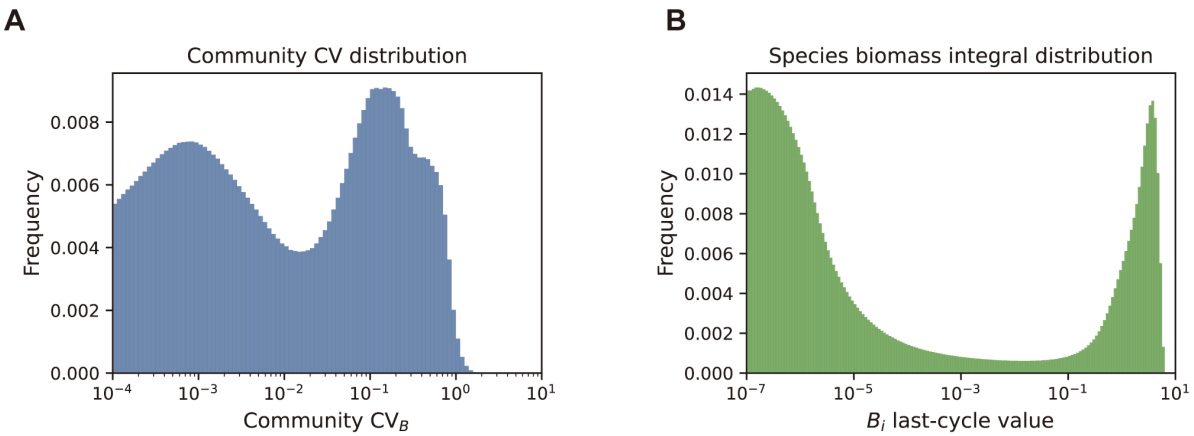

Fig. S16. Operational separation of model outcomes. (A) Distribution of model population instability (CV) calculated from B_i over the final three passages, showing a stable bimodal structure across the ensemble. (B) Distribution of final-passage species biomass, used together with mean total B to distinguish surviving from collapsed simulations. Main analyses used population instability (CV) > 0.1 for fluctuating non-collapsed trajectories and mean total $B < 10^{-3}$ for collapse.

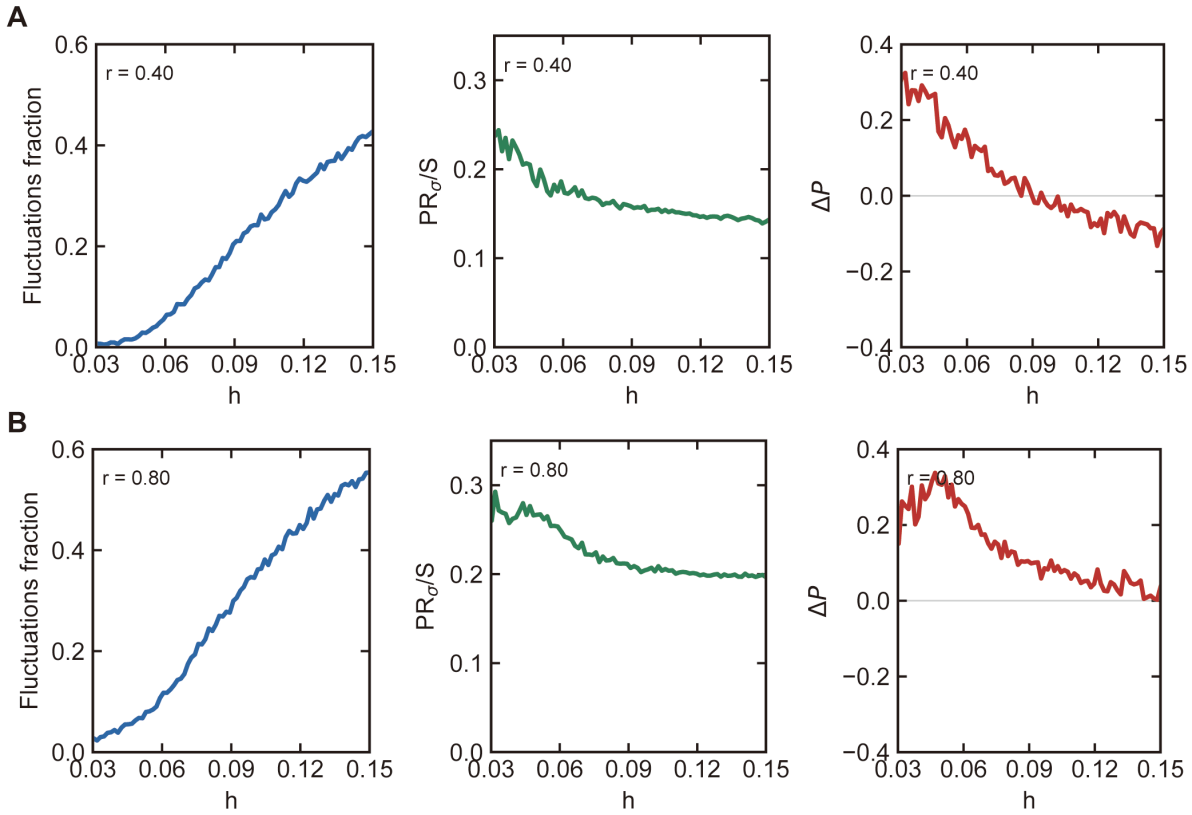

Fig. S17. Fixed-growth slices through the model parameter map. The fluctuating fraction, normalized participation ratio among fluctuating communities and ΔP are shown as h increases from 0.03 to 0.15 at mean growth rates $\langle r_i \rangle$ = 0.40 (A) and 0.80 (B). All quantities were computed from the final three passages of the S = 12, M = 24, R_0 = 1 ensemble.

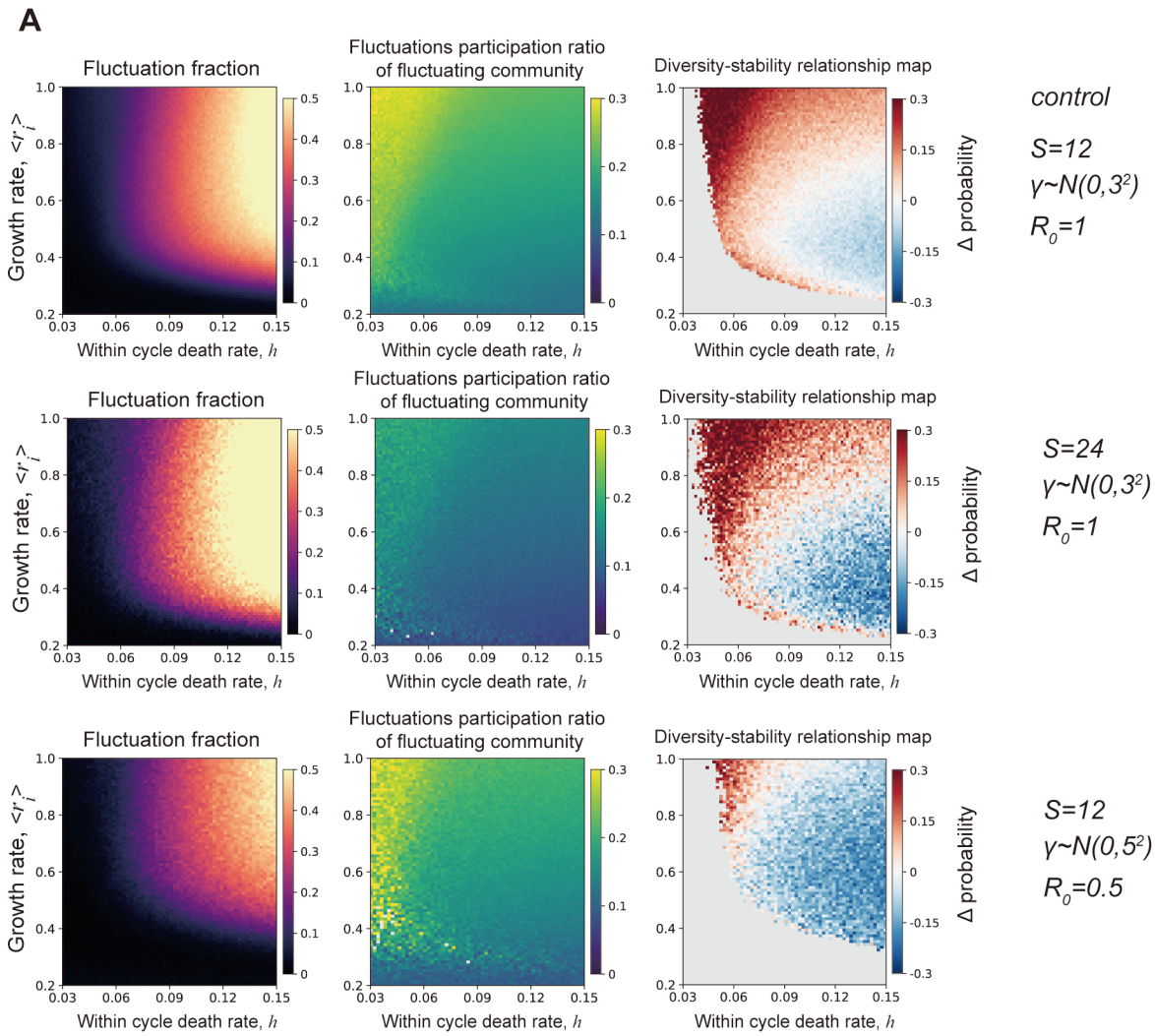

Fig. S18. Robustness of model ensemble trends to alternative parameterizations. Columns show the fraction of fluctuating communities, the normalized fluctuation-participation ratio (PR) among fluctuating communities, and the pairwise probability difference in Shannon diversity between fluctuating and stable communities (ΔP). Rows show the primary parameterization (S = 12, $\gamma\alpha \sim N(0, 3^2)$, $R_0 = 1$), an expanded species pool (S = 24, $\gamma\alpha \sim N(0, 3^2)$, $R_0 = 1$), and altered pH-modification variance and resource supply (S = 12, $\gamma\alpha \sim N(0, 5^2)$, $R_0 = 0.5$). Across these parameterizations, increasing within-cycle death strength increased the prevalence of fluctuations, reduced their participation across species, and weakened or reversed the diversity ordering between fluctuating and stable communities.

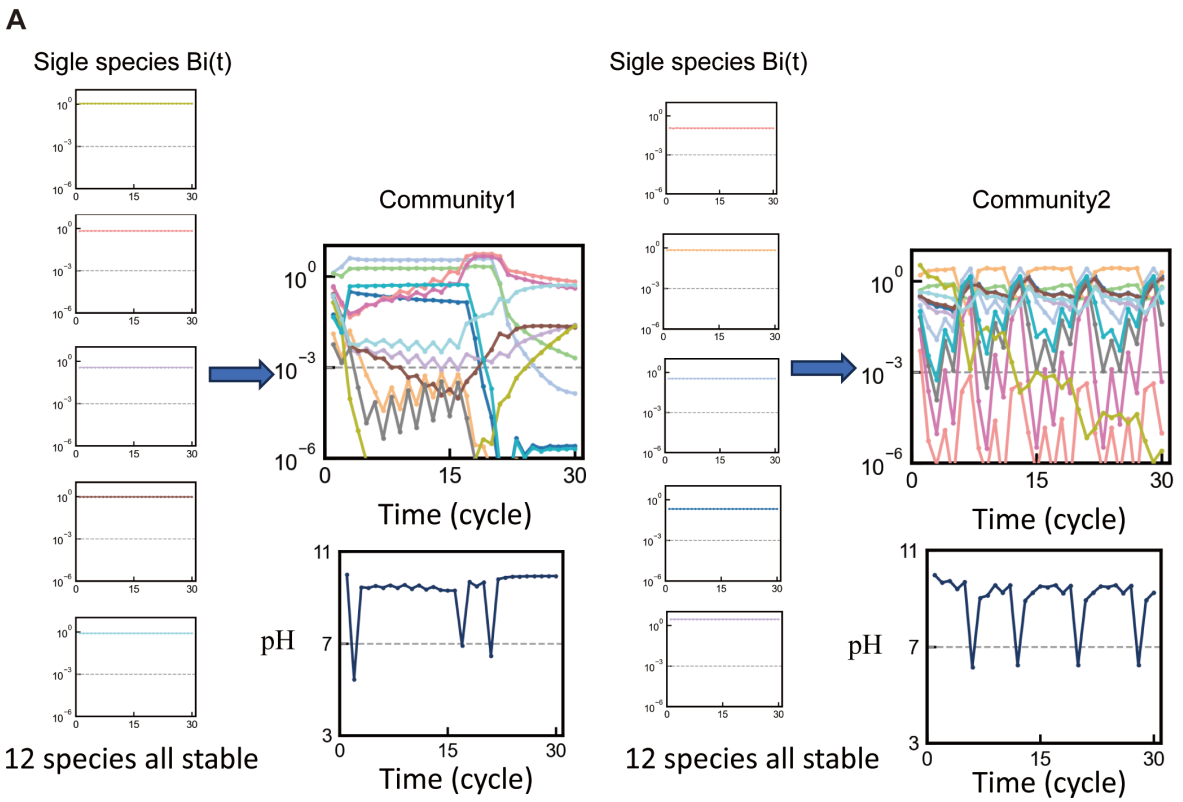

Fig. S19. Model robustness and examples of emergent fluctuation. The parameter maps show the fluctuating fraction, normalized participation ratio and ΔP for the main S = 12 ensemble and for ensembles with S = 24 or an alternative γα variance and resource supply. Parameter changes shift quantitative boundaries but retain the qualitative change in fluctuation participation and diversity ordering. The trajectory panels show two quenched-disorder realizations in which every constituent monoculture remains stable over 30 passages, whereas assembly of the same 12 species produces persistent irregular multispecies fluctuations; the lower panels show the corresponding end-of-cycle pH. These examples establish emergence within the model, not the mechanism of an experimental trajectory.

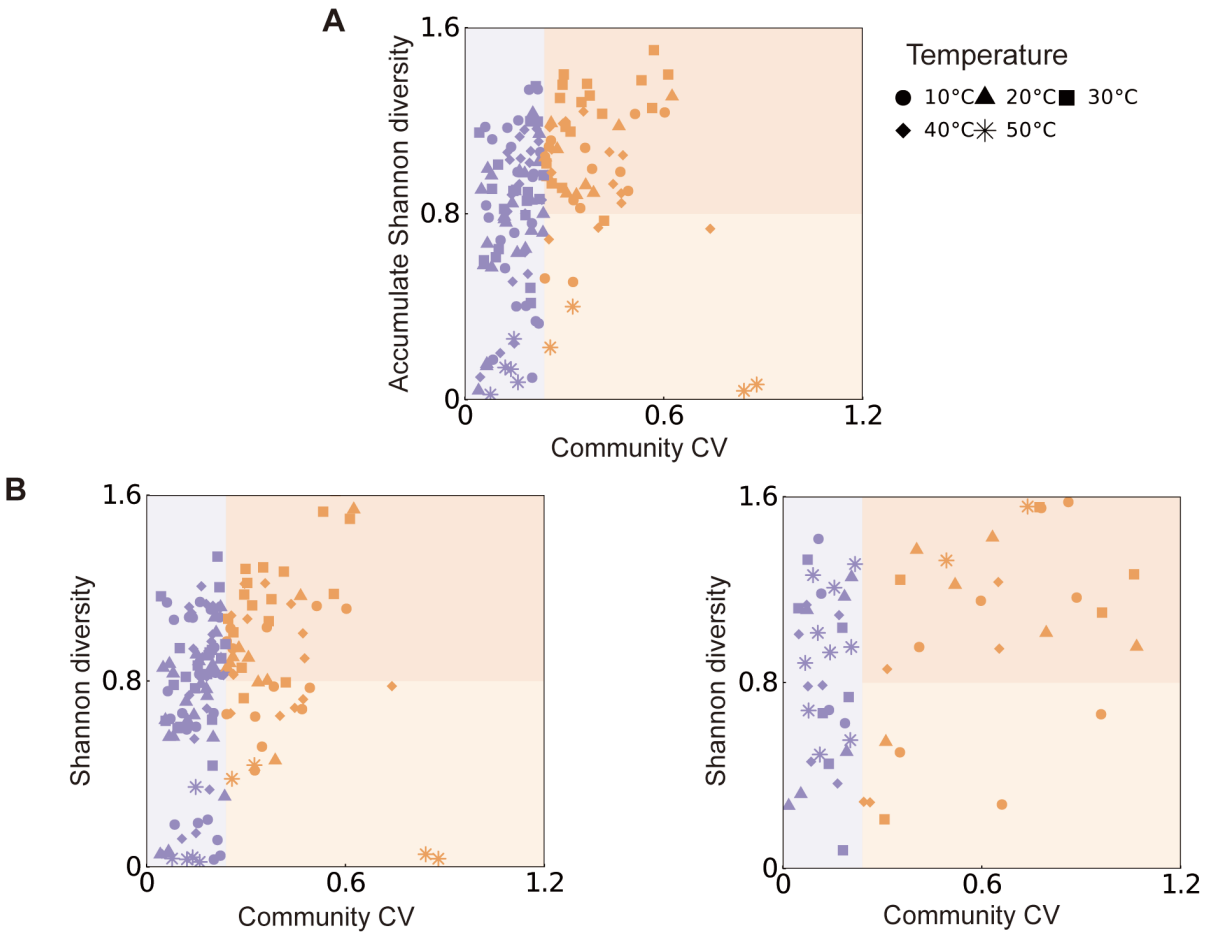

Fig. S20. Descriptive organization of temperature trajectories in the accumulated-diversity versus population-instability plane and its sensitivity to alternative diversity summaries or analysis subsets. Marker shape denotes temperature. Background regions are visual guides based on population instability (CV) = 0.25 and diversity = 0.8 and are not clustering results.