

Modeling convergent scale-by-scale skin color patterning in multiple species of lizards

Highlights

- Multiple lizard species convergently evolved dynamics of scale-by-scale color change
- Mesoscopic models predict pattern statistics without cellular and molecular variables
- Turing models generate robust predictions of individual scale-by-scale patterns
- Uncertainties in skin color and geometry explain patterns' residual unpredictability

Authors

Ebrahim Jahanbakhsh,
Michel C. Milinkovitch

Correspondence

michel.milinkovitch@unige.ch

In brief

Jahanbakhsh and Milinkovitch analyze convergent scale-by-scale skin color patterning dynamics in five divergent species of lizards and show that reaction-diffusion models robustly yield scale-by-scale predictability of individual adult patterns without the need to parametrize the system down to its many cellular and molecular variables.



Article

Modeling convergent scale-by-scale skin color patterning in multiple species of lizards

Ebrahim Jahanbakhsh^{1,2} and Michel C. Milinkovitch^{1,2,3,4,*}¹Laboratory of Artificial & Natural Evolution (LANE), Department of Genetics & Evolution, University of Geneva, 1211 Geneva, Switzerland²SIB Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland³Twitter: @LANEVOL⁴Lead contact*Correspondence: michel.milinkovitch@unige.ch<https://doi.org/10.1016/j.cub.2022.10.044>**SUMMARY**

Skin color patterning in vertebrates emerges at the macroscale from microscopic cell-cell interactions among chromatophores. Taking advantage of the convergent scale-by-scale skin color patterning dynamics in five divergent species of lizards, we quantify the respective efficiencies of stochastic (Lenz-Ising and cellular automata, sCA) and deterministic reaction-diffusion (RD) models to predict individual patterns and their statistical attributes. First, we show that all models capture the underlying microscopic system well enough to predict, with similar efficiencies, neighborhood statistics of adult patterns. Second, we show that RD robustly generates, in all species, a substantial gain in scale-by-scale predictability of individual adult patterns without the need to parametrize the system down to its many cellular and molecular variables. Third, using 3D numerical simulations and Lyapunov spectrum analyses, we quantitatively demonstrate that, given the non-linearity of the dynamical system, uncertainties in color measurements at the juvenile stage and in skin geometry variation explain most, if not all, of the residual unpredictability of adult individual scale-by-scale patterns. We suggest that the efficiency of RD is due to its intrinsic ability to exploit mesoscopic information such as continuous scale colors and the relations among growth, scales geometries, and the pattern length scale. Our results indicate that convergent evolution of CA patterning dynamics, leading to dissimilar macroscopic patterns in different species, is facilitated by their spontaneous emergence under a large range of RD parameters, as long as a Turing instability occurs in a skin domain with periodic thickness.

INTRODUCTION

Skin color patterning in vertebrates, as extensively studied in zebrafish,^{1–7} is a self-organized process, i.e., the spatial segregation of chromatophores observed during development does not follow pre-patterned positional information but autonomously emerges from cell-cell interactions among the chromatophores themselves.^{1,8} Because some of these interactions occur at short ranges and others at long ranges,^{1,6} the process can be efficiently described in the reaction-diffusion (RD) mathematical framework.^{1,8–14} In addition, we have demonstrated^{12,14} that, in the ocellated lizard (*Timon lepidus*), reduction of skin thickness at the borders of skin scales robustly transforms^{12,14} the RD color patterning process into a stochastic cellular automaton (sCA) where the neighborhood of monochromatic (either green or black) skin scales defines their color-flipping probabilities. Furthermore, we have shown¹⁵ that the 16-parameter sCA model (8 parameters for the color-flipping probabilities of green scales with 0–7 green neighbors, and 8 equivalent parameters for black scales) can be mapped to a simpler 2-parameter Lenz-Ising model^{16,17} (developed in the 1920s for describing the behavior of ferromagnetic materials),

which produces the observed steady-state neighborhood distribution of ocellated lizard adult patterns.

However, it remains unclear if the deterministic and continuous-state RD framework allows capturing features of the scale-by-scale skin color patterning process beyond those efficiently described with stochastic and discrete-state sCA and Lenz-Ising models. First, using a deterministic RD model might seem counter-intuitive because reactions and diffusion can be argued to be stochastic processes at the microscopic scale. However, RD is a continuous model that describes how components' concentrations at the mesoscopic/macroscopic scales, i.e., averages of molecule/cell densities at the nanoscopic/microscopic scales, deterministically vary in time and space. This is similar to deterministic thermodynamical properties emerging at the macroscale from statistical mechanical properties of microscopic entities. Hence, if initial components' concentrations are identical, RD simulations will always produce the same single mesoscopic trajectory. Second, as the RD,^{1,8–12,14} sCA, and Lenz/Ising models used to describe skin color patterning all equally ignore the unidentified or unmeasured underlying molecular and cellular variables (generally ill-defined as “noise”), it can be argued that the former has no



a priori reason to better capture, than the two latter, the dynamics of the system. Still, we reason here that RD might provide an intrinsically better description of the scale-by-scale skin color patterning process because it implicitly integrates a relation between the pattern length scale and geometrical parameters (such as scale geometries and their growth), and it allows to exploit the continuous-state distribution of scale colors (especially important at initial condition). To test this conjecture, we investigate here if deterministic RD allows describing individual patterning trajectories beyond their statistical properties, i.e., can RD efficiently predict, in multiple species, the positions of black and green/yellow scales in individual adult lizards despite that underlying unknown cellular and molecular variables are ignored?

We consider RD models that involve time, state, and spatial discretization at different levels: (1) a discrete RD model in two dimensions (2D-dRD), which is discrete in space but continuous in state and time, formally derived from Turing's spatially continuous RD equations by renormalizing the diffusion term;¹² (2) a 2D continuous RD model (2D-cRD), which is continuous in space, state, and time; and (3) a bona fide tridimensional cRD model (3D-cRD¹⁴). We have shown^{12,14,15} that all these RD models compute dynamics that are either constitutionally (dRD) or effectively (2D cRD and 3D cRD) discretized in space (i.e., scales are essentially monochromatic at any time point) and near discretized in state and time (i.e., color switching, between green and black, is not instantaneous but is much faster than the overall patterning process). These discretizations make the dynamics particularly suitable to unambiguous quantitative investigation: as the number of scales and the topology of the lattice are invariants throughout the life of a lizard, one can unequivocally identify the positions of black scales and green/yellow scales at any given time point during the post-hatching development of the animal.

Here, we investigate whether these various stochastic and RD models of patterning can be generalized to five species, belonging to five divergent lineages, exhibiting largely different adult patterns (Figure 1A), but that all convergently evolved dynamics of post-hatching scale-by-scale color change: the ocellated lizard (*T. lepidus*), the Argentine black and white tegu (*Salvator merianae*), the Gila monster (*Heloderma suspectum*), the mangrove monitor (*Varanus indicus*), and the Standing's day gecko (*Phelsuma standingi*). First, we show that all models predict neighborhood statistics of adult patterns with similar efficiencies. Second, we show that RD models substantially improve predictability of actual patterns beyond their statistical features without the need to parametrize the system down to its many cellular and molecular variables.¹⁸ More specifically, we show, in all five species, that continuous-state color information (i.e., estimates of RD-component concentrations based on the observed "greenness"/"blackness" of scales) from real juvenile lizards provides a substantial gain in scale-by-scale predictability of their corresponding adult patterns by RD over the stochastic Lenz-Ising and CA models. Third, we show that a substantial proportion of the residual unpredictability is due to heterogeneity in the variation of skin thickness among scale borders and among scales. Fourth, using Lyapunov spectrum analyses, we show that the uncertainty in color measurement in our experimental setup is sufficient to explain most, if not all, of the

remaining residual unpredictability due to the non-linear dynamical nature of the system.

RESULTS

Acquisition of lizard skin-scale color texture

Using a robotic system¹⁹ implementing the photometric stereo (PS) approach,²⁰ the surface microgeometry (normal map) and color texture (RGB color albedo) of a large dorsal skin patch (Figure 1B) were acquired fourteen to forty-five times over a period of 2–4 years (starting from the juvenile stage) in individuals of five divergent species of lizards (Figure 1A): two ocellated lizards (TL1 and TL2, about 2,000 scanned scales each), two Argentine tegus (SM1 and SM2, about 1,600 scales each), four Gila monsters (HS1, HS2, HS3, and HS4, about 900 scales each), two mangrove monitors (VI1 and VI2, about 1,400 scales each), and two Standing's day geckos (PS1 and PS2, about 3,000 scales each). Scale boundaries were identified as the edges of the Voronoi diagram (Figure 1C) partitioned using the scale centers defined by the watershed algorithm applied on the filtered curvature field derived from the initial normal map. This procedure both identifies scales and determines their neighborhood connectivity. Matching of scales across time points was performed by using local affine transformations producing a space-time network of scales (Figure 1D) with associated colors (albedo was averaged among pixels within each corresponding skin scale). The space-time network was used for finding optimal parameters in all stochastic and RD models. Additional details on animals, image acquisition, data processing, scale detection, scale matching, and iterative addition model parameter optimization are provided in the STAR Methods and in Tables S1 and S2.

Prediction of neighborhood statistics

Stochastic cellular automaton

Our analyses of time series of ocellated lizard scale-color switching previously indicated¹² that these dynamics can be effectively described as a stochastic cellular automaton (sCA), i.e., the stochastic automaton dynamically computes the adult labyrinthine pattern.¹² Indeed, the probability of color change of a green/black scale at each time-discrete iteration of the sCA is a function of $n_{g/b}$, i.e., the number of green/black direct neighbors. In real lizards, any black or green scale can have 0–7 isochromatic neighbors, such that the sCA model can be defined with $2 \times 8 = 16$ parameters. Here, we refine this model by using Bayesian inference to derive $\mathbb{E}[p]$, the expectation of the sCA probabilities (STAR Methods; Figure S1A), which are then used as CA probabilistic "rules" to simulate pattern time evolution. Each simulation is iterated until the number of scale-color flips reaches about the total number of flips observed in the real animal (between juvenile and adult stages); note that a perfect correspondence is usually not observed because one iteration of the sCA involves the color switch of more than one scale.

We then formulate an objective measure of how different the neighborhood statistics of each simulated pattern are from those of a real adult lizard. To this end, we simply calculate the normalized amplitude (E_{16D}) of the combined differences of simulated versus observed numbers of black and green scales with each of the 16 possible configurations of isochromatic neighbors.

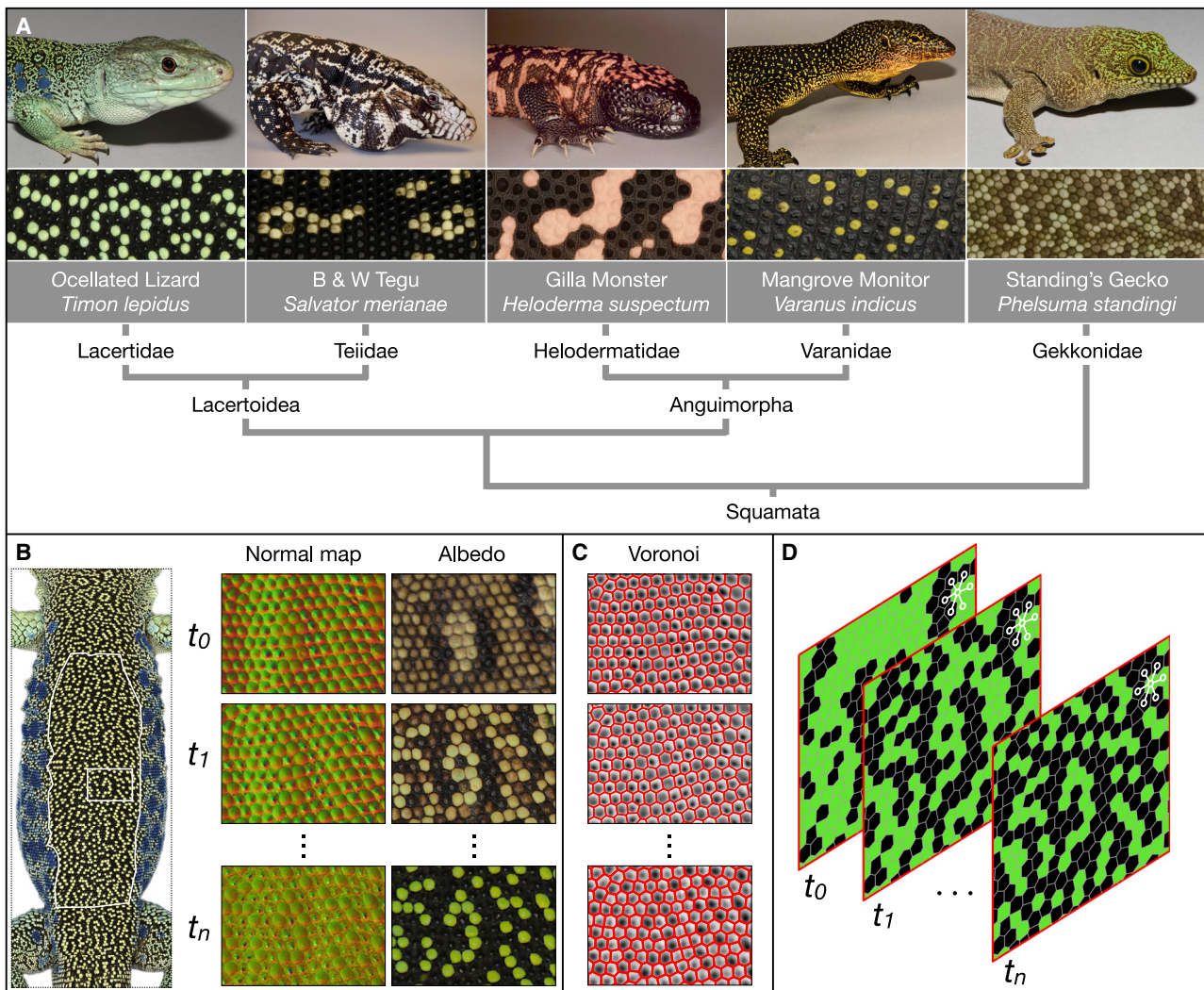


Figure 1. Acquisition of skin-scale color texture in five lizard species

(A) The five species investigated here (top, image of an adult individual; middle, close up on the adult pattern) belong to divergent squamate lineages (bottom, phylogenetic cladogram).

(B) A large dorsal skin patch (white outline in left panel) is photographed from different orientations and under different directions of incident illumination to generate high-resolution microgeometry (normal map, central panels, is shown for the small rectangular region framed in the left panel) and color texture (albedo, right panels) at multiple time points t_i . The individual shown in the left panel is TL1.

(C) The filtered curvature field (shading) allows identifying scale centers that are then used for building a Voronoi diagram, identifying scale boundaries (red lines), and scale lattice connectivity.

(D) Matching of scales across time points produces a space-time network of scales; white lines show one scale with one of its six neighbors shifting color from green to black between time points t_1 and t_n .

See also [Figure S6](#).

More technically, we compute, for each simulated adult pattern, the L^2 -norm of the 16-dimensional difference error vector (E_{16D}) of nearest-neighbor configuration statistics in comparison with the single observed adult pattern:

where n_s is the number of scales whereas $n(\text{Green}, i)$ and $n(\text{Black}, i)$ are, respectively, the number of green and black scales with i isochromatic neighbors; superscripts ^{sim} and ^{obs} denote the simulation and observation, respectively.

$$E_{16D} = \frac{1}{n_s} \sqrt{\sum_{i=0}^7 (n^{\text{sim}}(\text{Green}, i) - n^{\text{obs}}(\text{Green}, i))^2 + (n^{\text{sim}}(\text{Black}, i) - n^{\text{obs}}(\text{Black}, i))^2}$$

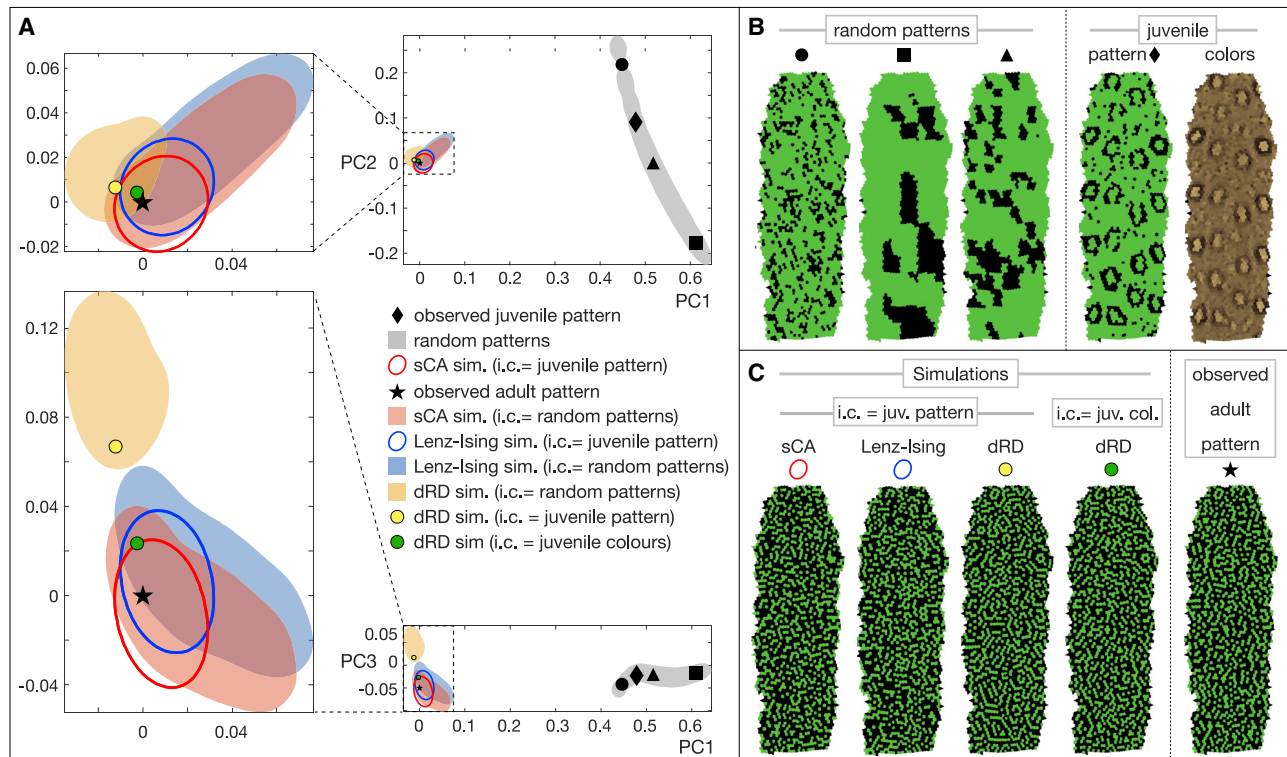


Figure 2. Prediction of neighborhood statistics in ocellated lizard individual TL1

(A) Projections on the principal component planes PC1-PC2 (top) and PC1-PC3 (bottom) of the 16D nearest-neighbor error vectors (in comparison with the observed adult pattern, black star) of patterns simulated with sCA (red ellipse and red shading), Lenz-Ising (blue), and dRD (yellow) models; interactive 3D graphs in PC1-PC2-PC3 space available as [Data S1](#). Red ellipse, blue ellipse, and yellow spot show adult patterns simulated from the observed juvenile pattern (black diamond), whereas red, blue, and yellow shadings show adult patterns simulated from random patterns (gray area) as initial condition. The green spot shows the adult dRD-pattern simulated from the juvenile colors shown in (B) (rightmost panel). Ellipses and border of shadings indicate 1% density isolines. (B) Different initial conditions used for simulations; their localization in PC1-PC2-PC3 space is shown with the corresponding geometrical symbols. (C) Adult patterns simulated with different initial conditions (i.c.): juvenile (juv.) pattern (=scale colors thresholded to green or black) and juvenile colors (col.) are both shown in (B). Similar results are obtained for individual TL2 ([Figure S2](#); [Data S2](#)). See also [Figure S1](#) and [Table S4](#).

We then use principal component analysis (PCA; [Figure 2A](#)) to identify the 3 largest modes (PC1, PC2, and PC3, corresponding to the 3 eigenvectors with largest eigenvalues) that are jointly capturing >90% of the variance of all error vectors. Performing 5,000 simulations, each starting from the real observed juvenile pattern (black diamond in [Figures 2A](#) and [2B](#)), indicates that the sCA model is robust: multiple runs starting from this same initial condition evolve into a population of simulated adult patterns (one example is shown in [Figure 2C](#)) whose E_{16D} values are restricted to an ellipsoid (red ellipses in [Figure 2A](#); mean $E_{16D} \pm SD = 0.037 \pm 0.012$) that also contains the zero-error coordinate (0,0,0) of the PC1-PC2-PC3 space, i.e., the position of the observed adult pattern (black star in [Figures 2A](#) and [2C](#)). We then produce 5,000 random patterns that span all length scales (from one skin scale to about one-fourth of the length of the analyzed dorsal skin patch) and are characterized by a large range of initial E_{16D} values (gray area in [Figure 2A](#)). Remarkably, the 5,000 simulations initiated from these random patterns (one simulation for each random pattern) show that the very large spread of initial conditions evolves into a population of patterns that are restricted to an area (red shading in [Figure 2A](#); mean $E_{16D} \pm SD = 0.065 \pm 0.05$) slightly larger than, but that strongly

overlaps with, the patterns (red ellipses) simulated from the juvenile pattern. Very similar results are obtained for the second ocellated lizard individual analyzed here (TL2; [Figure S2](#)).

Lenz-Ising model

The Lenz-Ising model^{16,17,21} is a statistical mechanical model developed for describing the states of magnetic materials as a lattice of sites, each with one of two possible orientations (+1 or -1) of electronic/nucleic magnetic dipole moments. The model requires only two parameters: J , which defines how much neighbors with the same orientation are favored ($J > 0$, ferromagnetic materials) or disfavored ($J < 0$, antiferromagnetic materials), and B , which correspond to an “external” magnetic field favoring one or the other state (depending on the sign of B). We have recently shown¹⁵ that the time evolution of the Lenz-Ising model in thermal equilibrium effectively describes the dynamics of scale-by-scale skin color patterning of the ocellated lizard when we treat green and black scales as if they were +1 and -1 dipoles, respectively. More technically, the late-time probability distribution of patterns is equivalent to the canonical probability distribution of the antiferromagnetic Ising model at finite temperature ($1/k_B T > 0$) with black scales being favored.¹⁵ Hence, the Lenz-Ising model is the simplest

formalism (2 parameters instead of the sCA 16 parameters) describing the skin color pattern self-organization in the ocellated lizard.¹⁵

Similar to the sCA analyses above, we perform simulations using the Lenz-Ising model with βB and βJ parameters and dynamics,¹⁵ optimized for each individual (Figures S1B and S2B). The 5,000 Lenz-Ising simulations starting from the real juvenile pattern (black diamond) and the 5,000 simulations starting each from one of the random patterns (gray area) generate populations of E_{16D} nearest-neighbor errors (Figure 2A, blue ellipses with mean $E_{16D} \pm SD = 0.048 \pm 0.016$ and blue area with mean $E_{16D} \pm SD = 0.091 \pm 0.058$, respectively) very similar to those obtained with the sCA model (red ellipses and red shading). These results indicate that, for ocellated lizards, the sCA and Lenz-Ising models exhibit similar performances in robustly predicting the nearest-neighbor statistics, hence the general labyrinthine look, of real adult patterns (Figure 2C).

Optimized reaction-diffusion models

RD models, first introduced by Turing,⁹ compute the concentrations of diffusing and reacting molecular species (RD components) through time and space. In the framework of skin color patterning, “diffusion” should be understood as “effective diffusion”²² of signals originating from short- and long-range cell-cell contacts.¹ Note that the long-range interactions involve macrophage-like cells⁶ that recognize and drag plasma membrane blebs on xanthoblasts, generating and extending thin cellular projections (airinemes), before performing a random walk and ultimately depositing the airineme vesicle on the surface of a melanophore at a distance up to several xanthoblast-body lengths. It is likely that these cell-cell interactions have the capacity of generating long-ranged gradients (i.e., larger than the typical 1 mm skin scale of an adult ocellated lizard).

Here, we use a three-component (u, v, w) 2D continuous RD (cRD) model, initially developed to study skin color patterning in zebrafish.¹ The components u and v represent the densities of melanophores and xanthophores (and their corresponding short-range factors), respectively, whereas w represents a long-range factor with diffusion coefficient D_w much larger than $D_u = D_v$. We previously adapted this model for describing skin-scale color change dynamics in the ocellated lizard¹² and extended it to 3D¹⁴ to take into account variation of the domain thickness. Here, we develop a new implementation of the 3D-cRD model in which skin geometry is represented as a curvilinear 3D grid (more accurate than the rectilinear grid used in Fofonjka and Milinkovitch¹⁴) yielding lower errors in computing components’ concentrations, hence allowing for the use of more efficient coarse-grain grids (STAR Methods).

Simpler modeling of the scaled skin color patterning process in 2D (i.e., in the plane of the skin) rather than in 3D can be achieved¹² by taking into account the reduction of skin thickness at scale borders through scaling of the continuous RD diffusion coefficients (by a factor P) at the one-dimensional edges of the 2D scales. Beside this 2D continuous model (2D-cRD), we also introduced a discrete RD model (2D-dRD) where each scale is represented by a single node.¹² This discretization is justified by the 3D geometry of the skin: as scale skin is much thicker than interscale skin, the concentrations of morphogens within a given skin scale tend to spatially homogenize, hence, each of these concentrations can be represented by a single value (i.e.,

the average concentration within the corresponding skin scale). This assumption is validated by bona fide 3D-RD simulations¹⁴ indicating that scales tend to effectively change color as a whole entity, rather than exhibit sustained gradients. We have previously shown how to renormalize the diffusion term when deriving the 2D-cRD model for regular hexagonal lattices.¹² Here, we generalize the 2D-dRD model by deriving the discrete RD equations for arbitrary polygons in order to simulate the ocellated lizard skin color change dynamics on realistic (non-strictly hexagonal) lattices of skin scales. Details on the derivations of the 2D-cRD and 2D-dRD models are provided in the STAR Methods.

Furthermore, we suggest here improvements of the 2D-cRD and 2D-dRD models to increase their predictive power by (1) integrating growth of the domain in which patterning is taking place, (2) considering continuous variation of color states, and (3) optimizing RD parameters using Bayesian machine-learning global minimization based on a Gaussian process regression approach. To this end, we first assume that scales grow isotropically following a logistic function (parameters fitted using measurements from real lizards) that we then integrate in the RD equations. Second, using linear stability analysis close to the homogeneous steady state (HSS), we develop transformation functions to transfer juvenile scale CIELAB color information to the space of RD variables and use PCA to perform transformations, at later (post-juvenile) time points, from the u, v, w RD space to the CIELAB color space. Additional details on the growth-integrated RD models, transformation functions, linear stability analysis, PCA, and Bayesian optimization are given in the STAR Methods.

We then run 2D-dRD simulations from the same initial conditions as for the sCA and Lenz-Ising simulations. The 5,000 2D-dRD simulations starting each from one of the random patterns (gray area in Figure 2A) generate a population of patterns exhibiting error vectors (orange shading; mean $E_{16D} \pm SD = 0.11 \pm 0.029$) similar to those obtained with the stochastic models (sCA and Lenz-Ising; red and blue shadings, respectively). When using the same single observed juvenile pattern (black diamond; Figure 2A) as that used for sCA and Lenz-Ising simulations (i.e., with scale colors thresholded to green or black by applying K-mean clustering), the resulting dRD-simulated adult pattern (yellow spot; Figure 2C) exhibits a nearest-neighbor error vector ($E_{16D} = 0.074$; yellow spot; Figure 2A) situated in the same region of the PCA graph. However, one of the interests of RD is the possibility to exploit the information included in the continuous-state values of scales at the observed initial condition: the dRD simulation starting from these real colors of the juvenile pattern (top right image in Figure 2B), instead of a thresholded initial pattern with only two states, produces a pattern (green spot in Figures 2A and 2C; $E_{16D} = 0.033$) that is closer to the observed adult pattern. Note that sCA and Lenz-Ising simulations cannot be performed with juvenile continuous color values as initial condition because these are intrinsically two-state models. Very similar results are obtained for the second ocellated lizard individual (TL2; Figure S2C).

Comparable neighborhood predictions of stochastic and RD models

The analyses above indicate that, for ocellated lizards, simulations initiated from the observed juvenile pattern, and run

using any of the above models, generate simulated adult patterns with neighborhood statistics, hence, qualitative looks (Figures 2C and S2E), remarkably similar to those of the real adult pattern. In other words, although RD can exploit the information associated with the continuous color values observed in juveniles, it does not perform spectacularly better than sCA and Lenz-Ising two-state models for predicting the neighborhood statistics of adult patterns.

Prediction of neighborhood statistics in other species

Our analyses indicate that both stochastic models (sCA and Lenz-Ising) are robust in all four additional species in predicting neighborhood statistics (Figure 3): performing 5,000 simulations, each starting from the real observed juvenile pattern (black diamond), generates a population of simulated adult patterns that are restricted to zones (Figure 3A, red and blue ellipses for sCA and Lenz-Ising, respectively) close to the zero-error coordinate (0,0) of the PC1-PC2 plane, i.e., close to the position of the observed adult pattern (black star). Note, however, that different species exhibit different lengths of E_{16D} trajectory between juvenile and adult patterns because juveniles are closer to the HSS in some species (e.g., ocellated lizard) than in others (e.g., Gilla monster and mangrove monitor), such that they differ in their typical number of post-hatching scale-color flips. Regarding Lenz-Ising optimization, our analyses indicate that the model is antiferromagnetic for the gecko (as in the ocellated lizard), i.e., favoring labyrinthine patterns in which direct neighbors tend to exhibit opposite colors, and ferromagnetic for the tegu, i.e., favoring the existence of larger areas of isochromatic scales, whereas it is virtually on the boundary between ferromagnetic and antiferromagnetic in the Gila monster and mangrove monitor (Figure S3A).

As with the ocellated lizard (Figure 2), the dRD simulations initiated from either the thresholded yellow/orange versus black juvenile pattern or the real juvenile colors (Figure 3B, first and second column of images, respectively) produce adult patterns exhibiting a low nearest-neighbor error (yellow and green spots, respectively), indicating that RD is an excellent neighborhood configuration predictor in all species analyzed here, despite the substantially different motifs of their patterns. Also similar to the ocellated lizard, the error vectors generated in the four additional species with 2D-dRD simulations (orange areas in Figure 3A) starting from 5,000 random patterns are not particularly better than those obtained with the sCA and Lenz-Ising models (red and blue areas, respectively). Very similar results are obtained for the additional individuals analyzed here for the four species (SM2, PS2, HS2, HS3, HS4, and VI2; Figure S3B).

Prediction of individual scale-by-scale patterns

We define for all models the “scale-by-scale error” between a simulated and observed pattern as

$$E_{\text{sbs}}^{(k)} = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{\|\mathbf{C}_i^{\text{sim}}(\tau_k) - \mathbf{C}_i^{\text{obs}}(\tau_k)\|}{\|\mathbf{C}_b(\tau_k) - \mathbf{C}_g(\tau_k)\|}$$

where n_s is the number of scales; \mathbf{C}_i denotes the color of the i^{th} scale; superscripts sim and obs denote the simulation and observation, respectively; τ_k denotes the time point; and \mathbf{C}_g and \mathbf{C}_b are the green and black state colors, respectively. For dRD analyses, \mathbf{C}_i are computed according to Equation 31 (STAR Methods), whereas \mathbf{C}_g and \mathbf{C}_b are mean values computed by K-mean

clustering in which scale colors are defined in CIELAB coordinates. For sCA and Ising simulations, \mathbf{C}_i , \mathbf{C}_g , and \mathbf{C}_b are all replaced by the scalars representing the binary (“green” or “black”) state of scales.

Stochastic versus deterministic models

Stochastic models (sCA and Lenz-Ising) are susceptible to generating substantial errors in the prediction of individual scale-by-scale patterns despite the fact that they excel in predicting the bulk statistics inferred from them. Indeed, given enough iterations of the sCA process or long enough thermalization of the Lenz-Ising model, the exact positions of green and black scales on a simulated adult pattern can, in principle, greatly differ from those observed on a real pattern, even if the neighborhood distributions between simulated and observed patterns are highly similar. We confirm this prediction by showing that sCA-simulated or Lenz-Ising-simulated patterns selected for exhibiting very low nearest-neighbor error vectors in the PC1-PC2-PC3 space (red and blue dots in Figure 4A, left) do not show E_{sbs} values particularly smaller than those observed for simulated patterns with larger E_{sbs} errors (Figure 4A, ellipses in left panel and distributions in central panel). This result is confirmed by the absence of correlation between sCA- or Lenz-Ising-associated E_{16D} nearest-neighbor errors and their corresponding E_{sbs} scale-by-scale errors (Figure 4A, right).

On the other hand, RD models could, in principle, exhibit increased precision in predicting actual scale-by-scale patterns (and not only the bulk statistics inferred from them) for three reasons: (1) RD is deterministic; (2) RD integrates an implicit relation between, on one hand, the pattern typical length scale, and on the other hand, growth and the geometry of skin scales (i.e., their sizes as well as their numbers and lengths of edges); and (3) RD can exploit the information provided by the actual (u, v, w) RD-component concentrations at initial condition. Note that the (u, v, w) values can be estimated from the observed “greenness”/“blackness” of scales in real juvenile lizards using linear stability analysis and transformation functions (STAR Methods). Figures 4A–4C indicate that this conjecture of substantial gain in scale-by-scale predictability of observed adult patterns by RD over stochastic models is correct: starting from the same juvenile K-mean thresholded (green and black) pattern as for the stochastic models, RD generates a simulated pattern (yellow spot in Figures 4A and 4B) with a scale-by-scale error ($E_{\text{sbs}} = 35.5\%$) smaller than those obtained with the stochastic models. Although this improvement is statistically significant ($p < 4.2 \times 10^{-4}$), its amplitude is somewhat small (Figure 4C). Conversely, the improvement of scale-by-scale predictability becomes much more prominent ($>15\%$, $p < 9.8 \times 10^{-57}$; green spot in Figure 4A) when using the juvenile unthresholded colors as initial condition. This result is unanticipated as neighborhood statistics, and *not* scale-by-scale error, were used as the objective function during Bayesian optimization of RD parameters. The same error reduction of about 15% is observed between stochastic models and the full RD analysis for the second ocellated lizard analyzed here (TL2; Figure S2F).

Residual unpredictability and skin thickness spatial distribution

Our experimental setup does not provide one key piece of information relevant to RD: the exact spatial distribution of skin thickness across the RD field. To quantify how much real skin

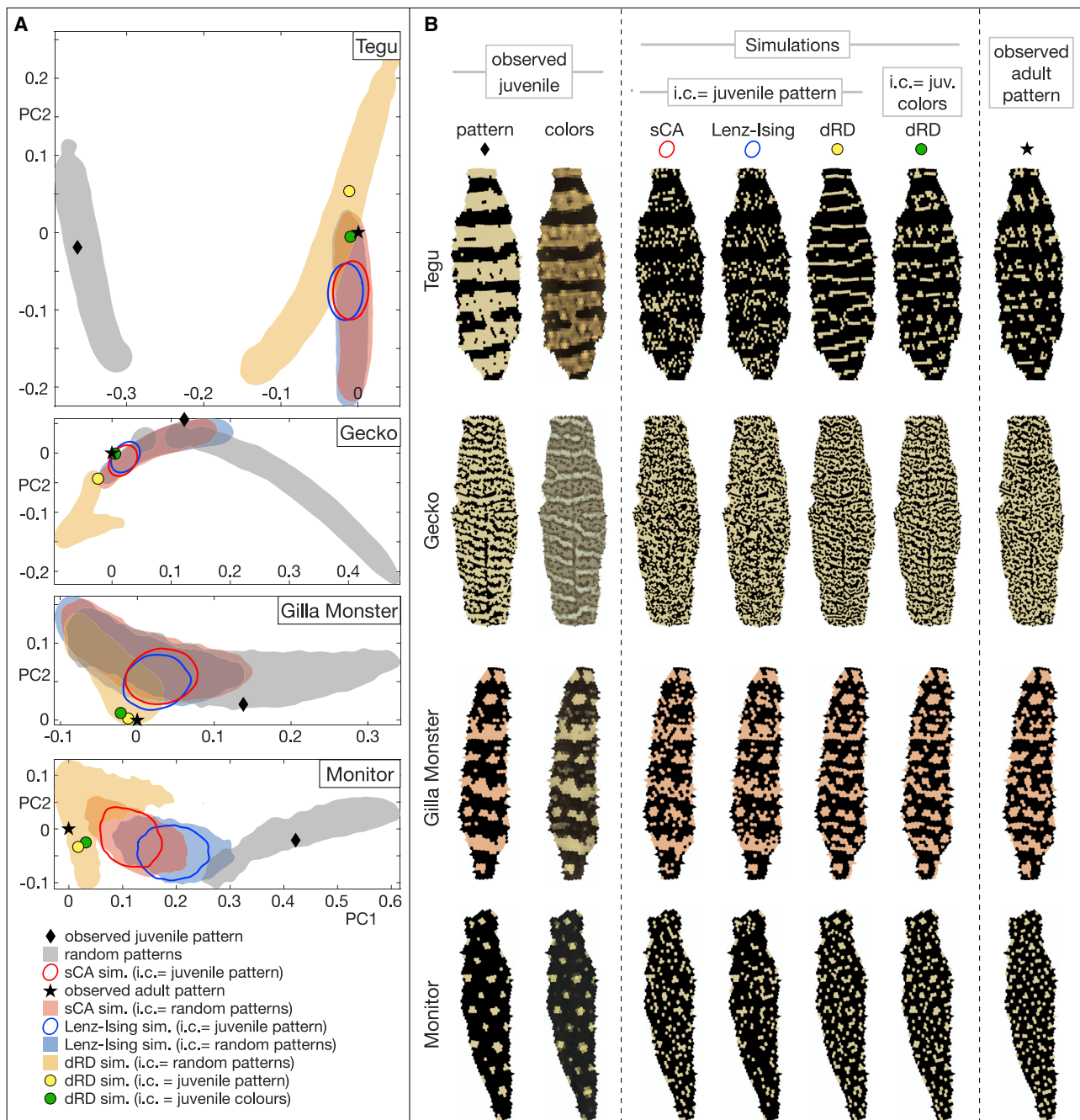


Figure 3. Prediction of neighborhood statistics in four additional species

(A) Projections on the PC1-PC2 plane of the 16D nearest-neighbor error vectors (in comparison with observed adult pattern, black stars) of patterns simulated with sCA (red ellipse and shading), Lenz-Ising (blue), and dRD (yellow) models in black and white tegu (SM1), Standing's gecko (PS1), Gila monster (HS1), and mangrove monitor (VI1). Red ellipses, blue ellipses, and yellow spots show adult patterns simulated from the observed juvenile pattern (black diamonds), whereas red, blue, and yellow shadings show patterns simulated from random patterns (gray areas) as initial condition. The green spots show the dRD-pattern simulated from the juvenile color shown in (B).

(B) Left: observed juvenile patterns and colors. Center: adult patterns simulated with juvenile (juv.) pattern (=scale colors thresholded to green or black) or juvenile colors as initial condition (i.c.). Right: observed adult patterns. Ellipses and border of shadings indicate 1% density isolines.

See also [Tables S3](#) and [S4](#) and [Figure S3](#).

geometry deviates from the assumption of identical skin thickness reduction at all scale borders, we perform here 3D geometry reconstruction of a patch of ocellated lizard dorsal skin

using high-resolution episcopic microscopy (HREM^{14,23}). [Figure 5A](#) shows a portion of the HREM-acquired skin 3D domain (black pixels correspond to melanophores) and the heatmap of

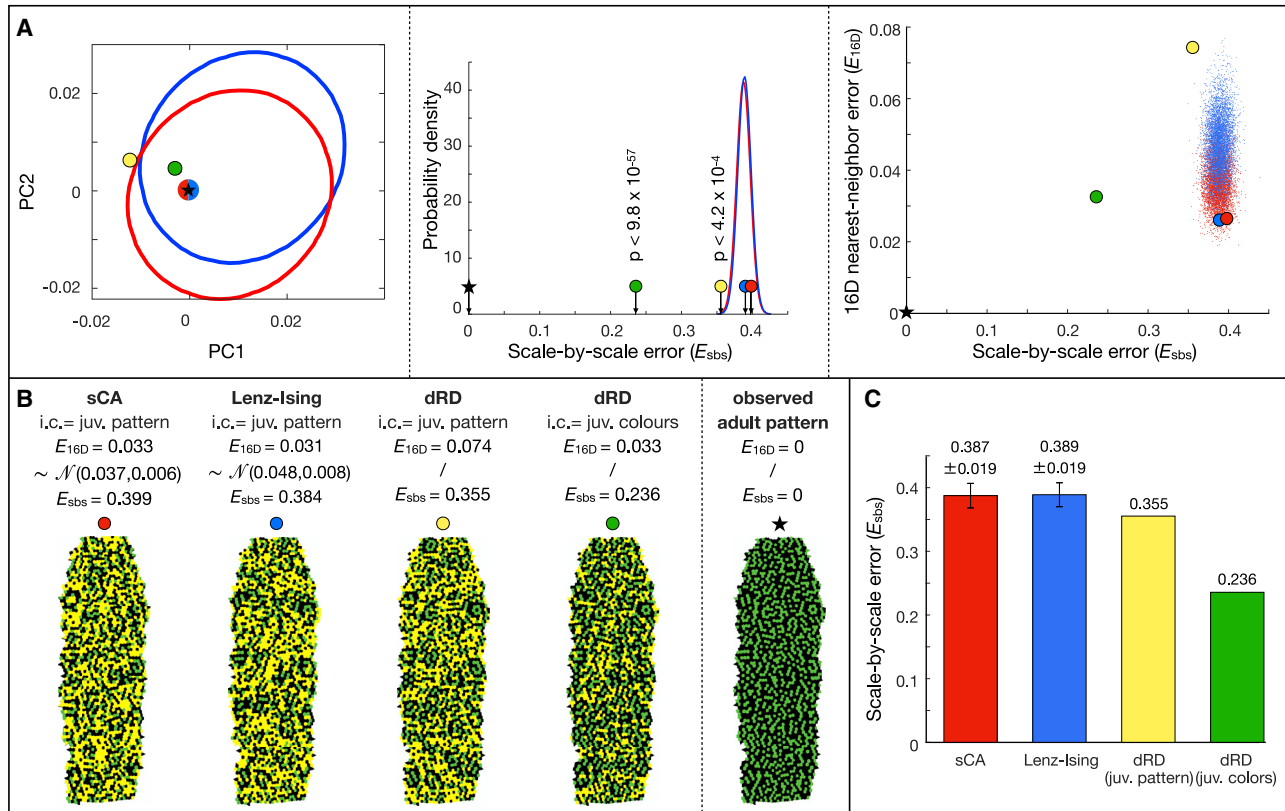


Figure 4. Prediction of scale-by-scale patterns in ocellated lizard individual TL1

(A) Simulated adult patterns with low neighborhood errors do not show particularly low scale-by-scale errors (E_{sbs}) within the distribution of simulated patterns: red and blue spots indicate zero-error patterns (in the PC1-PC2 plane, left) among the 5,000 simulated with sCA and Lenz-Ising, respectively. Ellipses in left panel are 1% density isolines of neighborhood error vectors whereas the Gaussian distributions in the central panel are the E_{sbs} probability density functions. The adult patterns simulated from the juvenile pattern or juvenile colors (yellow and green spots, respectively) exhibit statistically smaller E_{sbs} (p values shown in central panel) than the sCA and Lenz-Ising probability density functions. Right: no correlation is observed between neighborhood error and scale-by-scale error; red points and blue points are 5,000 simulations with sCA and Lenz-Ising, respectively.

(B) Simulated adult patterns (corresponding to red, blue, yellow, and green spots in A); scales whose color is erroneous (in comparison with observed adult, rightmost pattern) are indicated in yellow.

(C) Histogram (mean \pm SD) comparing scale-by-scale errors of adult patterns simulated with different models. Similar results are obtained for individual TL2 (Figure S2).

See also Figure S5.

black pixel positions (d_m) with respect to skin thickness (d). Whereas extraction of the spatial distribution of d is straightforward, the distribution of d_m is obtained by mapping d to the second-order polynomial fit (orange line in Figure 5A) of 75% deepest observed black pixel positions for all scales. From these HREM data, we also extract the mean distance S among neighboring scales at their highest point, as well as the means among the top-surface highest point of scales (\bar{h}_c) and among the heights of edges (\bar{h}_e).

We then construct a 3D lattice of super-Gaussian bumps for which we optimize parameters σ and p (cf. equations in Figure 5B), as well as the standard deviations σ_c and σ_e , of bump heights (h_c) and scale edge heights (h_e), respectively, to obtain a depth-map histogram (i.e., a skin thickness distribution) highly similar to that observed in the HREM data (Figure 5C). This lattice allows us to construct 1,000 noisy domains by sampling h_c (for each bump) and h_e (for each edge) from the normal distributions $\mathcal{N}(\bar{h}_c, \sigma_c)$ and $\mathcal{N}(\bar{h}_e, \sigma_e)$. Finally, using the

mapping $d_m(d)$ discussed above, we compute the bottom boundary of each domain by subtracting d_m from the top boundary, hence producing models of the skin domain restricted to chromatophores (turquoise volume in Figure 5B). As these 1,000 domains are all different, but they each exhibit thickness heterogeneities (among scale centers and among scale edges) similar to those of real lizards, we can use them as follows to test for their impact on the predictability of RD modeling. For each generated domain, we run a bona fide 3D-cRD numerical simulation (parameter values obtained from Bayesian optimization in TL1; see above) and compare the resulting adult pattern with that obtained with the simulation performed on a reference homogeneous 3D domain of 64 identical super-Gaussian bumps (i.e., with the centers of all bumps set to \bar{h}_c and all the edges set to \bar{h}_e). Note that the 1,000 simulations, and the reference simulation, are all started from the colors of a patch of 64 scales observed in TL1 at the juvenile stage. Figure 5D shows the distribution of scale-by-scale error

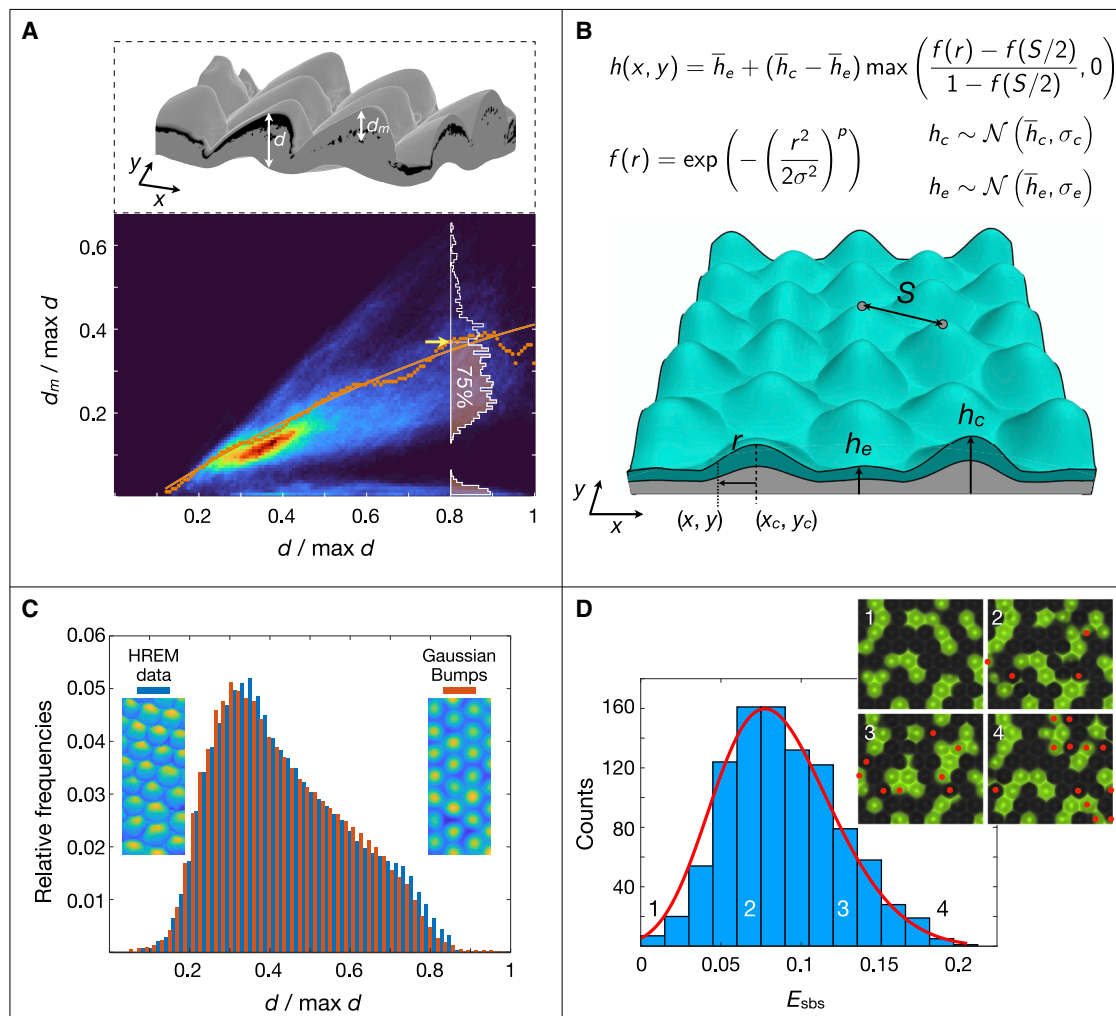


Figure 5. Heterogeneity in skin thickness variation affects RD predictability

(A) Top: ocellated lizard dorsal 3D skin patch reconstructed using HREM; d and d_m are, for each value of (x, y) , the skin thickness and black pixel lowest depth, respectively. Bottom: heatmap of normalized d_m versus normalized d ; orange dots indicate, for each value of $d/\max d$, the boundary (yellow arrow) at the 75th percentile in the corresponding distribution (white profile); orange line, second-order polynomial smooth fit across the orange dots.

(B) Example of an hexagonal 3D lattices of super-Gaussian bumps (mean distance among neighboring scales S is from HREM data; σ and p are Gaussian parameters in the equation on top) with h_c (height for each bump) and h_e (for each edge) are sampled from $\mathcal{N}(\bar{h}_c, \sigma_c)$ and $\mathcal{N}(\bar{h}_e, \sigma_e)$, where $\bar{h}_{c,e}$ and $\sigma_{c,e}$ are means and variances, respectively. The skin domain restricted to chromatophores (turquoise) is computed using the mapping $d_m(d)$; see text for details.

(C) Histogram of domain thickness of the 3D lattice shown in (B) after optimizing σ , p , $\bar{h}_{c,e}$, and $\sigma_{c,e}$ to obtain a profile (red) highly similar to the HREM data histogram (blue).

(D) Histogram of scale-by-scale error (in comparison with the pattern simulated with a reference homogeneous 3D domain of identical Gaussian bumps, i.e., $\sigma_c = \sigma_e = 0$) in adult patterns simulated on 1,000 heterogeneous 3D domains similar to the one shown in (B). The fitted generalized extreme-value distribution (red line) indicates that heterogeneity in skin thickness variation produces a mean $E_{sbs} \approx 8.5\%$. Inset shows examples of patterns generated with four heterogeneous domains (red dots indicate scales with wrong color, and numbers refer to their E_{sbs} value in the histogram).

associated to the 1,000 heterogeneous domains: the mean of the expected scale-by-scale error exclusively due to heterogeneity in skin thickness variation is $E_{sbs} = 8.5\%$. In other words, given the incomplete information on the thickness variation among scales and among scale borders, a hypothetical numerical model with perfect intrinsic predictability would, on average, still not allow predicting exact individual patterns in ocellated lizards with scale-by-scale error $< 8.5\%$. Details on HREM data acquisition and generation of noisy super-Gaussian 3D lattices are provided in the [STAR Methods](#).

Residual unpredictability and color measurement uncertainty

In order to generate Turing instabilities, the partial differential equations used in RD must be non-linear. In the model we use, the non-linearity resides in the piecewise form of the reaction terms (Equation 13). Non-linearity could make the dynamical system sensitive to initial conditions: given that the colors of the scales cannot be known with infinite precision, the corresponding initial uncertainty in measurement is likely to grow with time. However, given that it generates Turing instabilities,

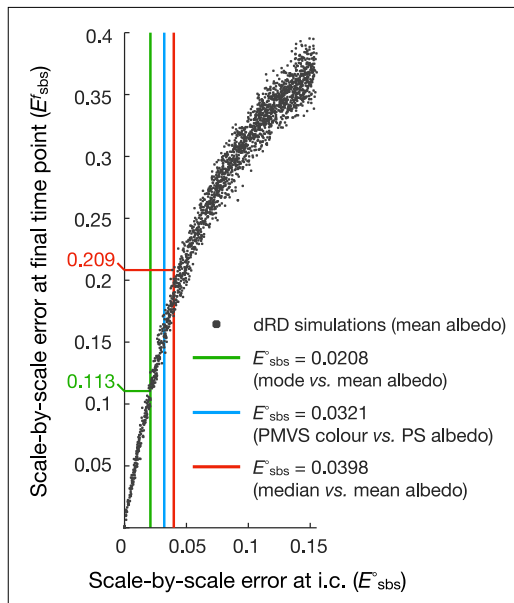


Figure 6. Scale-by-scale error due to juvenile color measurement uncertainty

Plot of the scale-by-scale error at the final (adult) time point versus error at initial condition (i.c.) for the 2,000 simulations shown in Figure S4 (Lyapunov spectrum analysis). At initial (juvenile) condition, median albedo (red line) or mode albedo (green) or mean RGB colors (blue) give differences of scale colors (against mean albedo) that generate 11.3%–20.9% scale-by-scale error at the adult stage.

the system we study must also be far from the chaotic regime where non-linearity would make it irreducibly unpredictable. As the dRD model exhibits a minimum predictability error of 23.6% in the analyses above, and about 8.5% of it is due to uncertainty regarding the spatial variation of skin thickness, what is the cause of the other 15.1%? Below, we show that most, if not all, of this residual unpredictability is due to uncertainty in color measurements at initial condition.

One powerful approach to estimate the degree of sensitivity of a dynamical system is to use Lyapunov spectrum analysis (STAR Methods), i.e., computing how fast two trajectories diverge when they initially differ by an arbitrary small difference. To this aim, we generate 2,000 “noisy initial conditions” by adding random noise to the colors of all scales of the observed juvenile lizard and study their time evolution as 2,000 trajectories in phase space. We then (1) measure the time evolution of the Euclidean distance $\delta_\tau = \|\mathbf{S}_\tau - \mathbf{S}_\tau^{\text{ref}}\|$ between the state of each perturbed system (\mathbf{S}_τ) and the state ($\mathbf{S}_\tau^{\text{ref}}$) of the reference (i.e., the latter is the trajectory starting from the observed unperturbed juvenile colors) and (2) use the Lyapunov exponent λ as an estimate of the rate of divergence between the two trajectories (Figure S4).

This analysis allows us to readily quantify the expected error at the adult stage given a specific uncertainty at the initial condition: we compute, for each of the 2,000 simulations, the scale-by-scale color errors, E_{sbs}^o and E_{sbs}^f , i.e., at the juvenile and adult stages, respectively. Note that these errors are computed by comparing each simulation (starting with noisy initial condition) with the reference simulation (starting with the observed juvenile colors). Figure 6 shows the variation of E_{sbs}^f with respect to E_{sbs}^o

across all 2,000 simulations: a larger initial error generates a larger error at the adult stage. To estimate the expected scale-by-scale error at the adult stage given the actual uncertainty in color measurements at the juvenile stage, we reflect that the color of each scale is measured as the mean albedo among pixels within that scale (STAR Methods). Although the procedure of using a single color for the whole scale is justified by the well-known behavior of the system^{12,14} (i.e., each of the (u, v, w) components tends to spatially homogenize within a scale), using the mean is arbitrary. Similarly, one could argue that using the mean RGB color observed on the pictures with homogeneous illumination is a more direct color measurement. Figure 6 indicates (vertical lines) that using the mode or the median (instead of the mean) of the PS albedo at initial condition would translate to an initial difference (E_{sbs}^o) of 2.08% and 3.98%, respectively, whereas the use of RGB colors would correspond to an intermediate E_{sbs}^o value of 3.21%. This range of uncertainties at initial condition generates a range of scale-by-scale errors in the adult (E_{sbs}^f) of 11.3% to 20.9%. Hence, the imprecision in color measurement of juveniles is sufficient to explain most, if not all, of the remaining residual unpredictability (15.1%) of the actual adult lizard patterns.

Residual unpredictability in other species

Below, we investigate whether these results in ocellated lizards can be generalized to the process generating largely different adult patterns in four other species (Figure 1A): the interrupted black banding of the Argentine black and white tegu, the low-contrast highly labyrinthine pattern of the Standing’s day gecko, the large black-and-orange meanders of the Gila monster, and the yellow-on-black speckles of the mangrove monitor. Remarkably, as in the ocellated lizard, RD modeling generates a substantial gain (over stochastic models) in scale-by-scale predictability of observed adult patterns in these four additional species. Indeed, starting from the juvenile thresholded (yellow/orange versus black) pattern, RD systematically generates simulated adult patterns (yellow spots in Figure 3) with a scale-by-scale error statistically smaller than those obtained with the stochastic models (Figure 7A, yellow bar). Similarly again to the ocellated lizard, the gain in predictability by the RD model substantially increases in amplitude and statistical significance when using the juvenile (un-thresholded) real colors as initial condition, leaving a residual scale-by-scale unpredictability (Figure 7A, green bar) of 14.2% for the black and white tegu, 22.3% for the Standing’s gecko, 12.7% for the Gila monster, and 19.1% for the mangrove monitor. Using 3D simulations with heterogeneous skin geometry variation (Figure 7B) and Lyapunov spectrum analyses (Figures 7C and 7D), we show that the residual error in all species is explained by a combination of uncertainty in skin thickness spatial distribution and in color measurement at initial condition: for the ocellated lizard, the mean of the scale-by-scale errors due to uncertainties in skin geometry variation and juvenile color measurement (Figures 7B and 7D) is 8.5% and 16.1%, summing up to 24.6%, i.e., close to the total residual error of 23.6% (Figure 7A). For the tegu, Standing’s gecko, Gila monster, and mangrove monitor, these numbers are 7.1% + 7.7% \approx 14.2%, 5.8% + 6.3% \approx 22.3%, 7.4% + 4.5% \approx 12.7%, and 3.1% + 16.2% \approx 19.1%, respectively. Only for Standing’s gecko, the means of geometry-induced and of color-induced errors do not approximately sum up to the observed

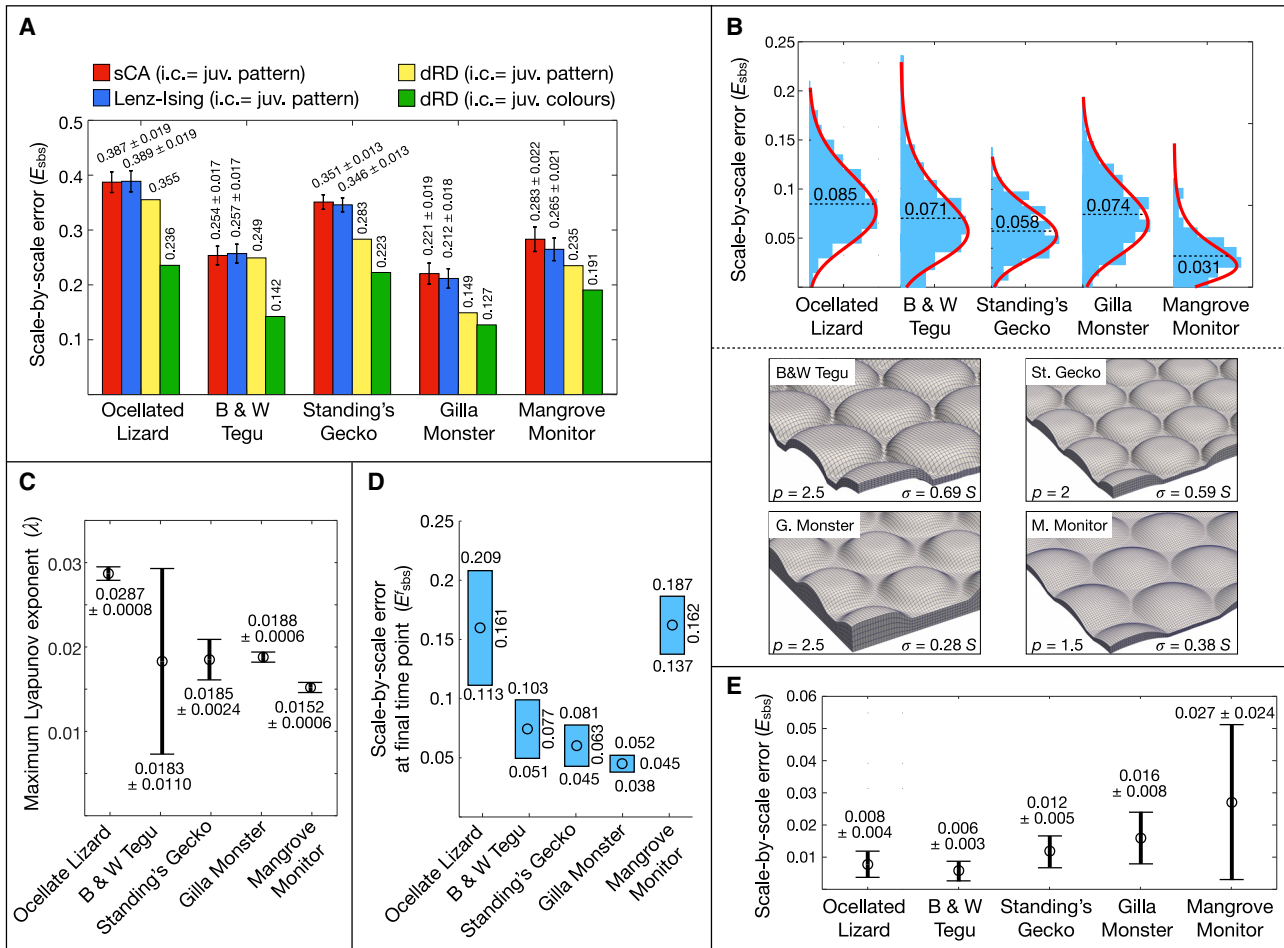


Figure 7. Residual unpredictability and measurement uncertainties

(A) Scale-by-scale errors (mean ± SD) of adult patterns simulated with different stochastic and RD models in five lizard species: green columns represent the residual unpredictability of the RD model with juvenile (juv.) colors at initial condition (i.c.).
 (B) Top: distribution and mean (dotted lines) of scale-by-scale error caused by heterogeneity in skin thickness variation (distributions generated as in Figure 5D) in five species of lizards. Lower panels show the reference hexagonal lattices of 3D super-Gaussian bumps with optimized p and σ ($p = 1$ and $\sigma = 0.28S$ for ocellated lizard). S = mean distance among neighboring scales on a real skin patch of the corresponding species. Heterogeneity in skin thickness variation generates a mean unpredictability of 8.5%, 7.1%, 5.8%, 7.4%, and 3.1% in ocellated lizard, tegu, gecko, Gila monster, and monitor, respectively.
 (C) Mean and SD of Lyapunov exponent (computed as in Figure S4) for five species of lizards.
 (D) Range (circles indicate middle of ranges) of scale-by-scale error (at adult time point) obtained with simulations initiated with juvenile colors differing by a small amount E_{abs}^* within the range $[\|C^{\text{mode}} - C^{\text{ref}}\|, \|C^{\text{median}} - C^{\text{ref}}\|]$, where C^{ref} , C^{mode} , and C^{median} are the mean, mode, and median albedos of juvenile colors. Uncertainties in juvenile color measurements generate a mean unpredictability of 16.1%, 7.7%, 6.3%, 4.5%, and 16.2% in ocellated lizard, tegu, gecko, Gila monster, and monitor, respectively.
 (E) RD is robust to parameters variation in all five species: scale-by-scale error (mean ± SD) is computed from 5,000 simulations with parameters uniformly sampled in a range covering ±10% of the RD parameter normalization factor estimated during Bayesian optimization.

residual error. Although it could be argued that the latter is compatible with the ranges of estimated errors (upper bounds are 14% and 8.1% for geometry- and color-induced errors, respectively), the result above prompted us to further test the validity of the current RD model for describing the Standing's gecko skin color patterning process. Indeed, performing Gaussian mixture model clustering with 2–6 clusters (using a cluster-validation approach²⁴ to evaluate their relative qualities) identified that the optimal number of color clusters is 2 for all individuals of all species, except Standing's gecko for which it is 3 or 4, depending on the individual. Hence, description of the color change process in Standing's gecko is likely to require a more

complex RD model in which more than two steady-state colors are emerging.

Robustness of the RD model

Finally, we conduct in all species a sensitivity analysis in order to estimate how much perturbations of the RD parameters values would affect the scale-by-scale error. To this end, we perform in each species 5,000 “impaired” simulations in which perturbed, rather than optimal, parameter values are used: for each simulation, parameter values are jointly sampled from a uniform distribution covering 20% (i.e., ±10%) of the RD parameters normalization factor (in the covariance function; Equation 32 in STAR Methods) estimated during Bayesian optimization. We

then compute, for each perturbed simulation, its scale-by-scale error in comparison with the unperturbed simulation (i.e., with optimal RD parameters) and fit a gamma distribution to the resulting histogram of errors, from which a mean and standard deviation are computed. These analyses indicate that simulations with suboptimal RD parameter values generate smaller mean scale-by-scale errors ($0.6\% < E_{\text{sbs}}^f < 2.8\%$; Figure 7E) than uncertainties in skin geometry and in juvenile colors (Figures 7B and 7D).

DISCUSSION

On evolutionary convergence

Our analyses reveal that two connected spectacular phenotypic characters associated to skin color, scale-by-scale color patterns and cellular automaton dynamics of patterning, have evolved independently in multiple divergent lineages of squamates (Figure 1A). This conclusion has important consequences for our understanding of phenotypic evolution in a developmental context. Indeed, the sequence of events that led to such convergences is, in principle, easy to understand: all that is needed for both phenotypic characters to emerge is that a Turing instability mechanism of skin color patterning occurs in a skin domain with a periodic thickness (here, caused by skin scales). Note that two additional requirements must be met. First, the intrinsic self-organizational Turing length scale of the color pattern should be larger than the length scale of the thickness period. Second, the relative timings of scale development and of skin color patterning are important. In principle, scale-by-scale coloration could occur with both scales developing before or after skin color patterning. Indeed, even if color patterning has reached steady state before the onset of scale development, the pattern should “snap” to the position of scales if the dynamical system of chromatophore interactions remains active. On the other hand, if the color pattern gets fixed (because of the arrest of the underlying microscopic dynamical system) before emergence of scales, then the newborn will not exhibit a scale-by-scale color pattern. Such a situation is observed in ball pythons (*Python regius*). We therefore conjecture that convergences and reversals of these two phenotypic characters, hence, the existence of closely related species differing by the presence/absence of scale-by-scale skin color and/or CA dynamics, can easily occur through heterochrony, i.e., by changes in the relative timings or rates of skin scales development and skin color patterning.

Prediction of individual scale-by-scale patterns and their neighborhood statistics

Our results show that, in the case of scale-by-scale skin color patterning, simple statistical (such as sCA and Lenz-Ising) models capture well enough the underlying non-linear dynamical system to predict, with high precision, the time evolution of the pattern statistical features. Using the assumption that the functions of skin color patterns are associated to their statistical features,¹⁵ such as their length scale or their general “look” (e.g., spots versus stripes versus labyrinths), we argue that statistical and RD numerical models can capture most of the functionally relevant behavior of these specific morphogenetic self-organizational systems.

Conversely, we show that (1) the improvement of predictability of the exact scale-by-scale dynamics of individual animals in species exhibiting vastly different color patterns is large for deterministic RD over simpler statistical models and (2) the residual error in exact scale-by-scale color predictability is explained mostly, if not entirely, by uncertainties in skin geometry variation and in color measurements at the initial (juvenile) condition. One could argue that the higher efficiency of RD is due to the incorporation of a large number of parameters (overfitting) and/or to fine-tuning of these parameters. The overfitting argument could hold for the Lenz-Ising but not for the sCA model as the former includes only two parameters (J and B), but the latter incorporates 16 parameters, all of which are optimized in our analyses. Remarkably, parameter fine-tuning can also be ruled out as an explanation for the efficiency of RD. Indeed, the sensitivity analysis reported here (Figure 7E) indicates that the high performance of RD to predict actual scale-by-scale patterns is robust to RD model parameter perturbations, hence, does not require parametrizing the system down to its many nanoscopic/microscopic variables.¹⁸

In addition, iterative addition of RD model parameters during Bayesian optimization (Table S1) identified the necessity to jointly optimize only 3–5 parameters (Table S2) among a total of 16 (all others are set to the values of Manukyan et al.¹²). This result indicates that the values of the remaining parameters either cannot be optimized beyond those estimated in Manukyan et al.,¹² or are somewhat unimportant. Examples of the latter are evident in the pairwise parameter gradient plots of Figure S5: e.g., c_w and c_2 affect the pattern much more than c_7 (cf. 3rd and 6th graphs on the first line of gradient plots) and, similarly, the pattern is very sensitive to c_3 and relatively insensitive to c_u (1st column, 5th row). Note also that obtaining a given low-error pattern depends on the joint optimization of multiple parameters such that single-parameter values within versus among species should not be over-interpreted. For example, despite that optimized values of the growth parameter q (Table S2) differ more between black and white tegu SM1 and SM2 than between SM2 and V11 (a mangrove monitor lizard), simulating the adult pattern of SM2 using the set of optimized parameter values from SM1 (or vice versa) generates an error (E_{16D} value) smaller than the difference of nearest-neighbor configuration statistics between the two tegu individuals (Table S3). In other words, the joint set of optimized parameters of SM1 or of SM2 generate tegu-specific patterns, whereas the set of parameters of V11 does not (Table S3). A similar comparison (Table S4) is performed for one ocellated lizard (TL1) with an optimized value of the parameter $D_{u,v}$ (Table S2), closer to that of a tegu (SM1) than to that of another ocellated lizard (TL2). Hence, interpreting RD parameters in isolation is difficult because the RD model is phenomenological rather than built bottom-up from quantified molecular or cellular processes.

In conclusion, the superior efficiency of RD over stochastic models in describing the scale-by-scale skin color patterning process in multiple divergent species of lizards is due to intrinsic properties of RD, including the ability to exploit mesoscopic information such as (1) continuous character states (here, scale colors) and (2) the implicit relation between the pattern length scale and geometrical parameters (here, growth and

scale/edge geometries). In other words, spatial discretization of the RD model and/or reduction of its spatial dimensions (from 3D to 2D) do not affect much its efficiency in predicting scale-by-scale patterns, whereas discretization of states (to two colors) in the sCA and Ising models, as well as the inability of these approaches to incorporate diffusion and growth, generate a larger negative impact.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
 - Animal scanning and identification of the boundaries and colors of skin scales
 - Scale matching
 - Stochastic Cellular Automaton
 - Reaction-diffusion
 - Enhanced RD
 - Bayesian optimization of RD parameters
 - Uncertainty in skin thickness spatial distribution and production of 3D super-Gaussian heterogeneous networks
 - Color measurement uncertainty and Lyapunov spectrum analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.10.044>.

A video abstract is available at <https://doi.org/10.1016/j.cub.2022.10.044#mmc5>.

ACKNOWLEDGMENTS

We thank Szabolcs Zakany for commenting on the manuscript and on technical aspects of simulations. We thank Adrien Debry and Florent Montange for technical assistance with scanning animals, as well as Carine Langrez and Grigori Timin for HREM analyses. This work was supported by grants to M.C.M. from the Georges & Antoine CLARAZ foundation, the Swiss National Science Foundation (FNSNF, grants 31003A and CR32I3), the International Human Frontier Science Program Organisation (HFSP RGP0019/2017), and the European Research Council (ERC, advanced grant EVOMORPHYS) under the European Union's Horizon 2020 research and innovation program. The funding bodies played no role in the design of the study; collection, analysis, and interpretation of data; and writing the manuscript.

AUTHOR CONTRIBUTIONS

M.C.M. conceived and supervised all aspects of the study. E.J. performed all coding and numerical simulations. E.J. and M.C.M. analyzed the data and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 15, 2022

Revised: July 25, 2022

Accepted: October 19, 2022

Published: November 14, 2022

REFERENCES

1. Nakamasu, A., Takahashi, G., Kanbe, A., and Kondo, S. (2009). Interactions between zebrafish pigment cells responsible for the generation of Turing patterns. *Proc. Natl. Acad. Sci. USA* 106, 8429–8434. <https://doi.org/10.1073/pnas.0808622106>.
2. Inaba, M., Yamanaka, H., and Kondo, S. (2012). Pigment pattern formation by contact-dependent depolarization. *Science* 335, 677. <https://doi.org/10.1126/science.1212821>.
3. Frohnhöfer, H.G., Krauss, J., Maischein, H.M., and Nüsslein-Volhard, C. (2013). Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. *Development* 140, 2997–3007. <https://doi.org/10.1242/dev.096719>.
4. Hamada, H., Watanabe, M., Lau, H.E., Nishida, T., Hasegawa, T., Parichy, D.M., and Kondo, S. (2014). Involvement of Delta/Notch signaling in zebrafish adult pigment stripe patterning. *Development* 141, 318–324. <https://doi.org/10.1242/dev.099804>.
5. Singh, A.P., Schach, U., and Nüsslein-Volhard, C. (2014). Proliferation, dispersal and patterned aggregation of iridophores in the skin prefigure striped coloration of zebrafish. *Nat. Cell Biol.* 16, 607–614. <https://doi.org/10.1038/ncb2955>.
6. Eom, D.S., and Parichy, D.M. (2017). A macrophage relay for long-distance signaling during postembryonic tissue remodeling. *Science* 355, 1317–1320. <https://doi.org/10.1126/science.aal2745>.
7. Patterson, L.B., and Parichy, D.M. (2019). Zebrafish pigment pattern formation: insights into the development and evolution of adult form. *Annu. Rev. Genet.* 53, 505–530. <https://doi.org/10.1146/annurev-genet-112618-043741>.
8. Kondo, S., and Miura, T. (2010). Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* 329, 1616–1620. <https://doi.org/10.1126/science.1179047>.
9. Turing, A.M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. Series B, Biol. Sci.* 237, 37–72.
10. Gierer, A., and Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik* 12, 30–39.
11. Murray, J.D. (2002). *Mathematical Biology II*, 3rd Edition] (Springer).
12. Manukyan, L., Montandon, S.A., Fofonjka, A., Smirnov, S., and Milinkovitch, M.C. (2017). A living mesoscopic cellular automaton made of skin scales. *Nature* 544, 173–179. <https://doi.org/10.1038/nature22031>.
13. Schweisguth, F., and Corson, F. (2019). Self-organization in pattern formation. *Dev. Cell* 49, 659–677. <https://doi.org/10.1016/j.devcel.2019.05.019>.
14. Fofonjka, A., and Milinkovitch, M.C. (2021). Reaction-diffusion in a growing 3D domain of skin scales generates a discrete cellular automaton. *Nat. Commun.* 12, 2433. <https://doi.org/10.1038/s41467-021-22525-1>.
15. Zakany, S., Smirnov, S., and Milinkovitch, M.C. (2022). Lizard skin patterns and the Ising model. *Phys. Rev. Lett.* 128, 048102.
16. Lenz, W. (1920). Beiträge zum Verständnis der magnetischen Eigenschaften in Festen Körpern. *Phys. Z.* 21, 613–615.
17. Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* 31, 253–258. <https://doi.org/10.1007/BF02980577>.
18. Milinkovitch, M.C. (2021). Emergence of self-organizational patterning at the mesoscopic scale. *Dev. Cell* 56, 719–721. <https://doi.org/10.1016/j.devcel.2021.03.003>.
19. Martins, A.F., Bessant, M., Manukyan, L., and Milinkovitch, M.C. (2015). R(2)OBBIIE-3D, a fast robotic high-resolution system for quantitative

- phenotyping of surface geometry and colour-texture. *PLoS One* 10. e0126740. <https://doi.org/10.1371/journal.pone.0126740>.
20. Woodham, R.J. (1980). Photometric methods for determining surface orientation from multiple images. *Opt. Eng.* 19, 139–144.
21. Onsager, L. (1944). Crystal statistics I A two-dimensional model with an order-disorder transition. *Phys. Rev.* 65, 117–149. <https://doi.org/10.1103/PhysRev.65.117>.
22. Dickmann, J.E.M., Rink, J.C., and Jülicher, F. (2022). Long-range morphogen gradient formation by cell-to-cell signal propagation. Preprint at bioRxiv. <https://doi.org/10.1101/2022.04.11.487794>.
23. Geyer, S.H., and Weninger, W.J. (2019). High-resolution episcopic microscopy (HREM): looking back on 13 years of successful generation of digital volume data of organic material for 3D visualisation and 3D display. *Appl. Sci. Basel* 9, 3826. <https://doi.org/10.3390/app9183826>.
24. Rousseeuw, P.J. (1987). Silhouettes—a graphical aid to the interpretation and validation of cluster-analysis. *J. Comput. Appl. Math.* 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
25. van der Vorst, H.A. (1992). Bi-Cgstab—a fast and smoothly converging variant of bi-Cg for the solution of nonsymmetric linear-systems. *SIAM J. Sci. Stat. Comput.* 13, 631–644. <https://doi.org/10.1137/0913035>.
26. Frazier, P.I. (2018). A tutorial on Bayesian optimization. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1807.02811>.
27. Jones, D.R., Schonlau, M., and Welch, W.J. (1998). Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* 13, 455–492. <https://doi.org/10.1023/A:1008306431147>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Skin biopsies for HREM	University of Geneva	N/A
Experimental models: Organisms/strains		
2 individuals of <i>T. lepidus</i>	University of Geneva	LLEP_00038; LLEP_00066
2 individuals of <i>S. merianae</i>	University of Geneva	TMER_0001; TMER_0002
4 individuals of <i>H. suspectum</i>	University of Geneva	HSUS_001; HSUS_002; HSUS_003; HSUS_004
2 individuals of <i>V. indicus</i>	University of Geneva	VIND_0002; VIND_0003
2 individuals of <i>P. standingi</i>	University of Geneva	PSTA_00021; PSTA_00022
Software and algorithms		
MATLAB R2021a (version 9.10)	https://www.mathworks.com/products/matlab.html	N/A
MATLAB Image Processing Toolbox (version 11.3)	https://www.mathworks.com/products/matlab.html	N/A
MATLAB Statistics and Machine Learning Toolbox (version 12.1)	https://www.mathworks.com/products/statistics.html	N/A
MATLAB Parallel Computing Toolbox (version 7.4)	https://www.mathworks.com/products/parallel-computing.html	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and material should be directed to and will be fulfilled by the lead contact, Michel C. Milinkovitch (Michel.Milinkovitch@unige.ch).

Materials availability

This study did not generate new materials.

Data and code availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the [supplemental information](#). Executables implementing the 2D and 3D RD processes are available at <https://github.com/LANEvol/Discrete-RD.git> and <https://github.com/LANEvol/RD-Curvilinear.git>, respectively, for repeating the numerical simulations presented here.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ocellated lizards, Argentinian tegus, Gila monsters, mangrove monitors and Standing's day geckos are housed in Milinkovitch and Tzika's laboratory, Department of Genetics and Evolution, University of Geneva, Switzerland. Maintenance of, and experiments on animals were approved by the Geneva Veterinary Cantonal authorities (authorizations GE/82/14, GE/169/17 and GE24/33145) and performed according to the Swiss law. These guidelines meet international standards.

METHOD DETAILS

Animal scanning and identification of the boundaries and colors of skin scales

Animals were scanned with R²OBBIE-3D, a robotic system¹⁹ integrating a six-axis robotic arm, a 36-megapixel digital single-lens reflex color camera (Nikon D810) and an illumination basket of 30 high-intensity light-emitting diodes (LEDs). For each position of the camera, 30 images were taken (each with a different LED, *i.e.*, a different lighting direction) and combined to infer the normal map of the corresponding 3D surface with the photometric stereo (PS) approach.²⁰ Depending on the animal size at the time of scanning (they were scanned 14 to 45 times over a period of 2 to 4 years starting from the juvenile stage), 90 to 780 images were taken as

the back and flanks were covered by using 3 to 26 different camera positions. The corresponding multiview-PS approach generated 3D models with a resolution of >400 pixels per scale, allowing for the micro-geometry reconstruction of these skin appendages, a pre-requisite for automated identification of skin scales described below. One advantage of the PS approach is that it produces both the normal and the albedo of each pixel. We stored the normal vectors of each PS patch as a 16-bits RGB color image in which each channel contains a component of the normal vector linearly transferred from range $(-1,1)$ to $(0,65535)$.

We then used the normal map to compute the mean curvature of the skin surface, parametrizing the surface $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ with the variables u and v :

$$\Phi(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \end{pmatrix} \quad \text{(Equation 1)}$$

where $x = u$, $y = v$, $z = z(u, v)$. Normals to the surface are then computed as

$$\mathbf{n} = \frac{\Phi_u \times \Phi_v}{\|\Phi_u \times \Phi_v\|} \quad \text{(Equation 2)}$$

where $\Phi_u = \frac{\partial \Phi}{\partial u}$ and $\Phi_v = \frac{\partial \Phi}{\partial v}$. Substituting Equation 1 into Equation 2 yields

$$\mathbf{n} = \frac{1}{\sqrt{z_u^2 + z_v^2 + 1}} \begin{pmatrix} -z_u \\ -z_v \\ 1 \end{pmatrix} \quad \text{(Equation 3)}$$

where $z_u = \frac{\partial z}{\partial u}$ and $z_v = \frac{\partial z}{\partial v}$. If the normal vector is $\mathbf{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$, according to Equation 3, we can write $z_u = -n_1/n_3$, $z_v = -n_2/n_3$, we can compute Φ_u and Φ_v as

$$\Phi_u = \begin{pmatrix} 1 \\ 0 \\ -\frac{n_1}{n_3} \end{pmatrix}, \Phi_v = \begin{pmatrix} 0 \\ 1 \\ -\frac{n_2}{n_3} \end{pmatrix} \quad \text{(Equation 4)}$$

and we can write

$$\Phi_{uu} = \begin{pmatrix} 0 \\ 0 \\ -\frac{\partial}{\partial u} \left(\frac{n_1}{n_3} \right) \end{pmatrix}, \Phi_{vv} = \begin{pmatrix} 0 \\ 0 \\ -\frac{\partial}{\partial v} \left(\frac{n_2}{n_3} \right) \end{pmatrix}, \Phi_{uv} = \begin{pmatrix} 0 \\ 0 \\ -\frac{\partial}{\partial u} \left(\frac{n_2}{n_3} \right) \end{pmatrix} \quad \text{(Equation 5)}$$

where $\Phi_{uu} = \frac{\partial^2 \Phi}{\partial u^2}$, $\Phi_{vv} = \frac{\partial^2 \Phi}{\partial v^2}$, and $\Phi_{uv} = \frac{\partial^2 \Phi}{\partial u \partial v}$.

We then compute the mean curvature by using the first and second fundamental form, F_1 and F_2 .

$$F_1 = \begin{bmatrix} E & F \\ F & G \end{bmatrix}, F_2 = \begin{bmatrix} L & M \\ M & N \end{bmatrix}$$

in which

$$\begin{aligned} E &= \Phi_u \cdot \Phi_u, & F &= \Phi_u \cdot \Phi_v, & G &= \Phi_v \cdot \Phi_v \\ L &= \Phi_{uu} \cdot \mathbf{n}, & M &= \Phi_{uv} \cdot \mathbf{n}, & N &= \Phi_{vv} \cdot \mathbf{n} \end{aligned} \quad \text{(Equation 6)}$$

After substituting the derivatives of Φ from Equations 4 and 5 into Equation 6, we have

$$\begin{aligned} E &= 1 + \left(\frac{n_1}{n_3} \right)^2, & F &= \frac{n_2 n_1}{n_3^2}, & G &= 1 + \left(\frac{n_2}{n_3} \right)^2 \\ L &= -\frac{\partial}{\partial u} \left(\frac{n_1}{n_3} \right) n_3, & M &= -\frac{\partial}{\partial v} \left(\frac{n_1}{n_3} \right) n_3, & N &= -\frac{\partial}{\partial v} \left(\frac{n_2}{n_3} \right) n_3 \end{aligned} \quad \text{(Equation 7)}$$

The mean curvature H is then computed as

$$H = \frac{EN - 2FM + GL}{2(EG - F^2)} \quad \text{(Equation 8)}$$

By substituting Equation 7 into Equation 8, we compute the mean curvature (from the normal map) as

$$H = -\frac{n_3}{2}(n_3^2 + n_1^2) \frac{\partial}{\partial V} \left(\frac{n_2}{n_3} \right) + n_3 n_2 n_1 \frac{\partial}{\partial V} \left(\frac{n_1}{n_3} \right) - \frac{n_3}{2}(n_3^2 + n_2^2) \frac{\partial}{\partial U} \left(\frac{n_1}{n_3} \right)$$

Figures 1B and 1C show the normal map and the mean curvature for a rectangular patch of skin of an ocellated lizard (TL1) at different post-hatching developmental stages. We then apply both a Gaussian and a Median filter on the mean curvature field to remove very short wavelength noises while preserving the sharpness of scales boundaries. Note that, as the average scale size in pixels is about s , we find that smoothing is efficiently applied by the Gaussian and Median filters with a variance $\sigma = s/10$ and window size $\mathbf{w} = (s, s)^T$, respectively. Figure 1C shows the mean curvature after applying the two filters. Each local minimum in the filtered curvature field indicates an individual scale. However, the actual centroid of the corresponding scale is not necessarily located at this minimum because scales in all the species investigated (except Standing's gecko) are slightly asymmetrical and tend to have their lowest curvature slightly displaced distally (*i.e.*, caudally). Hence, we use the watershed algorithm to segment the filtered mean curvature grayscale images with the darkest regions representing the local minima. Although the resulting segmentation of scales is approximate, the centroids of the watershed segments better correspond to the actual center of scales. Proper scale boundaries (red lines in Figure 1C) are then identified by applying Voronoi tessellation on the centroids of the watershed segments.

After identifying scale borders as described above, scale colors are determined by averaging color across pixels within each scale. Note that pixels too close to scales boundaries (distance $< s/10$) are excluded to avoid artifactual shading due to the high curvature of scale borders.

Scale matching

In order to trace, across time points, all scales in the Voronoi meshes generated for each time-point, we first attempt to match all possible pairs of Voronoi meshes. This is achieved by finding for each scale an affine transformation (allowing for translations, rotations, scaling and shear) which transforms the scale centers between two Voronoi meshes at times r and t . The affine matrix M in 2D transforms the target coordinates $\begin{pmatrix} x_t \\ y_t \end{pmatrix}$ into the reference coordinate $\begin{pmatrix} x_r \\ y_r \end{pmatrix}$ according to

$$\begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix} = M \begin{pmatrix} x_t \\ y_t \\ 1 \end{pmatrix}$$

where M is a 3×3 matrix which contains 6 variable coefficients as

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

If we know at least three matching points between the reference and the target spaces, we can find M as

$$M = \begin{pmatrix} x_{r1} & x_{r2} & x_{r3} \\ y_{r1} & y_{r2} & y_{r3} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_{t1} & x_{t2} & x_{t3} \\ y_{t1} & y_{t2} & y_{t3} \\ 1 & 1 & 1 \end{pmatrix}^{-1}$$

If we use >3 points to compute M , we obtain an over-determined system yielding a least-square approximation of the affine transformation. After initiating the matching algorithm by manually selecting the centers of at least 3 scales in all time points, the first iteration consists into computing an affine transformation for each scale (independently for each pair of time points) that is a first neighbor of any of the manually-selected scales, *i.e.*, affine matrices are computed independently for each neighbor scale. The center of each of these newly-considered scales is then transferred from the source to the target space by using the corresponding affine matrix. If the distance between a transferred point and a scale center in the target lattice is smaller than ϵ pixels (*i.e.*, 0.05 times the average distance between neighboring scales), the two centers are registered as a match between the source and target lattices. Once all transferred points have been tested, the patch in the source lattice is updated by adding the new matches. The process is then iterated by finding all first neighbors of scales in the updated patch and computing the corresponding affine transformations. The affine transformation for each new scale considers at least 3 and at most 10 closest neighbors already matched. Note that points that have been transferred but not matched at iteration i will necessarily be first-neighbors of the updated patch and will therefore be transferred again at iteration $i+1$. These points will be more likely to find a match in the target lattice because the corresponding affine matrix will involve more neighbors already matched. The matching algorithm stops when no new match is found.

After matching all pairs of time points, we combine all pairwise matches into a global graph that we term a 'space-time Voronoi network', in which nodes are scales (each scale receives a unique label at each time-point) and edges are matches. Each scale for which the number of nodes in the global graph equals the number of time-points is selected for the next step. In other words, the selected connected components should represent the scales for which at least one match exists in each time-point. The final patch of matched scales generated by the algorithm is a lattice with a jagged boundary and possibly some missing elements inside. More specifically, remaining unmatched points correspond to real scales that cannot be matched with the use of an affine transformation. At this stage, we find a very small proportion of these errors (average = 5%) which are then easily corrected individually by manually identifying the corresponding scale in one time-point such that the matches are automatically generated at all other time-

points using local affine transformations. Next, we correct for errors in connectivity. According to our observations, the number of scales as well as their connectivity are invariant in time. Hence, the approximately 10% mismatch in connectivity that we observe between pairs of time-points are errors that should be corrected. For that purpose, we generate a unified space-time Voronoi network by averaging (after correcting for scaling, translation and rotation among time-points) the centers of each scale across time points. We exclude shear and anisotropic scaling because they would transform the Voronoi network topology in favor of the reference time-point. Given these constraints, we can write the transformation as

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} s \cos(\theta) & -s \sin(\theta) \\ s \sin(\theta) & s \cos(\theta) \end{pmatrix} \begin{pmatrix} x'_t \\ y'_t \end{pmatrix}, \begin{pmatrix} x'_t \\ y'_t \end{pmatrix} = \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \mathbf{t} \quad (\text{Equation 9})$$

in which s is the isotropic scaling factor, θ is the rotation angle, $\mathbf{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ is the translation vector and subscript t and p denote the coordinates of target scales before and after transformation, respectively. The translation vector \mathbf{t} can be easily found from the centroids of the scales as

$\mathbf{t} = \begin{pmatrix} \bar{x}_r - \bar{x}_t \\ \bar{y}_r - \bar{y}_t \end{pmatrix}$, where subscripts r denotes the coordinates of the reference scales and $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ denotes to the average in which N is the number of scales. In order to find θ and s , we use the least-squares approach to minimize the error function defined as

$$E = \sum_{i=1}^N \left((x_{r,i} - x_{p,i})^2 + (y_{r,i} - y_{p,i})^2 \right) \quad (\text{Equation 10})$$

Substituting Equation 9 into Equation 10 and solving for $\partial E / \partial \theta = 0$ and $\partial E / \partial s = 0$, we get

$$\theta = \arctan \left(\frac{\bar{y}_r \bar{x}_t - \bar{x}_r \bar{y}_t}{\bar{x}_r \bar{x}_t + \bar{y}_r \bar{y}_t} \right), s = \frac{\cos(\theta) (\bar{x}_r \bar{x}_t + \bar{y}_r \bar{y}_t) + \sin(\theta) (\bar{y}_r \bar{x}_t - \bar{x}_r \bar{y}_t)}{\bar{x}_t^2 + \bar{y}_t^2}$$

By knowing θ , s and \mathbf{t} , we can use Equation 9 to transfer all time-points to the reference space and then take, for each scale, its across-time-point average position. The averaged scale centers are then used to generate the final space-time Voronoi network (including color of scales) for the corresponding animal. Figure S6 shows the resulting matched scales for 25 time-points of the ocellated lizard TL1.

Stochastic Cellular Automaton

First, we label all the scales in the Voronoi lattice by one of the nearest-neighbor configurations in the set

$$C = \{ (S, i) | S \in \{ \text{Green}, \text{Black} \}, i \in \{ 0, 1, 2, 3, 4, 5, 6, 7 \} \}$$

where S indicates the state (green or black) of the scale and i denotes the number of isochromatic neighbors. We assume that the probability p of a scale changing color in the space-time network (Figure 1F) only depends on its nearest-neighbor configuration, *i.e.*, $p = p(S, i)$, such that the sCA model can be defined with $2 \times 8 = 16$ parameters. As frequencies of rare color changes, and/or of changes that involve rare nearest-neighbor configurations, are poor estimates of the corresponding color-change probabilities p , we use Bayesian inference to compute the expectation $\mathbb{E}[p]$ of these probabilities. Assuming that color change follows a Binomial distribution with probability of k successes from n trials, the posterior probability of p (assuming flat priors) is

$$f_{p|k,n} = (n+1) \binom{n}{k} p^k (1-p)^{n-k}$$

with the maximum of that posterior distribution function positioned at

$$p_{MAP} = \operatorname{argmax}_p f_{p|k,n} = \frac{k}{n}$$

i.e., at the corresponding observed color-change frequency. On the other hand, the expectation of p is:

$$\mathbb{E}[p] = \frac{k+1}{n+2}$$

The sCA rules are obtained simply by summing up trials and successes across time-points and deriving probabilities for each of the 16 possible configurations. Note that, during numerical simulations, we adapt the number of sCA iterations such that the cumulative sum of color flips between initial and final time-points is approximately equal to the number of flips observed in the corresponding real animal. As an example, the time-evolution of neighborhood statistics in two ocellated lizard individuals (TL1 and TL2) are shown in Figures S1A and S2A.

Reaction-diffusion

Following the notation of Manukyan et al.¹² and Fofonjka and Milinkovitch,¹⁴ we use a RD model with three components u, v, w such that the three coupled partial differential equations (PDEs) can be written in matrix form as

$$\partial_t \mathbf{u} = \mathbf{F}(\mathbf{u}) - \mathbf{c}\mathbf{u} + \mathbf{D}\nabla^2 \mathbf{u} \quad (\text{Equation 11})$$

in which

$$\mathbf{u} = \begin{pmatrix} u \\ v \\ w \end{pmatrix}, \mathbf{F}(\mathbf{u}) = \begin{pmatrix} F(u, v, w) \\ G(u, v, w) \\ H(u, v, w) \end{pmatrix}, \mathbf{c} = \begin{pmatrix} c_u & 0 & 0 \\ 0 & c_v & 0 \\ 0 & 0 & c_w \end{pmatrix}, \mathbf{D} = \begin{pmatrix} D_u & 0 & 0 \\ 0 & D_v & 0 \\ 0 & 0 & D_w \end{pmatrix}$$

and the Laplacian operator can be written in the compact form

$$\nabla^2 \mathbf{u} = \nabla \cdot (\nabla \mathbf{u}) = \begin{pmatrix} \partial_{xx}^2 u + \partial_{yy}^2 u + \partial_{zz}^2 u \\ \partial_{xx}^2 v + \partial_{yy}^2 v + \partial_{zz}^2 v \\ \partial_{xx}^2 w + \partial_{yy}^2 w + \partial_{zz}^2 w \end{pmatrix} \quad (\text{Equation 12})$$

Note that the non-linearity of the reaction terms is due to their step-wise function form:

$$F(u, v, w) = \begin{cases} 0 & : & c_1 v + c_2 w + c_3 < 0 \\ c_1 v + c_2 w + c_3 & : & 0 \leq c_1 v + c_2 w + c_3 \leq F_{\max} \\ F_{\max} & : & F_{\max} < c_1 v + c_2 w + c_3 \end{cases}$$

$$G(u, v, w) = \begin{cases} 0 & : & c_4 u + c_5 w + c_6 < 0 \\ c_4 u + c_5 w + c_6 & : & 0 \leq c_4 u + c_5 w + c_6 \leq G_{\max} \\ G_{\max} & : & G_{\max} < c_4 u + c_5 w + c_6 \end{cases} \quad (\text{Equation 13})$$

$$H(u, v, w) = \begin{cases} 0 & : & c_7 u + c_8 v + c_9 < 0 \\ c_7 u + c_8 v + c_9 & : & 0 \leq c_7 u + c_8 v + c_9 \leq H_{\max} \\ H_{\max} & : & H_{\max} < c_7 u + c_8 v + c_9 \end{cases}$$

3D continuous RD (3D-cRD)

The Laplacian operator in Equation 12 is defined in the Cartesian space. We then rewrite the Laplacian operator in curvilinear 3D space (accurately representing the skin geometry) as follows. We define curvilinear space coordinates y_i , parametrized by Cartesian space coordinates, x_i , as

$$y_i = y_i(x_1, x_2, x_3)$$

in which $i = 1, 2, 3$ is the dimension index. In order to find the equivalent operator in the curvilinear space, we first rewrite the gradient operator as

$$\nabla u = \frac{\partial u}{\partial x_i} \mathbf{e}_i = \frac{\partial u}{\partial y_i} \mathbf{b}^i$$

in which \mathbf{e}_i are the basis vectors in the Cartesian system and

$$\mathbf{b}^i = \nabla y_i = \frac{\partial y_i}{\partial x_j} \mathbf{e}_j = b_j^i \mathbf{e}_j$$

is the covariant basis in the curvilinear system.

Then, the Laplacian of u reads

$$\begin{aligned} \nabla^2 u &= \nabla \cdot (\nabla u) = \nabla \cdot \left(\frac{\partial u}{\partial y_i} \mathbf{b}^i \right) \\ &= \nabla \left(\frac{\partial u}{\partial y_i} \right) \cdot \mathbf{b}^i + \frac{\partial u}{\partial y_i} \nabla \cdot \mathbf{b}^i \\ &= \frac{\partial^2 u}{\partial y_j \partial y_i} \mathbf{b}^j \cdot \mathbf{b}^i + \frac{\partial u}{\partial y_i} \frac{\partial b_j^i}{\partial y_k} \frac{\partial y_k}{\partial x_j} \\ &= \frac{\partial^2 u}{\partial y_j \partial y_i} \mathbf{b}^j \cdot \mathbf{b}^i + \frac{\partial u}{\partial y_i} \frac{\partial b_j^i}{\partial y_k} b_j^k \end{aligned} \quad (\text{Equation 14})$$

The covariant basis components, b_j^i , appearing in Equation 14 are computed for all nodes in the structured grid. The nodes located at the grid borders are named 'boundary nodes' whereas all others are 'interior nodes'. The latter are assumed to be at the center of a

3×3×3 stencil for which the finite-difference approximation is computed and the discrete Laplacian operator is written in a matrix form using the local indices of the nodes in the stencil. In order to approximate the covariant basis components of a node, we take into account the x_i and y_i variations of all 27 neighbor nodes in the corresponding stencil. More precisely, the metric tensor components are the slope of the lines calculated using the least-square linear regression on the 27 data points. For boundary nodes, there are 22 possible different stencils (all with <27 nodes), but the least-square linear regression remains valid for these boundary nodes. Then, we use finite-difference formulas to approximate the first, second and cross derivatives appearing in Equation 14. Finally, Equation 11 is time integrated using the Euler backward scheme. Then, the resulting linear system is solved with a stabilized variant of the Bi-conjugate gradient method (BiCGSTAB²⁵).

2D continuous RD (2D-cRD)

As the skin domain thickness (h in the Z dimension) is much smaller than its size in the two other spatial dimensions, it is reasonable to neglect variations of RD variables across Z . This assumption enables reducing the model from 3D to 2D by integrating in the Z direction¹⁴ the RD equations (Equation 11) to give

$$\partial_t \bar{\mathbf{u}} = \mathbf{F}(\bar{\mathbf{u}}) - \mathbf{c}\bar{\mathbf{u}} + \frac{1}{h} \nabla \cdot (h(\mathbf{x}) \mathbf{D} \nabla \bar{\mathbf{u}})$$

in which $\bar{\mathbf{u}} = \frac{1}{h} \int_{Z_0}^{Z_0+h} \mathbf{u} dz$ and $\bar{h} = \frac{1}{A_{\text{skin}}} \int_{A_{\text{skin}}} h(\mathbf{x}) dA$.

Using a position-dependent diffusion coefficient $\mathbf{D}(\mathbf{x}) = h(\mathbf{x})/\bar{h} \mathbf{D}$, we can rewrite the reduced-order RD equations as

$$\partial_t \bar{\mathbf{u}} = \mathbf{F}(\bar{\mathbf{u}}) - \mathbf{c}\bar{\mathbf{u}} + \nabla \cdot (\mathbf{D}(\mathbf{x}) \nabla \bar{\mathbf{u}})$$

As we are building a 2D model, we replace hereafter $\bar{\mathbf{u}}$ by \mathbf{u} . We can further simplify the model by assuming that scales are prismatic (*i.e.*, thickness is reduced only at discrete scale borders). In 2D, this is translated into taking into account the reduction of skin thickness through scaling of the continuous RD diffusion coefficients (by a factor P) at the one-dimensional edges of the 2D scales,¹² *i.e.*, we introduce a position-dependent diffusion coefficient in the 2D model, with the diffusion matrix reading

$$\mathbf{D}(\mathbf{x}) = \begin{cases} P\mathbf{D} & \mathbf{x} \in \text{scale boundary} \\ \mathbf{D} & \text{else} \end{cases}$$

Then, similar to the 3D-cRD case above, we use the finite-difference method to discretize the diffusion term. For that purpose, we consider that the physical domain is discretized into a square grid with spacing ε . As both the physical and computational domains are equivalent square grids, the metric tensor simplifies into a unit tensor. Similarly to Manukyan et al.,¹² the 2D-cRD equations become

$$\frac{d\mathbf{u}_i}{dt} = \mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i + \frac{1}{\varepsilon^2} \sum_j \mathbf{D}(\bar{\mathbf{x}}_{ij}) (\mathbf{u}_j - \mathbf{u}_i) \quad (\text{Equation 15})$$

where $\bar{\mathbf{x}}_{ij} = (\mathbf{x}_i + \mathbf{x}_j)/2$ and j represents the neighbors of the i^{th} node.

2D Discrete RD (2D-dRD)

We derive here the discrete-RD equations by using the finite-volume method for arbitrary polygons. The governing equations (Equation 1) are integrated over the skin domain Ω .

$$\int_{\Omega} \partial_t \mathbf{u} dA = \int_{\Omega} [\mathbf{F}(\mathbf{u}) - \mathbf{c}\mathbf{u} + \nabla \cdot (\mathbf{D}(\mathbf{x}) \nabla \mathbf{u})] dA$$

Using the Divergence theorem, we obtain

$$\frac{d}{dt} \int_{\Omega} \mathbf{u} dA = \int_{\Omega} (\mathbf{F}(\mathbf{u}) - \mathbf{c}\mathbf{u}) dA + \int_{\partial\Omega} \mathbf{D}(\mathbf{x}) \frac{\partial \mathbf{u}}{\partial n} dL$$

Then we write this integral for a single scale labeled by index i :

$$A_i \frac{d\mathbf{u}_i}{dt} = A_i [\mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i] + P\mathbf{D} \sum_j \left(\frac{\partial \mathbf{u}}{\partial n} \right)_{ij} L_{ij} \quad (\text{Equation 16})$$

where $A_i = \int_{\Omega_i} dA$, $L_{ij} = \int_{\partial\Omega_i \cap \partial\Omega_j} dL$, and $u_i = \frac{1}{A_i} \int_{\Omega_i} \mathbf{u} dA$

are the scale area, the length of the edge between scales i and j , and the area-averaged values of component u , respectively. To solve Equation 16, we need to compute $(\partial \mathbf{u} / \partial n)_{ij}$, *i.e.*, the diffusion flux exchanged between neighboring scales. Recalling that the scales boundary thickness is ε , we can approximate the normal-oriented flux as

$$\left(\frac{\partial \mathbf{u}}{\partial n} \right)_{ij} \approx \frac{\mathbf{u}_j - \mathbf{u}_i}{\varepsilon}$$

This approximation implies that the RD component concentrations are changing linearly across interscale edges. Finally, substituting the normal flux into Equation 16, we obtain the discrete RD equations for arbitrary polygonal scales:

$$\frac{d\mathbf{u}_i}{dt} = \mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i + \frac{PD}{A_i \varepsilon} \sum_j (\mathbf{u}_j - \mathbf{u}_i) L_{ij} \quad (\text{Equation 17})$$

Note that, if we assume a regular hexagonal lattice, with edge length S , we have

$$A_i = \frac{3\sqrt{3}}{2} S^2, L_{ij} = S$$

such that Equation 17 simplifies to the discrete equation derived in Manukyan et al.¹²:

$$\frac{d\mathbf{u}_i}{dt} = \mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i + \frac{2PD}{3\sqrt{3}S\varepsilon} \sum_j (\mathbf{u}_j - \mathbf{u}_i) \quad (\text{Equation 18})$$

Enhanced RD

Here we integrate growth of the animal as well as variation of color state at initial condition.

Growth model

We assume growth to change as a logistic function

$$S(\tau) = \frac{k_1}{\exp(-k_2\tau + k_3) + 1} \quad (\text{Equation 19})$$

in which S is the average (across all scales) of edges' lengths, τ is the growth time variable and $k_i = 1,2,3$ are the constants determined by fitting the logistic curve to the measurements made on the corresponding real animals. Figure S7A shows the measured scales size as well as the fitted growth functions for the two ocellated lizards (TL1 and TL2). To integrate isotropic growth in the RD equations, we must take into account the different time-scales of the growth *versus* RD processes. Here, we consider that they are related by a rate factor q as

$$q = \frac{dt}{d\tau}$$

in which τ and t denote the time variables of growth and RD processes, respectively. Then, we can rewrite the discrete RD Equation 17 as a function of τ .

$$\frac{1}{q} \frac{d\mathbf{u}_i}{d\tau} = \mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i + \frac{PD}{A_i(\tau)\varepsilon(\tau)} \sum_j (\mathbf{u}_j - \mathbf{u}_i) L_{ij}(\tau) \quad (\text{Equation 20})$$

in which ε , L and A are changing as a functions of τ . Knowing the scales growth function $S(\tau)$ from Equation 19, we determine the growth of all lengths and areas in proportion to the reference time-point geometry such that

$$\varepsilon(\tau) = \frac{S(\tau)}{S_{\text{ref}}} \varepsilon_{\text{ref}}, L_{ij}(\tau) = \frac{S(\tau)}{S_{\text{ref}}} L_{ij,\text{ref}}, A_i(\tau) = \left(\frac{S(\tau)}{S_{\text{ref}}}\right)^2 A_{i,\text{ref}} \quad (\text{Equation 21})$$

Here, we assume $\varepsilon_{\text{ref}} = 1$ and select the last measured time-point as the reference geometry. Substituting Equation 21 into Equation 20 results in a discrete RD (2D-dRD) model integrating growth:

$$\frac{1}{q} \frac{d\mathbf{u}_i}{d\tau} = \mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i + \left(\frac{S_{\text{ref}}}{S(\tau)}\right)^2 \frac{PD}{A_{i,\text{ref}}\varepsilon_{\text{ref}}} \sum_j (\mathbf{u}_j - \mathbf{u}_i) L_{ij,\text{ref}}$$

Similarly, we can integrate growth into the 2D continuous RD (2D-cRD) model by using $dt = qd\tau$ and substituting $\varepsilon(\tau) = \frac{S(\tau)}{S_{\text{ref}}} \varepsilon_{\text{ref}}$ into Equation 15 to obtain

$$\frac{1}{q} \frac{d\mathbf{u}_i}{d\tau} = \mathbf{F}(\mathbf{u}_i) - \mathbf{c}\mathbf{u}_i + \left(\frac{S_{\text{ref}}}{S(\tau)}\right)^2 \frac{1}{\varepsilon_{\text{ref}}^2} \sum_j \mathbf{D}(\bar{\mathbf{x}}_{ij,\text{ref}}) (\mathbf{u}_j - \mathbf{u}_i)$$

Juvenile initial condition (IC) and color transformation matrix

Determining the initial condition (corresponding to the juvenile pattern) requires transferring color information to the space of RD variables. Here, we detail how to obtain the transformation function. First, we linearize the RD equations close to the homogeneous steady state (HSS). By definition, at HSS the temporal and spatial derivatives in Equation 11 vanish, such that

$$\mathbf{F}(\mathbf{u}^*) = \mathbf{c}\mathbf{u}^* \quad (\text{Equation 22})$$

where the vector \mathbf{u}^* corresponds to the RD component concentrations (u^*, v^*, w^*) at the HSS. The reaction terms defined in Equation 13 partition the u, v, w space into 9 linear regions. As we assume that our RD system is initiated by values all located in the central region, the reaction terms F, G and H read

$$\begin{aligned} F(u, v, w) &= c_1v + c_2w + c_3 \\ G(u, v, w) &= c_4u + c_5w + c_6 \\ H(u, v, w) &= c_7u + c_8v + c_9 \end{aligned} \quad (\text{Equation 23})$$

Substituting Equation 23 into Equation 22, we have

$$\begin{pmatrix} u^* \\ v^* \\ w^* \end{pmatrix} = \begin{pmatrix} -c_u & c_1 & c_2 \\ c_4 & -c_v & c_5 \\ c_7 & c_8 & -c_w \end{pmatrix}^{-1} \begin{pmatrix} c_3 \\ c_6 \\ c_9 \end{pmatrix}$$

In addition, as we assume that the HSS belongs to the central region, valid values of the parameters ($c_1, \dots, c_9, c_u, c_v, c_w$) must respect

$$\begin{aligned} 0 < c_1v^* + c_2w^* + c_3 &\leq F_{\max} \\ 0 < c_4u^* + c_5w^* + c_6 &\leq G_{\max} \\ 0 < c_7u^* + c_8v^* + c_9 &\leq H_{\max} \end{aligned} \quad (\text{Equation 24})$$

such that \mathbf{u}^* is a point in 3D u, v, w space inside the region delimited by the six planes defined by Equation 24. We then linearize the governing 2D-dRD equations on the hexagonal lattice: substituting Equation 23 into Equation 18 yields

$$\frac{d\mathbf{u}_i}{dt} = \begin{pmatrix} -c_u & c_1 & c_2 \\ c_4 & -c_v & c_5 \\ c_7 & c_8 & -c_w \end{pmatrix} \mathbf{u}_i + \begin{pmatrix} c_3 \\ c_6 \\ c_9 \end{pmatrix} + \frac{2P}{3\sqrt{3}S\varepsilon} \mathbf{D} \left(\sum_j \mathbf{u}_j - 6\mathbf{u}_i \right) \quad (\text{Equation 25})$$

We can then approximate the summation terms in Equation 25 as $\sum_j \mathbf{u}_j = 6\mathbf{u}^*$ and define

$$\mathbf{b} = \begin{pmatrix} c_3 \\ c_6 \\ c_9 \end{pmatrix} + \frac{4PD}{\sqrt{3}S\varepsilon} \mathbf{u}^*, \mathbf{M} = \begin{pmatrix} -c_u & c_1 & c_2 \\ c_4 & -c_v & c_5 \\ c_7 & c_8 & -c_w \end{pmatrix} - \frac{4PD}{\sqrt{3}S\varepsilon}$$

to write Equation 25 in a compact form representing a system of linear ordinary differential equations

$$\partial_t \mathbf{u} = \mathbf{M}\mathbf{u} + \mathbf{b} \quad (\text{Equation 26})$$

whose solution describing the time evolution of \mathbf{u} is

$$\mathbf{u}(t) = e^{\mathbf{M}t} (\mathbf{u}_0 + \mathbf{M}^{-1}\mathbf{b}) - \mathbf{M}^{-1}\mathbf{b} \quad (\text{Equation 27})$$

in which \mathbf{u}_0 is the initial condition. We can then use spectral decomposition to rewrite Equation 27 as $\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{\Lambda}$ is the diagonal matrix which contains eigenvalues of \mathbf{M} .

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

and \mathbf{V} is a matrix in which columns are $\mathbf{V} = (\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3)$, i.e., the eigenvectors of \mathbf{M} . Then, the exponential of the matrix \mathbf{M} can be written as $e^{\mathbf{M}} = \mathbf{V}e^{\mathbf{\Lambda}}\mathbf{V}^{-1}$ which enables us to rewrite Equation 27 as

$$\begin{aligned} \mathbf{u}(t) &= \mathbf{V}e^{t\mathbf{\Lambda}}\mathbf{V}^{-1} (\mathbf{u}_0 + \mathbf{M}^{-1}\mathbf{b}) - \mathbf{M}^{-1}\mathbf{b} \\ &= \mathbf{V}e^{t\mathbf{\Lambda}} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} - \mathbf{M}^{-1}\mathbf{b} \\ &= e^{\lambda_1 t} a_1 \mathbf{v}_1 + e^{\lambda_2 t} a_2 \mathbf{v}_2 + e^{\lambda_3 t} a_3 \mathbf{v}_3 - \mathbf{M}^{-1}\mathbf{b} \end{aligned} \quad (\text{Equation 28})$$

whereas $a_{1,2,3}$ are constant coefficients computed as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{V}^{-1} (\mathbf{u}_0 + \mathbf{M}^{-1}\mathbf{b})$$

Equation 28 is an approximation of a scale color trajectory in u, v, w space. When t is large enough, the maximum eigenvalue term (λ_1) dominates and the trajectory simplifies to the dominating eigenvector line

$$\mathbf{u}(t') = \mathbf{v}_1 t' - \mathbf{M}^{-1} \mathbf{b} \quad (\text{Equation 29})$$

where $t' = e^{2t} a_1$ can take negative or positive values depending on the sign of a_1 which itself depends on the initial value \mathbf{u} . If $\lambda_1 < 0$, the system is stable and the trajectory converges to the steady state \mathbf{u}^* , whereas, if $\lambda_1 > 0$, the scale color trajectory is directed towards one of two 'extremities' of the dominating eigenvector line, *i.e.*, the 3D points where the line defined by Equation 29 intersects the planes described by Equation 24. Figure S7B shows the color trajectories in u, v, w space starting from different initial conditions. Note that performing this analysis on the RD model integrating growth (see previous section) produces the same dominating eigenvector line.

After performing the linear stability analysis above, we use the dominating eigenvector line Equation 29 to convert the color data of all juvenile scales to u, v, w space (Figure S7C). Indeed, assuming that juvenile scale colors correspond to color variables close to the HSS and vary along the dominating eigenvector line, we construct a transformation matrix that converts color data to u, v, w (and *vice versa*) as follows. First, we convert both juvenile and adult scale colors from RGB to CIELAB color space which expresses color along three axes: perceptual lightness (L^* ; with black at $L^*=0$ and white at $L^*=100$), unbound green–red opponents (a^* ; with $a^*<0$ towards green and $a^*>0$ towards red) and unbound blue–yellow opponents (b^* ; with blue towards $b^*<0$ and yellow towards $b^*>0$). We use the CIELAB color space because it was designed to approximate the nonlinear response of the human eye. We then use the K-mean algorithm to divide color data into two clusters. The cluster centers are denoted by

$$\mathbf{C}_g = \begin{pmatrix} L \\ a \\ b \end{pmatrix}_g \text{ and } \mathbf{C}_b = \begin{pmatrix} L \\ a \\ b \end{pmatrix}_b$$

where $L_g > L_b$.

Using Equations 18 and 24, we compute the two extremities of the distribution line as

$$\mathbf{u}_g = \begin{pmatrix} u \\ v \\ w \end{pmatrix}_g \text{ and } \mathbf{u}_b = \begin{pmatrix} u \\ v \\ w \end{pmatrix}_b$$

where $v_g > v_b$.

We then construct two affine transformation matrices: the first projects every point in color space on the line passing through \mathbf{C}_g and \mathbf{C}_b whereas the second transfers the projected points to the line passing through \mathbf{u}_g and \mathbf{u}_b in u, v, w space. The two corresponding matrices are denoted by $\mathbf{T}_1(\mathbf{C}_g, \mathbf{C}_b)$ and $\mathbf{T}_2(\mathbf{C}_g, \mathbf{C}_b, \mathbf{u}_g, \mathbf{u}_b)$, respectively. For arbitrary vectors $\mathbf{x}, \mathbf{x}', \mathbf{y}$, and \mathbf{y}' , the matrices \mathbf{T}_1 and \mathbf{T}_2 read

$$\mathbf{T}_1(\mathbf{x}, \mathbf{x}') = \begin{pmatrix} \mathbf{e}(\mathbf{x}, \mathbf{x}') \mathbf{e}(\mathbf{x}, \mathbf{x}')^T & (\mathbf{I} - \mathbf{e}(\mathbf{x}, \mathbf{x}') \mathbf{e}(\mathbf{x}, \mathbf{x}')^T) \mathbf{x} \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{T}_2(\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}') = \begin{pmatrix} \mathbf{e}(\mathbf{y}, \mathbf{y}') \mathbf{e}(\mathbf{x}, \mathbf{x}')^T \frac{\|\mathbf{y} - \mathbf{y}'\|}{\|\mathbf{x} - \mathbf{x}'\|} & \mathbf{y} - \left(\mathbf{e}(\mathbf{y}, \mathbf{y}') \mathbf{e}(\mathbf{x}, \mathbf{x}')^T \frac{\|\mathbf{y} - \mathbf{y}'\|}{\|\mathbf{x} - \mathbf{x}'\|} \right) \mathbf{x} \\ 0 & 1 \end{pmatrix}$$

where $\mathbf{e}(\mathbf{x}, \mathbf{x}') = \frac{\mathbf{x} - \mathbf{x}'}{\|\mathbf{x} - \mathbf{x}'\|}$.

In order to allow optimizing the initial condition, all the transferred points are moved closer to the HSS by applying \mathbf{T}_3 which reads

$$\mathbf{T}_3(r, \mathbf{u}^*) = \begin{pmatrix} r\mathbf{I} & (1-r)\mathbf{u}^* \\ 0 & 1 \end{pmatrix}$$

where r is an adjustable parameter ranging between 0 (the point is moved to the HSS) and 1 (the point is not moved). Finally, the color of each juvenile scale is converted to the u, v, w space using the combination of the $\mathbf{T}_1, \mathbf{T}_2$, and \mathbf{T}_3 matrices (see Figure S7C):

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \mathbf{T}_3(r, \mathbf{u}^*) \mathbf{T}_2(\mathbf{C}_g, \mathbf{C}_b, \mathbf{u}_g, \mathbf{u}_b) \mathbf{T}_1(\mathbf{C}_g, \mathbf{C}_b) \begin{pmatrix} C \\ 1 \end{pmatrix} \quad (\text{Equation 30})$$

Color transformation matrix for later time-points

To allow for comparisons between simulated and real patterns at later time points, we define the transformation from the u, v, w to the color space through Principal Component Analysis (PCA) by finding the eigenvectors and eigenvalues of the covariance matrix defined as

$$\text{Cov}(\mathbf{X}) = \frac{1}{n} (\mathbf{X} - \bar{\mathbf{X}})^T (\mathbf{X} - \bar{\mathbf{X}})$$

where n denotes the number of scales and, given that there are three RD variables (and three color variables), \mathbf{X} is a $n \times 3$ data points matrix. For both color and u, v, w spaces, we can extract three orthogonal eigenvectors of which the one corresponding to the largest eigenvalue denotes the axis along which the data points exhibit the largest variation. Hence, the eigenvalues and eigenvectors form an ellipsoid approximately fitted to the data points: the ellipsoid center is the centroid of the data points; the ellipsoid axes and radii are the eigenvectors and the square root of eigenvalues, respectively. Hence, the transformation between the color and u, v, w spaces is defined as the matrix that maps the two ellipsoids from these two spaces. We then assume that (i) the eigenvectors of the color and RD variables are the columns of the matrices \mathbf{V}_C , and \mathbf{V}_u , respectively and (ii) the eigenvalues are the diagonal matrices \mathbf{D}_C and \mathbf{D}_u , respectively. Hence, the affine transformation matrix from RD variable to color space reads

$$\mathbf{T}_e(\mathbf{C}, \mathbf{u}) = \begin{pmatrix} \mathbf{Q}(\mathbf{C}, \mathbf{u}) & \bar{\mathbf{C}} - \mathbf{Q}(\mathbf{C}, \mathbf{u})\bar{\mathbf{u}} \\ 0 & 1 \end{pmatrix}$$

in which $\bar{\mathbf{C}}$ and $\bar{\mathbf{u}}$ are the color and RD data points centroids, respectively, and $\mathbf{Q}(\mathbf{C}, \mathbf{u})$ is the rotation and scaling matrix which reads

$$\mathbf{Q}(\mathbf{C}, \mathbf{u}) = \mathbf{V}_C \mathbf{D}_C^{\frac{1}{2}} \left(\mathbf{V}_u \mathbf{D}_u^{\frac{1}{2}} \right)^{-1}$$

Then, the color of each scale is computed as

$$\begin{pmatrix} \mathbf{C} \\ 1 \end{pmatrix} = \mathbf{T}_e(\mathbf{C}, \mathbf{u}) \begin{pmatrix} \mathbf{u} \\ 1 \end{pmatrix} \quad (\text{Equation 31})$$

Bayesian optimization of RD parameters

We optimize RD parameters using *Bayesian Optimization* (*bayesopt* library in MATLAB R2021a with parallel sampling), a machine-learning global minimization problem, defined as $\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$, in which $f(\mathbf{x})$ is the objective function. According to Frazier,²⁶ this approach is efficient in minimising objective functions that are continuous, expensive to evaluate, have less than 20 dimensions and whose feasible set \mathcal{X} is a hyper-rectangle or d-dimensional simplex.

Gaussian process regression

The core of the method is a random Gaussian process (i.e., a generalization of the multivariate normal distribution to infinite dimensions) which enables sampling random functions. Here, we use this process to approximate the objective function. Let's assume that we already know a 'training dataset' of N outputs of the function $f(\mathbf{x})$ as a vector \mathbf{f} associated with N input vectors stored in the matrix \mathbf{X} :

$$\mathbf{f} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}$$

in which, $\mathbf{x}_i \in \mathbb{R}^d$ is a row vector containing optimisable variables. We know that the function f is sampled from a Gaussian process defined as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

in which m and k are mean and covariance functions, respectively. Now, we would like to estimate $f_* = f(\mathbf{x}_*)$ at any arbitrary point \mathbf{x}_* , i.e., we look for $P(f_* | \mathbf{f})$, the probability of observing f_* given the known output vectors \mathbf{f} . By definition, any finite sets of data chosen from the Gaussian process exhibit a normal joint distribution. Since both f_* and the output vector \mathbf{f} are sampled from the same Gaussian process, they have a normal joint probability $P(\mathbf{f}, f_*)$ which reads

$$\begin{pmatrix} \mathbf{f} \\ f_* \end{pmatrix} \sim P(\mathbf{f}, f_*) = \mathcal{N} \left(\begin{pmatrix} \mathbf{m} \\ m(\mathbf{x}_*) \end{pmatrix}, \begin{pmatrix} \mathbf{k} & \mathbf{k}_*^T \\ \mathbf{k}_* & k(\mathbf{x}_*, \mathbf{x}_*) \end{pmatrix} \right)$$

in which \mathbf{m} is the mean vector and \mathbf{k}, \mathbf{k}_* are covariance block matrices defined as

$$\mathbf{m} = \begin{pmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \end{pmatrix}, \mathbf{k} = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{pmatrix}, \mathbf{k}_* = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_*) \\ \vdots \\ k(\mathbf{x}_N, \mathbf{x}_*) \end{pmatrix}$$

We can compute the conditional probability $P(f_* | \mathbf{f})$ as

$$P(f_* | \mathbf{f}) = \frac{P(f_*, \mathbf{f})}{P(\mathbf{f})} = \frac{P(f_*, \mathbf{f})}{\int P(f_*, \mathbf{f}) df_*}$$

It can be shown that $P(f_* | \mathbf{f})$, from which f_* can be sampled, is a normal distribution

$$f_* \sim P(f_* | \mathbf{f}) = \mathcal{N}(m_{\text{post}}(\mathbf{x}_*), k_{\text{post}}(\mathbf{x}_*, \mathbf{x}_*))$$

in which

$$m_{\text{post}}(\mathbf{x}_*) = m(\mathbf{x}_*) + \mathbf{k}_*^T \mathbf{k}^{-1} (\mathbf{f} - \mathbf{m})$$

$$k_{\text{post}}(\mathbf{x}_*, \mathbf{x}'_*) = k(\mathbf{x}_*, \mathbf{x}'_*) - \mathbf{k}_*^T \mathbf{k}^{-1} \mathbf{k}_*$$

The covariance function $k(\mathbf{x}, \mathbf{x}')$ has a crucial role on the performance of the Gaussian process and here we use:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \sqrt{5}h + \frac{5}{3}h^2 \right) \exp(-\sqrt{5}h) \quad (\text{Equation 32})$$

in which

$$h = \sqrt{(\mathbf{x} - \mathbf{x}')^T \boldsymbol{\sigma}^{-2} (\mathbf{x} - \mathbf{x}')}, \boldsymbol{\sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_d \end{pmatrix} \quad (\text{Equation 33})$$

The scalar σ_f and the diagonal matrix $\boldsymbol{\sigma}$ are called *hyper-parameters* (that allow handling different length scales in parameter space) whose values are computed within the optimization loop. Here, we define the objective function as

$$f = \frac{1}{n_k} \sum_k E_k^{\text{pattern}} \quad (\text{Equation 34})$$

where n_k is the total number of observation time-points and E_k^{pattern} reads

$$E_k^{\text{pattern}} = \frac{1}{n_s} \sum_{i=0}^7 |n_k^{\text{sim}}(\text{Green}, i) - n_k^{\text{obs}}(\text{Green}, i)| + |n_k^{\text{sim}}(\text{Black}, i) - n_k^{\text{obs}}(\text{Black}, i)|$$

where n_s is the total number of scales. The definition of nearest-neighbor statistics $n(S, i)$ is given above in the section [stochastic cellular automaton](#).

Note that the sensitivity analysis over the parameters of the RD model (Figure 7E) was performed by running, in each species, 5,000 simulations with perturbed (unoptimal) parameter values sampled from a uniform random distribution within the range $\pm 0.1\sigma_{1,\dots,d}$, where $\sigma_{1,\dots,d}$ are the hyper-parameters of the covariance function (Equation 33).

Gaussian process optimization loop

Starting with an initial parameter vector \mathbf{x}_0 , evaluation of the objective function $f(\mathbf{x}_0)$ is iteratively accomplished through the following steps: (i) Generate initial condition $\mathbf{u}(\tau_0)$ by converting the observed colors $\mathbf{C}(\tau_0)$ and using the transformation defined in Equation 30; (ii) Solve the discrete-RD equations (Equation 20) for all time points ranging from τ_0 to τ_k ; (iii) Convert the simulation results $\mathbf{u}(\tau_k)$ to the Lab coordinate colors using the observed colors $\mathbf{C}(\tau_k)$ and the transformation defined in Equation 31 and (iv) Evaluate the objective function f according to Equation 34. After point (iv), the output vector and input matrix are updated by adding the corresponding values to the end of \mathbf{f} and \mathbf{X} , respectively. We assume that $f(\mathbf{x})$ can be drawn from the Gaussian process in which $m(\mathbf{x}) = 0$ and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function given in Equation 32. We can use \mathbf{f} and \mathbf{X} to obtain the mean m_{post} and covariance k_{post} at any point in the parameter space. The latter two are used to guess the next sampling point. For that purpose, we define the improvement function as:

$$I(\mathbf{x}_*) = \begin{cases} f_{\min} - f_* & f_{\min} > f_* \\ 0 & f_{\min} \leq f_* \end{cases}$$

in which f_{\min} is the minimum values of the objective function evaluated so far. The improvement function simply indicates the amount by which the objective function is reduced with respect to the best value evaluated so far. For region in which the objective function exceeds f_{\min} , an improvement value of zero is assigned. The expectation of $I(f_*)$ has the following closed form²⁷

$$\mathbb{E}[I(\mathbf{x}_*)] = \int_{-\infty}^{\infty} I(\mathbf{x}_*) P(f_* | \mathbf{f}) df_* = (f_{\min} - m_{\text{post}}(\mathbf{x}_*)) \Phi\left(\frac{f_{\min} - m_{\text{post}}(\mathbf{x}_*)}{k_{\text{post}}(\mathbf{x}_*, \mathbf{x}_*)}\right) + k_{\text{post}}(\mathbf{x}_*, \mathbf{x}_*) \varphi\left(\frac{f_{\min} - m_{\text{post}}(\mathbf{x}_*)}{k_{\text{post}}(\mathbf{x}_*, \mathbf{x}_*)}\right) \quad (\text{Equation 35})$$

where φ and Φ are the probability density function and cumulative density function of the standard normal distribution, respectively. Note that this function is much cheaper to evaluate than the objective function. The point to sample in the next iteration is the one that minimizes the expected improvement function in Equation 35. This classical minimization problem is efficiently handled by a gradient descent algorithm. The optimization loop is repeated until 1000 successive iterations have not generated any improvement.

Narrowing parameter space

Even-though the dRD solver is computationally very efficient, running the optimization loop over the full space of parameters

$$\mathbb{S}_{full} = \{[c_u, c_v, c_w, c_1, \dots, c_9, D_u, D_v, D_w, P, q, r] : c_u, c_v, c_w, c_1, \dots, c_9 \in \mathbb{R}, D_u, D_v, D_w, P, q, r \in \mathbb{R}^+\}$$

is not feasible. To identify a suitable subset of parameters $\mathbb{S} \subset \mathbb{S}_{full}$ on which to perform optimisation, we first take into account the redundancy in the dRD formulation. Indeed, the parameter P can be removed from the optimization loop as it is a multiplier of the diffusion coefficients D_u, D_v, D_w . Furthermore, we remove D_v as $D_u = D_v$. Second, we use the 2-D continuous RD formulation to study the effect of varying pairs of decay and reaction coefficients $c_u, c_v, c_w, c_1, \dots, c_9$ on pattern generation. The corresponding 55 simulations are displayed as gradient plots (Figure S5) in which two coefficients are varying in space. For example, in the simulation labeled $c_i - c_j$, the variation of coefficients of variation is defined as

$$c_i(x) = c_i + \left(\frac{|c_i|}{x_{max} - x_{min}} \right) \left(x - \frac{(x_{min} + x_{max})}{2} \right)$$

$$c_j(y) = c_j + \left(\frac{|c_j|}{y_{max} - y_{min}} \right) \left(y - \frac{(y_{min} + y_{max})}{2} \right)$$

where $x_{min}, x_{max}, y_{min}$ and y_{max} are the domain limits and, c_i and c_j are the coefficient values proposed in Manukyan et al.¹², i.e.,

$$\begin{aligned} c_u &= 0.020 & , & & c_v &= 0.025 & , & & c_w &= 0.06, \\ c_1 &= -0.04 & , & & c_2 &= -0.056 & , & & c_3 &= 0.382, \\ c_4 &= -0.05 & , & & c_5 &= 0.0 & , & & c_6 &= 0.25, \\ c_7 &= 0.016 & , & & c_8 &= -0.03 & , & & c_9 &= 0.24 \end{aligned}$$

To produce these plots, we neglect domain growth and set the other coefficients to the values proposed in Manukyan et al.¹², i.e.,

$$D_u = D_v = 1.125, D_w = 12.5, P = 0.00889$$

All simulations start from the HSS, perturbed with small random fluctuations, and are run to steady-state. In most of the gradient plots, a similar diversity of patterns appears, suggesting that one might optimize the model with only a subset of parameters. To test this conjecture, we set up optimization experiments in which parameters are added iteratively to \mathbb{S} . First, we restrict \mathbb{S} to q and r and run the optimisation loop (all other parameters are set to the values proposed in Manukyan et al.¹²). Then, we repeat the optimization loop for all possible subsets corresponding to the inclusion of one additional parameter in \mathbb{S} . The parameter whose addition generates the lowest value of the objective function is permanently added to \mathbb{S} . We repeat this iterative process of adding the next best parameter until no reduction of the objective function is observed. Table S1 shows the objective function values obtained during the iterative addition process performed for TL1: the columns indicate the number of parameters in \mathbb{S} , and the rows are ordered by parameters iteratively identified as most useful (because exhibiting the lowest value, shown in bold, of the objective function). Table S1 indicates that the optimal choice of subset \mathbb{S} for TL1 is

$$\mathbb{S} = \{[c_v, D_u, D_w, q, r] : c_v \in \mathbb{R}, D_u, D_w, q, r \in \mathbb{R}^+\}$$

Note that we perform this identification of best subset \mathbb{S} independently for each individual of each species. After performing Bayesian optimisation (again, independently for each individual of each species, see above) on this parameter vector, we obtain the optimized dRD model parameters listed in Table S2. Note that an optimal subset \mathbb{S} was found for all individuals of all species after optimising only 3 to 5 parameters (all other parameters did not require modifying their values from Manukyan et al.¹²).

Uncertainty in skin thickness spatial distribution and production of 3D super-Gaussian heterogeneous networks

Although the acquisition of skin-scale surface micro-geometry captures the continuous variation of skin height, it does not capture the variation of the bottom boundary of the domain. In other words, we assume that the thinning of the skin at scale borders is the same at all borders of all scales, hence, the use of a single parameter P for the reduction of RD diffusion coefficients at all one-dimensional edges¹² in the 2D-cRD model. To quantify how much real skin geometry deviates from this assumption, we perform 3D geometry reconstruction of a patch of dorsal skin for each of the five species. For example, the patch for ocellated lizard comprises 24 scales (10 green and 14 black) and was reconstructed using 707 images (4082 × 3072 pixels each with 1 pixel = 1.2 μm) acquired with high-resolution episcopic microscopy (HREM^{14,23}). Sections generated by HREM are intrinsically-aligned, making the volume rendering straightforward. We then perform image analyses¹⁴ to identify the top and bottom boundaries of the skin as well as the positions occupied by melanophores. The former allows defining the full skin 3D domain whereas the latter provides an estimate of the restricted 3D domain populated by chromatophores. Indeed, as RD components are associated to chromatophores themselves, it is likely that relevant statistics on skin thickness apply to the restricted domain rather than to the whole skin depth.

The HREM data allows us to extract, for a dorsal patch of skin in each species, the spatial distributions of skin thickness (d) and of thickness d_m in the domain restricted to chromatophores. We then compute the mean distance S among neighboring scales at their highest point, as well as the means among the top-surface highest point of scales (\bar{h}_c) and among the heights of edges (\bar{h}_e). We then construct a ‘noisy’ 3D lattice (an example is shown in Figure 5B for *T. lepidus*) in which heights h_c and h_e (i.e., at the centers and at the edges of the super-Gaussian scales, respectively) vary with standard deviations σ_c and σ_e . Starting from the HREM-observed

average values of S , \bar{h}_c and \bar{h}_e , we optimise \bar{h}_c , \bar{h}_e , ρ , σ_c and σ_e to obtain a depth-map histogram of the noisy lattice (e.g., Figure 5C for *T. lepidus*), hence, a skin thickness distribution, highly similar to that observed in the HREM data. We then construct 1,000 domains of 64 super-Gaussian scales, represented as 3D curvilinear structured grids, by keeping S and ρ constant, but sampling h_c (for each bump) and h_e (for each edge) from the Normal distributions $\mathcal{N}(\bar{h}_c, \sigma_c)$ and $\mathcal{N}(\bar{h}_e, \sigma_e)$ with optimized parameters. Finally, using the mapping $d_m(d)$ discussed above, we compute the bottom boundary of each domain by subtracting d_m from the top boundary, hence, producing models of the skin domain restricted to chromatophores (turquoise volume illustrated in Figure 5B for *T. lepidus*). HREM data acquisition, optimization of parameters \bar{h}_c , \bar{h}_e , ρ , σ_c , as well as sampling of h_c and h_e from the Normal distributions $\mathcal{N}(\bar{h}_c, \sigma_c)$ and $\mathcal{N}(\bar{h}_e, \sigma_e)$, are performed independently for each species. We also generate for each species a reference hexagonal 3D lattice of identical super-Gaussian bumps, with super-Gaussian parameters σ and ρ , and with the centers of all bumps and all the edges set to optimized \bar{h}_c and \bar{h}_e , respectively.

Color measurement uncertainty and Lyapunov spectrum analysis

One powerful approach to estimate the degree of sensitivity of a dynamical system is to compute how fast two trajectories diverge when they initially differ by an arbitrary small difference. To this aim, we first define $\mathbf{C}^{\text{ref}} = (\mathbf{C}_1, \dots, \mathbf{C}_{n_s})$ as the reference $3 \times n_s$ matrix containing the states (in CIELAB colors) of all scales of the observed juvenile lizard. We then generate 2,000 ‘noisy initial conditions’ by adding random noise to the reference matrix, i.e., $\mathbf{C} = \mathbf{C}^{\text{ref}} + s\mathbf{C}'$, where the scalar s and the matrix \mathbf{C}' are randomly sampled from the uniform and 3-variate normal distributions, $\mathcal{U}(0, 1)$ and $\mathcal{N}(0, \Sigma^{\text{ref}})$, respectively. Note that Σ^{ref} is the covariance matrix of \mathbf{C}^{ref} . We then use a color transformation matrix (see STAR Methods, section Juvenile initial condition (IC) and color transformation matrix) to transfer each of these 2,000 noisy initial conditions to the space of RD variables for performing numerical simulations. For each initial condition, the state of the system at time τ is defined as:

$$\mathbf{S}_\tau = [u_1(\tau), v_1(\tau), w_1(\tau), \dots, u_{n_s}(\tau), v_{n_s}(\tau), w_{n_s}(\tau)]$$

and we study its time evolution as a trajectory in the $\mathbb{R}^{3 \times n_s}$ phase space parametrized by the time variable $\tau_0 \leq \tau \leq \tau_f$, where τ_0 and τ_f are the juvenile (initial condition) and final (adult) time points, respectively. We then measure, at each time τ , the Euclidean distance $\delta_\tau = \|\mathbf{S}_\tau - \mathbf{S}_\tau^{\text{ref}}\|$ between the state of the system (\mathbf{S}_τ) and the state ($\mathbf{S}_\tau^{\text{ref}}$) of the reference (i.e., the latter is the trajectory starting from the observed unperturbed juvenile colors). We then use the Lyapunov exponent λ as an estimate of the rate of divergence between the two trajectories:

$$\lambda = \lim_{\tau \rightarrow \infty} \lim_{\delta_0 \rightarrow 0} \frac{1}{\tau - \tau_0} \log \frac{\delta_\tau}{\delta_0}$$

where $\delta_0 = \delta_{\tau_0}$. Figure 6A shows, for each of the 2,000 simulations, the trajectory of $\log(\delta_\tau/\delta_0)$ as a function of time. The Lyapunov exponent is then simply derived as the slope of the curve (computed with least-square fit) for early time points, i.e., before saturation of δ_τ . The positive value of the mean (\pm SD) Lyapunov exponent ($\lambda = 0.0287 \pm 0.0004$ for TL1) confirms that the system is unstable.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed with MATLAB R2021a. Statistical details can be found in the figure legends, results section and the STAR Methods.

Current Biology, Volume 32

Supplemental Information

**Modeling convergent scale-by-scale skin color
patterning in multiple species of lizards**

Ebrahim Jahanbakhsh and Michel C. Milinkovitch

Current Biology, Volume 32

Supplemental Information

**Modeling convergent scale-by-scale skin color
patterning in multiple species of lizards**

Ebrahim Jahanbakhsh and Michel C. Milinkovitch

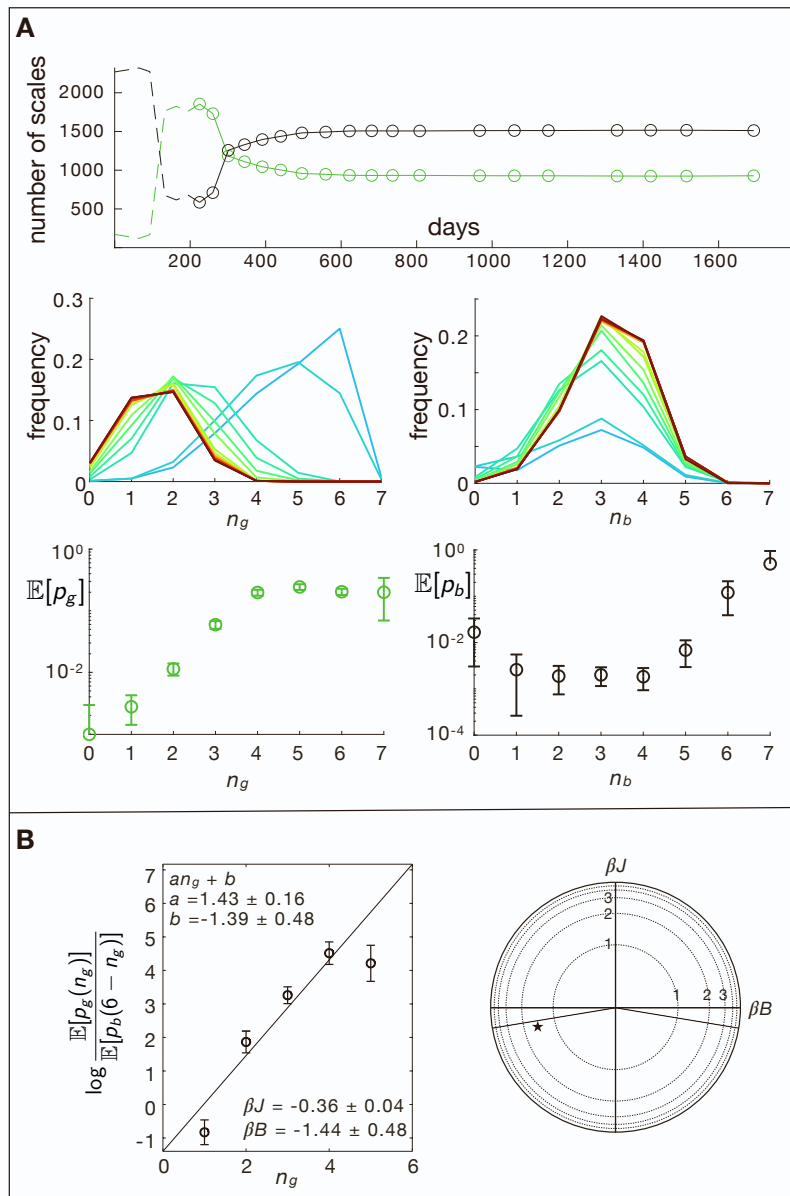


Figure S1 | Neighbourhood statistics and stochastic models for ocellated lizard; related to Figure 2 – (A) Top panel: time history of the number of green (green circles) and black (black circles) scales observed in ocellated lizard individual TL1; scale colours are thresholded to green or black by applying K-mean clustering. Middle panel: relative frequencies of green (left) and black (scales) as a function of $n_{g/b}$, *i.e.*, the number of isochromatic direct neighbours; lines of different colours correspond to different observed time-points (*i.e.*, circles in top panel), from blue (juvenile state) to red (adult). Lower panel: expected probability of colour change ($\mathbb{E}[p_{g/b}]$) for green/black scales as a function of $n_{g/b}$ (circles); error bars indicate the 0.95% confidence interval corresponding to the Binomial distribution. (B) Left panel: mean values (\pm SD) of $\log(\mathbb{E}[p_g(n_g)]/\mathbb{E}[p_b(6-n_g)])$ (using the data from (A)) with the best affine fit ($an + b$) and corresponding Lenz-Ising model parameters βB and βJ (non-Glauber dynamics¹⁴). Right panel: optimised Lenz-Ising model parameters (black star) at finite temperature, for individual TL1, plotted in polar coordinates with radial coordinate mapped to a finite range. Oblique solid lines indicate the separation between ferromagnetic and anti-ferromagnetic regions in the triangular Lenz-Ising model.

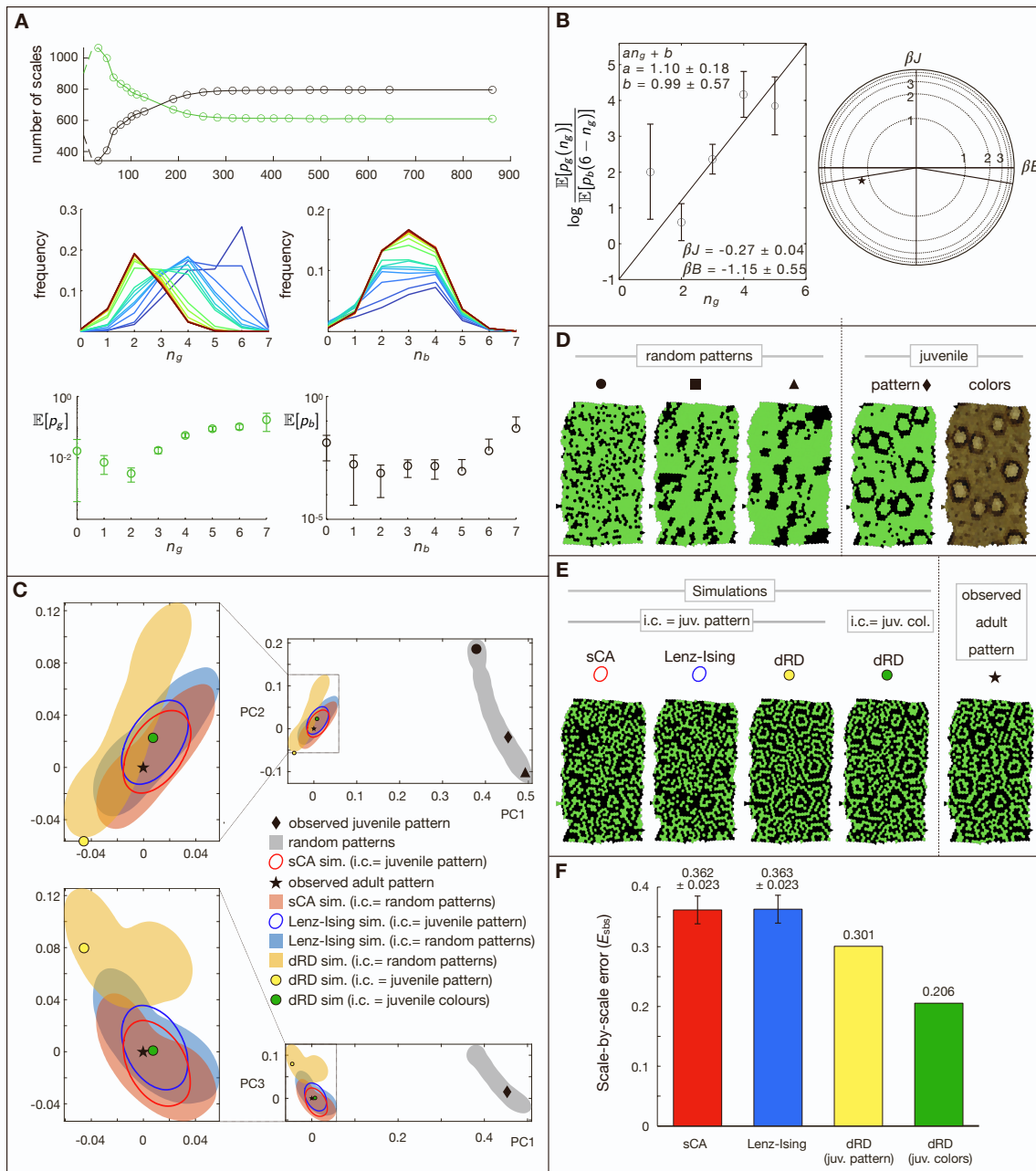


Figure S2 | Neighbourhood statistics and their predictability in ocellated lizard individual TL2; related to Figures 2 and 4 – (A) Same analyses as in figure S1A but for individual ocellated lizard TL2. (B) Same analyses as in figure S1B but for individual ocellated lizard TL2. (C) Same analyses as in main Figure 2 but for individual ocellated lizard TL2. (D) Different initial conditions used for simulations; their localisation in PC1-PC2-PC3 space are shown with the corresponding geometrical symbols. (E) Adult patterns simulated with different initial conditions (i.c.): juvenile (juv.) pattern (= scale colours thresholded to green or black) and juvenile colours (col.) are both shown in D. (F) Histogram comparing scale-by-scale errors (mean \pm SD) of adult patterns simulated with different models.

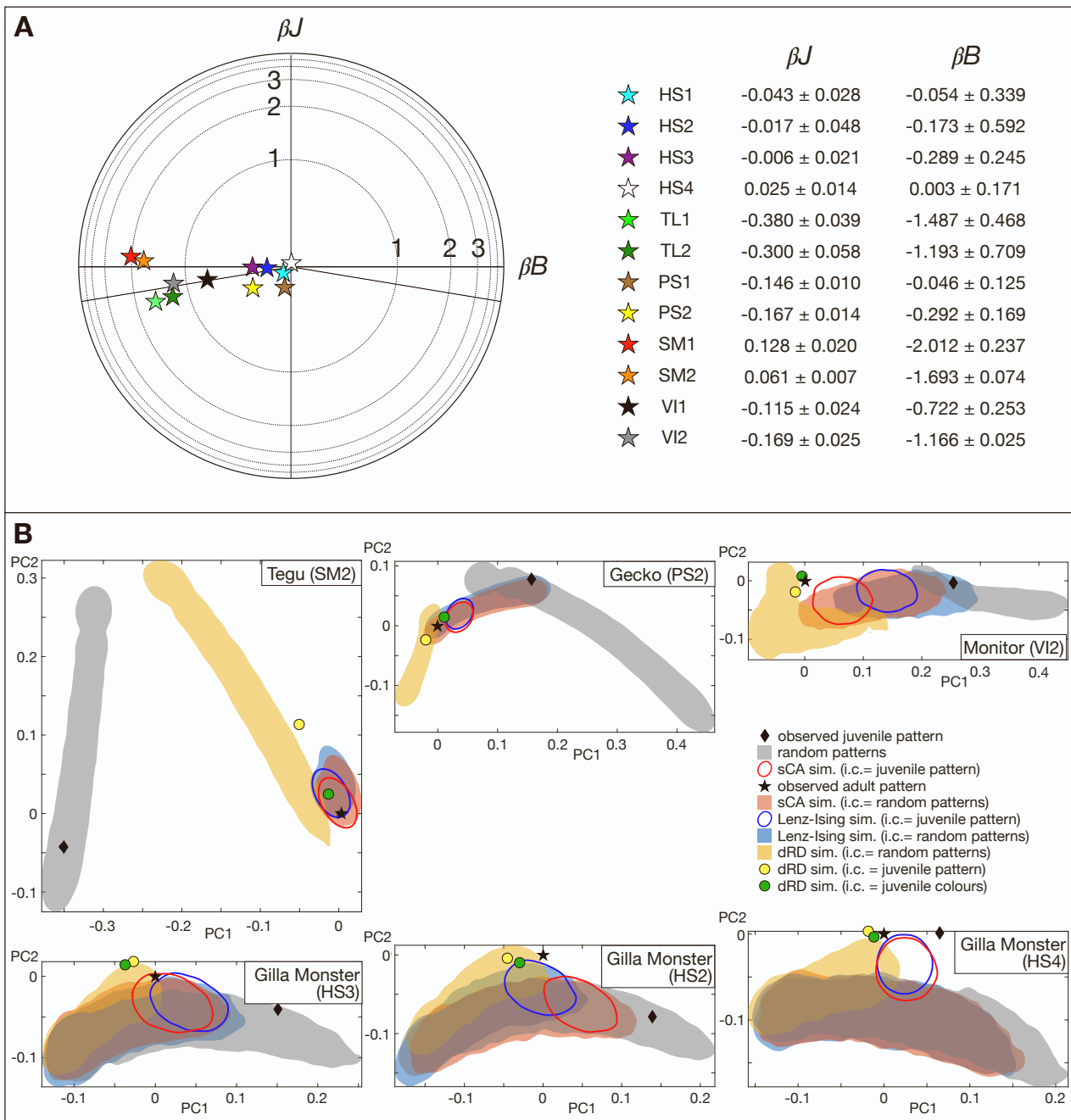


Figure S3 | Lenz-Ising parameters and prediction of neighbourhood statistics in four other species; related to Figure 3 — (A) Optimised Lenz-Ising model parameters (stars) for all individuals of all species plotted in polar coordinates with radial coordinate mapped to a finite range. Oblique solid lines indicate the separation between ferromagnetic and anti-ferromagnetic regions in the triangular Lenz-Ising model. **(B)** Projections on the PC1-PC2 plane of the 16D nearest-neighbour error vectors (in comparison to the corresponding observed adult pattern, black stars) of patterns simulated with sCA (red ellipse and red shading), Lenz-Ising (blue), and dRD (yellow) models. Red ellipses, blue ellipses and yellow spots show adult patterns simulated from the corresponding observed juvenile patterns (black diamonds), whereas red, blue and yellow shadings show adult patterns simulated from random patterns (grey areas) as initial condition. The green spots show the adult dRD-patterns simulated from the juvenile colours in the corresponding individual. Ellipses and border of shadings indicate 1% density isolines.

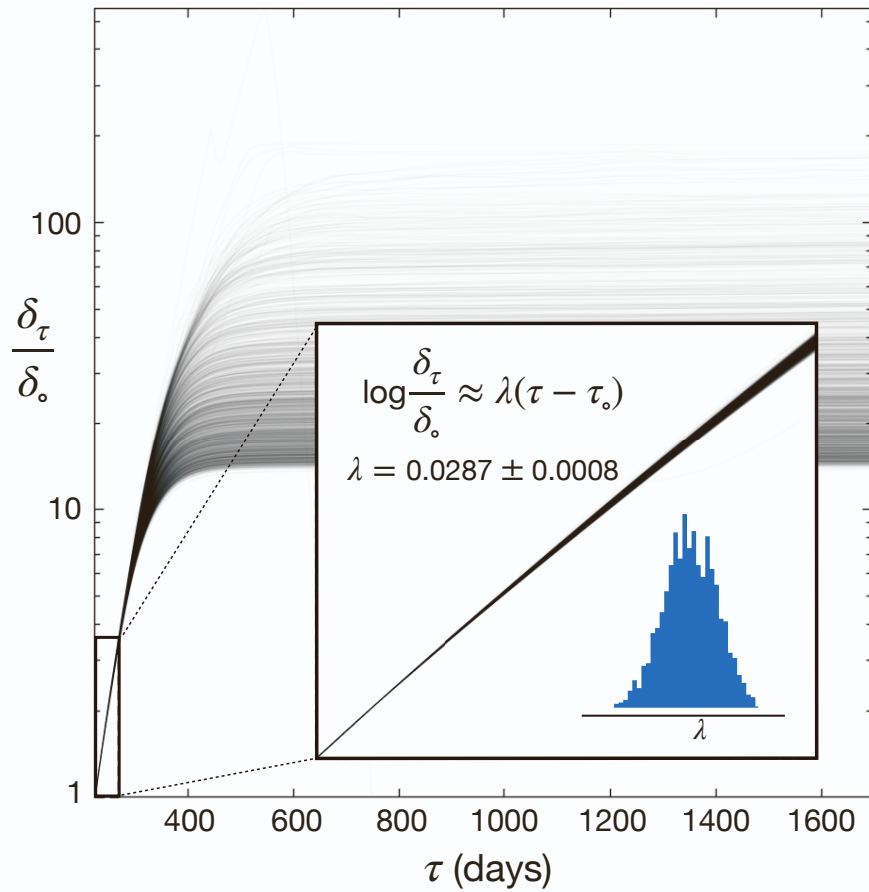


Figure S4 | Lyapunov spectrum analysis and colour measurement uncertainty in ocellated lizard individual TL1; related to Figure 6 and Data S1B — Time-evolution trajectories of 2,000 ‘perturbed’ simulations, *i.e.*, started at random small colour differences (E_{sbs}°) from the observed juvenile colours. E_{sbs}° is then translated to the Euclidean distances (δ_0) in RD space; δ_τ = distance at time τ between the state of each perturbed simulation and the state of the reference trajectory (starting from observed juvenile state); (Data S1B). The positive value of the mean (\pm SD) Lyapunov exponent (λ = slope of $\log(\delta_\tau/\delta_0)$ for early time points) confirms the instability of the system. The blue histogram in the inset indicates the distribution of λ across all simulations.

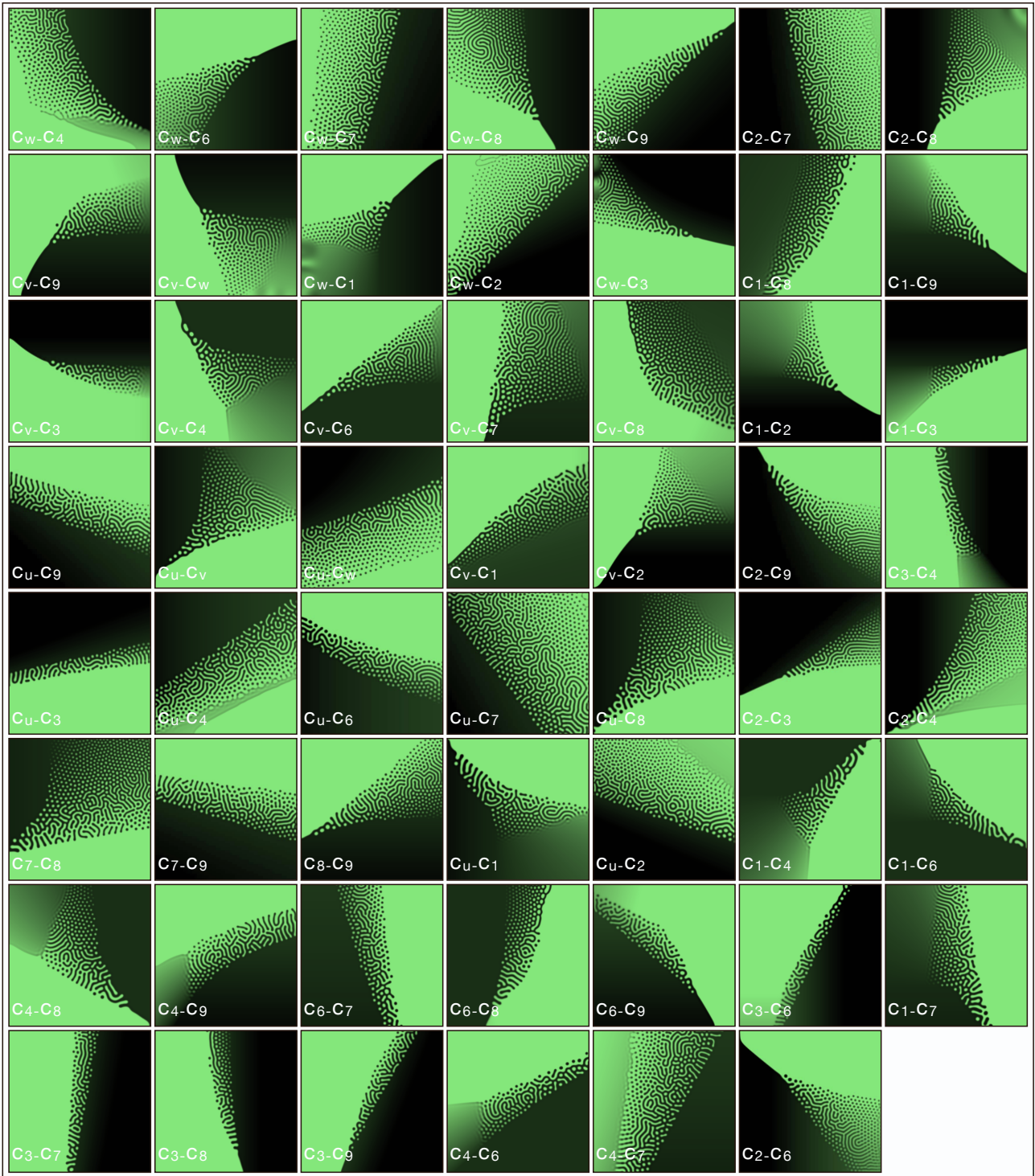


Figure S5 | Pairwise parameter gradient plots; related to Figure 4 and STAR methods
 — Different patterns generated by linear variation of RD model parameters. First and second varying parameters (white font) correspond to the horizontal and vertical directions, respectively. The limit of variations in each direction is $\pm 50\%$ of the absolute values given in Eq. (35).

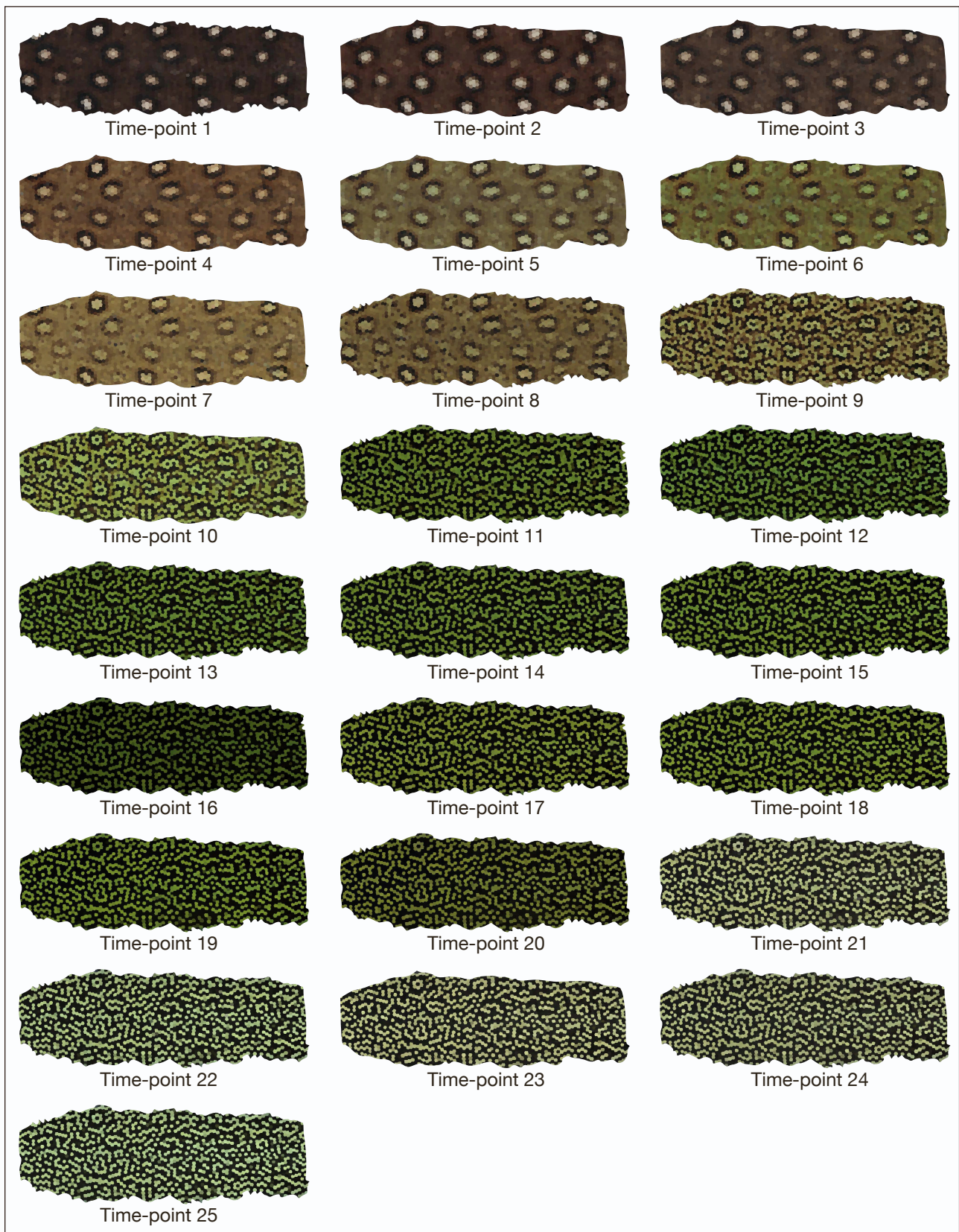


Figure S6 | Time evolution of observed scale colours of ocellated lizard; related to Figure 1B-D and STAR methods – Scale colours at 25 time-points for the patch of skin analysed in ocellated lizard TL1 (cf. Figure 1B). The colour of each polygonal scale is set to the mean albedo of its pixels at the corresponding time-point (see STAR methods).

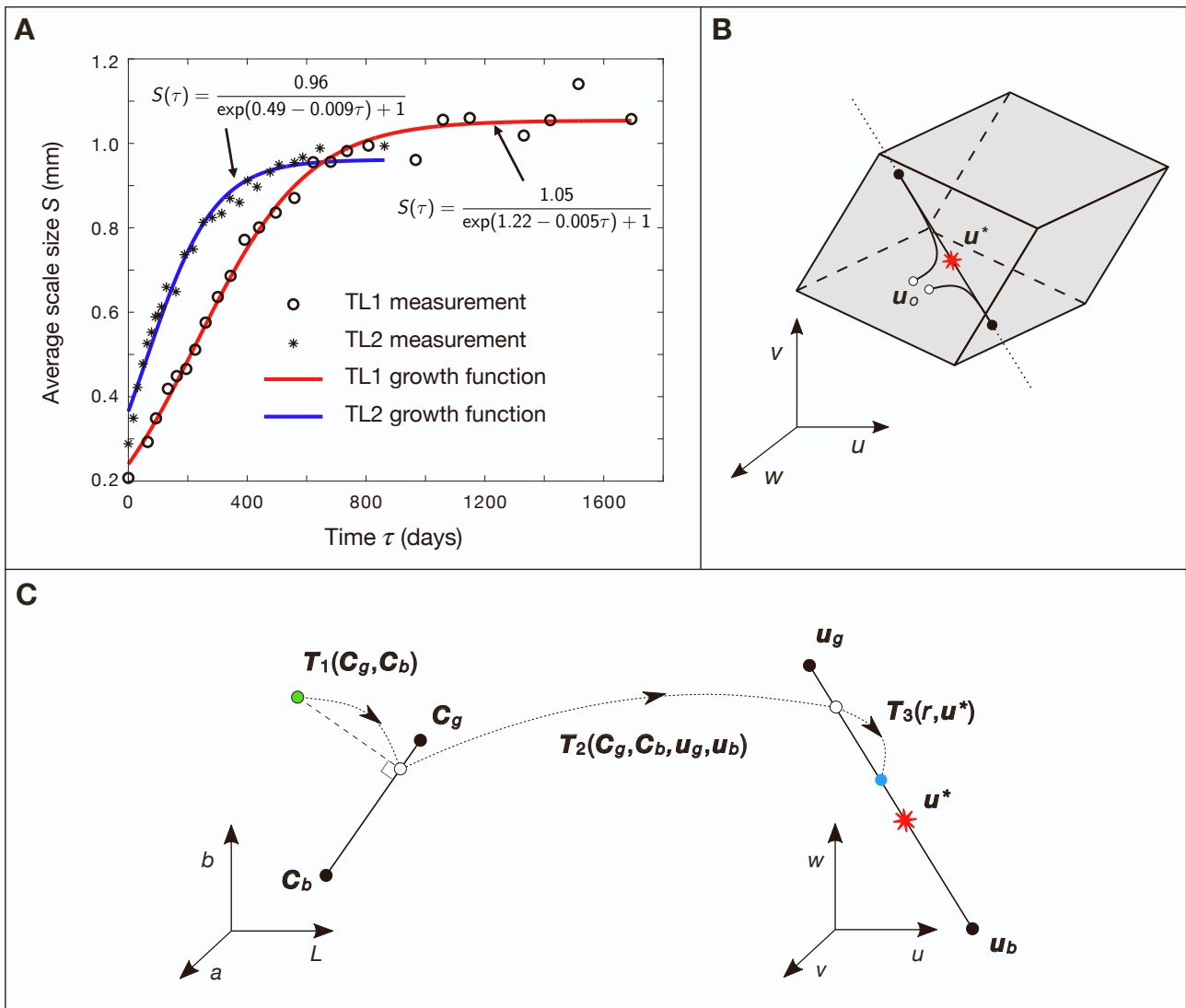


Figure S7 | Growth model and colour transformation to RD variables; related to STAR methods – (A) The time history of mean scale size (among edge lengths of all scales) for ocellated lizards TL1 and TL2. Solid lines show the best fitted logistic functions. (B) Schematic trajectories of scales in u, v, w space shown as curved lines; trajectories start from \mathbf{u}_0 (white circles) close to the HSS (red star) and approach the dominating eigenvector line (Eq. (29)) shown as a solid black line which extremities (black circles) intersect the grey planes described by Eq. (24). (C) Transformation between CIELAB colour space and u, v, w space using matrices T_1 , T_2 , and T_3 ; the green dot indicates, for a given scale, its observed juvenile colour and the blue dot, located on the $\mathbf{u}_g - \mathbf{u}_b$ line (*i.e.*, the dominating eigenvector line), indicates the corresponding initial RD component values used for simulations.

dimensions =	2	3	4	5	6
q, r	0.2448	—	—	—	—
c_v	—	0.1774	—	—	—
$D_{u,v}$	—	0.2373	0.1167	—	—
D_w	—	0.2308	0.1774	0.1086	—
c_u	—	0.2257	0.1774	0.1103	0.1086
c_w	—	0.2246	0.1774	0.1143	0.1086
c_1	—	0.1809	0.1774	0.1167	0.1086
c_2	—	0.2268	0.1774	0.1167	0.1086
c_3	—	0.2129	0.1774	0.1167	0.1086
c_4	—	0.2247	0.1774	0.1167	0.1086
c_5	—	0.2264	0.1774	0.1167	0.1086
c_6	—	0.2123	0.1774	0.1167	0.1086
c_7	—	0.2252	0.1774	0.1167	0.1086
c_8	—	0.1812	0.1774	0.1167	0.1086
c_9	—	0.2121	0.1774	0.1167	0.1086

Table S1 | Optimised values, for individual TL1, of the objective function f (equation 34 in STAR methods) during iterative addition of different parameters (left column) to \mathbb{S} ; related to STAR methods — We start by optimising 2 free parameters (q and r), while keeping all others at the values of ref. 12 in the main text, and obtain an f value of 0.2448. We then make all combinations of q , r and i (column 3), where i is each of the remaining 14 reaction, decay, and diffusion parameters. For each i , we optimise q , r and i , and select the parameter i that gives the best value of f (here, 0.1774 for c_v vs. higher values for all other parameters). We then make all combinations of q , r , c_v and i , where i is each of the remaining 13 parameters (column 4). The procedure is iterated until no improvement in the value of function f is obtained. In this example, (individual TL1), the optimal value is obtained after optimising 5 parameters highlighted by shaded cells.

	q [-]	r [-]	c_u [d ⁻¹]	c_v [d ⁻¹]	c_w [d ⁻¹]	$D_{u,v}$ [d ⁻¹ l ²]	D_w [d ⁻¹ l ²]	c_1 [d ⁻¹]	c_2 [d ⁻¹]	c_3 [z d ⁻¹]	c_4 [d ⁻¹]	c_5 [d ⁻¹]	c_6 [z d ⁻¹]	c_7 [d ⁻¹]	c_8 [d ⁻¹]	c_9 [z d ⁻¹]
TL1	2.9	0.23	0.02	0.03	0.06	0.07	21.8	-0.04	-0.056	0.382	-0.05	0	0.25	0.016	-0.03	0.24
TL2	2.25	0.33	0.02	0.025	0.06	0.33	22.9	-0.04	-0.056	0.382	-0.05	0	0.25	0.016	-0.03	0.239
PS1	3.15	0.63	0.02	0.022	0.06	1.125	12.5	-0.04	-0.056	0.384	-0.05	0	0.25	0.016	-0.03	0.24
PS2	4.9	0.85	0.02	0.025	0.06	1.125	21.5	-0.04	-0.056	0.382	-0.05	0	0.25	0.015	-0.03	0.24
SM1	4.8	0.95	0.02	0.025	0.06	0.055	12.5	-0.04	-0.056	0.382	-0.05	0	0.212	0.016	-0.03	0.24
SM2	1.04	0.93	0.02	0.025	0.06	0.06	12.5	-0.04	-0.056	0.382	-0.05	0	0.222	0.016	-0.03	0.24
HS1	1.9	0.07	0.02	0.025	0.06	0.27	12.5	-0.04	-0.056	0.382	-0.05	0	0.238	0.016	-0.03	0.24
HS2	1.7	0.23	0.018	0.025	0.06	0.43	12.5	-0.04	-0.056	0.382	-0.05	0	0.25	0.016	-0.03	0.24
HS3	2.5	0.16	0.02	0.025	0.06	0.21	12.5	-0.04	-0.056	0.382	-0.05	0	0.238	0.016	-0.03	0.24
HS4	1.13	0.23	0.02	0.025	0.06	0.27	12.5	-0.04	-0.056	0.382	-0.05	0	0.25	0.016	-0.03	0.24
VI1	1.1	0.05	0.02	0.025	0.06	0.11	12.5	-0.04	-0.056	0.382	-0.053	0	0.25	0.016	-0.039	0.24
VI2	2.7	0.61	0.02	0.025	0.06	0.43	12.5	-0.037	-0.056	0.382	-0.05	0	0.228	0.016	-0.03	0.24

Table S2 | RD parameters for all individuals; related to STAR methods – Shaded cells indicate values optimised using Bayesian machine-learning following the procedure illustrated in table S1. Other values are from ref. 12 in the main text. Units are indicated between brackets ([-] = dimensionless, d = day, $l \approx 86 \mu\text{m}$ and z = dimensions of the corresponding u , v and w RD variables).



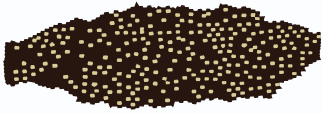


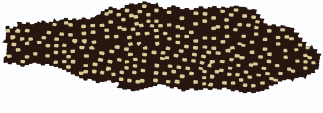



		Model parameters		
		SM1	SM2	VI1
Initial condition	SM1	 (0.037)	 (0.065)	 (0.296)
	SM2	 (0.070)	 (0.056)	 (0.281)
	VI1	 (0.407)	 (0.380)	 (0.077)

Table S3 | Exchanging sets of RD parameters among individuals within and between species (black and white tegu *versus* mangrove monitor lizard); related to Figure 3 — Colour patterns of two tegus ($E_{16D}= 0.081$ for SM1 *versus* SM2) and one monitor lizard (VI1) simulated after exchanging their optimised RD parameters. Numbers in parentheses indicate E_{16D} errors when compared to the observed adult pattern.

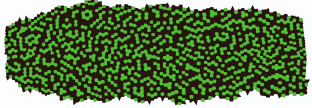
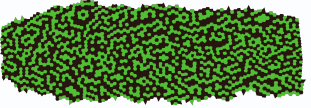
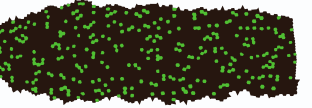
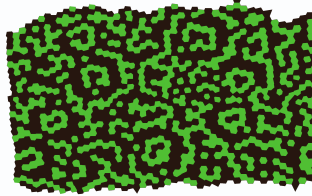
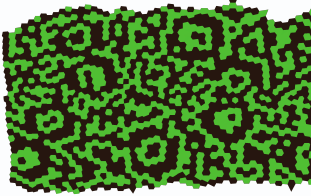
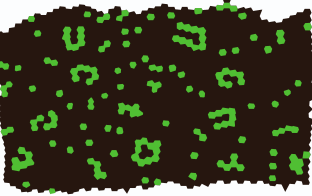



		Model parameters		
		TL1	TL2	SM1
Initial condition	TL1	 (0.029)	 (0.146)	 (0.526)
	TL2	 (0.078)	 (0.044)	 (0.520)
	SM1	 (0.372)	 (0.397)	 (0.037)

Table S4 | Exchanging sets of RD parameters among individuals within and between species (ocellated lizards vs. black and white tegu); related to Figures 2 and 3 — Colour patterns of two ocellated lizards ($E_{16D}= 0.16$ for TL1 versus TL2) and one tegu (SM1) simulated after exchanging their optimised RD parameters. Numbers in parentheses indicate E_{16D} errors when compared to the observed adult pattern.