

Evolutionary Branching in a Finite Population: Deterministic Branching vs. Stochastic Branching

Joe Yuichiro Wakano^{*,†,1} and Yoh Iwasa[‡]

^{*}Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, Kawasaki 214-8571, Japan, [†]Japan Science and Technology, Precursory Research for Embryonic Science and Technology, Tokyo 102-8666, Japan, and [‡]Department of Biology, Kyushu University, Fukuoka 812-8581, Japan

ABSTRACT Adaptive dynamics formalism demonstrates that, in a constant environment, a continuous trait may first converge to a singular point followed by spontaneous transition from a unimodal trait distribution into a bimodal one, which is called “evolutionary branching.” Most previous analyses of evolutionary branching have been conducted in an infinitely large population. Here, we study the effect of stochasticity caused by the finiteness of the population size on evolutionary branching. By analyzing the dynamics of trait variance, we obtain the condition for evolutionary branching as the one under which trait variance explodes. Genetic drift reduces the trait variance and causes stochastic fluctuation. In a very small population, evolutionary branching does not occur. In larger populations, evolutionary branching may occur, but it occurs in two different manners: in deterministic branching, branching occurs quickly when the population reaches the singular point, while in stochastic branching, the population stays at singularity for a period before branching out. The conditions for these cases and the mean branching-out times are calculated in terms of population size, mutational effects, and selection intensity and are confirmed by direct computer simulations of the individual-based model.

IN addition to models using quantitative genetics (Lande 1981; Barton and Turelli 1991; Iwasa *et al.* 1991; Pomiankowski *et al.* 1991), adaptive dynamics formalism has revealed diverse behaviors of evolutionary dynamics associated with continuous traits in various organisms (Metz *et al.* 1992; Dieckmann and Law 1996; Geritz *et al.* 1997, 1998). This approach allows us to analyze the evolution of many ecological traits for which frequency-dependent selection is prevalent without specifying the derlying genetic systems controlling the trait. Even in a perfectly constant environment, a population may show intriguing temporal behavior. For example, if the trait evolves by accumulation of mutations of small magnitude and if the fitness is frequency dependent, a continuous trait may show convergence to a singular point followed by spontaneous splitting of a unimodal trait distribution into a bimodal (or multimodal) one, referred to as “evolutionary branching” (Metz *et al.* 1992, 1996; Geritz *et al.* 1997). Evolutionary branching is predicted to occur

at an evolutionarily singular point that is approaching stable (or convergence stable, CS) but not evolutionarily stable (ES), and it is actually observed in individual-based simulations in many models for the evolution of ecological traits.

In evolutionary branching, an evolving population spontaneously changes its trait distribution from unimodal to bimodal (or multimodal). Although evolutionary branching has been regarded as a model of sympatric speciation by some authors (Dieckmann and Doebeli 1999; Doebeli and Dieckmann 2000), the relationship between evolutionary branching in adaptive dynamics formalism and speciation events has been the focus of much debate (for review, see Waxman and Gavrillets 2005).

Most previous analyses of adaptive dynamics have assumed an infinite population size and that the fitness landscape acts as the deterministic driving force of evolution. However, in a finite population, stochastic fluctuations are unavoidable. For example, an individual with a lower fitness may be able to produce more offspring just by chance. Thus, a finite population can evolve against the selection gradient. As a result, stochasticity allows a population to jump from a low local peak to a higher peak. Hence, stochastic fluctuation can be an important factor in the outcome of evolution. The importance of stochasticity, or

Copyright © 2013 by the Genetics Society of America
doi: 10.1534/genetics.112.144980

Manuscript received August 17, 2012; accepted for publication October 22, 2012
Available freely online through the author-supported open access option.

¹Corresponding author: Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, 1-1-1 Higashi Mita, Tama-ku, Kawasaki 214-8571, Kanagawa, Japan.
E-mail: joe@meiji.ac.jp

genetic drift, has been studied in detail in the context of quantitative population genetics (e.g., Lande 1976; Tazzyman and Iwasa 2000). On the other hand, in population genetics, random drift is known to decrease the genetic diversity of a population (e.g., Ewens 2004), which tends to slow down the action of natural selection. Hence, the effect of genetic drift is important to understand evolutionary dynamics, including evolutionary branching, in finite populations. In fact, Claessen *et al.* (2007) analyzed evolutionary branching in a model including three species and reported that genetic drift tends to cause a delay in evolutionary branching. However, a more detailed study of the relationship between evolutionary branching and genetic drift is needed.

In this article, we study the effect of stochastic fluctuation caused by the finiteness of population size (random genetic drift) on evolutionary branching. As an illustrative example, we adopt a model of a public goods game studied by Doebeli *et al.* (2004), who investigated evolutionary branching by using a corresponding individual-based simulation. The game model deals with a situation that is both biologically and socially interesting: a self-organized differentiation of players into a highly cooperative group and a noncooperative group. Concerning the genetics, we in effect assume that the population is haploid and that the focal trait is controlled by a locus with many alleles. By focusing on this simplest case, we can analyze the effect of random genetic drift on the evolutionary outcome most clearly. We first show the results of individual-based simulations, using different population sizes. We show that genetic drift is especially important when the population size is small and that evolutionary branching does not necessarily occur even when the standard theory of adaptive dynamics predicts that it should. We also derive deterministic and stochastic versions of the dynamics of the trait variance. We reveal that there are two different manners in which evolutionary branching occurs: in deterministic branching, branching may occur quickly after the population reaches the singular point, while in stochastic branching, the population stays at a singularity for a period before branching out suddenly. We derive the conditions and mean branching-out times for these cases and confirm our predictions by direct computer simulations of the individual-based model.

Evolutionary Branching in a Game in a Finite Population

Individual-based model

As a full stochastic model of a population under natural selection, mutation, and random genetic drift, we consider the following stochastic process. The population consists of N individuals reproducing asexually. Each individual has a genetically determined trait value $z \in [0, 1]$. At each discrete time step t , an individual i obtains a fecundity F_i from pairwise interactions with the other $N - 1$ individuals. Each interaction is a two-player game where a payoff given to an

individual i matched with an individual j is given by $f(z_i, z_j)$. Its fecundity F_i is denoted by

$$F_i = \sum_{j \neq i} f(z_i, z_j). \quad (1)$$

Our framework is valid for any payoff function $f(z_i, z_j)$. As an example, we adopt a model by Doebeli *et al.* (2004) in which trait z represents the amount of investment into public goods shared by both players. In this game, the payoff function is represented by

$$f(z_i, z_j) = 1 + B(z_i + z_j) - C(z_i), \quad (2a)$$

where the benefit function is

$$B(z_1 + z_2) = b_1(z_1 + z_2) + b_2(z_1 + z_2)^2, \quad (2b)$$

and the cost function is

$$C(z) = c_1 z + c_2 z^2. \quad (2c)$$

To keep the population size constant, the fitness of an individual with trait z_i is denoted as

$$w(z_i) = \frac{F_i}{(1/N) \sum_j F_j}. \quad (3)$$

When the population size is infinitely large, Doebeli *et al.* (2004) have shown that the singular strategy $z^* = (c_1 - b_1)/(4b_2 - 2c_2)$ is convergence stable when $Q_{CS} = 4b_2 - 2c_2 < 0$ and evolutionarily stable when $Q_{ES} = 2b_2 - 2c_2 < 0$. The singular strategy is referred to as an evolutionary branching point when $Q_{CS} < 0$ and $Q_{ES} > 0$. They have observed evolutionary branching in an individual-based simulation of a large population ($N = 10,000$).

In a finite population, the model includes random genetic drift, and the details of the updating rule and the order of different stages in population dynamics become important. Here, we assume a Wright–Fisher process. At each time step, fecundity is calculated for all N individuals. Then all individuals die and are replaced by their offspring, which are sampled multinomially, with the probability of an individual being the parent of each offspring individual proportional to the fecundity of the focal individual.

After reproduction, there is a small chance of mutation μ , and the mutant has a trait value slightly different from that of the parent. To be specific, it follows a normal distribution with a mean equal to the parent's and a variance σ^2 . Mutants are constrained to be between 0 and 1, which confines trait z within an interval $[0, 1]$.

Evolutionary branching depends on the population size

All simulations were run with parameter values for which a unique singular point exists that is convergence stable and evolutionarily nonstable (i.e., the evolutionary branching

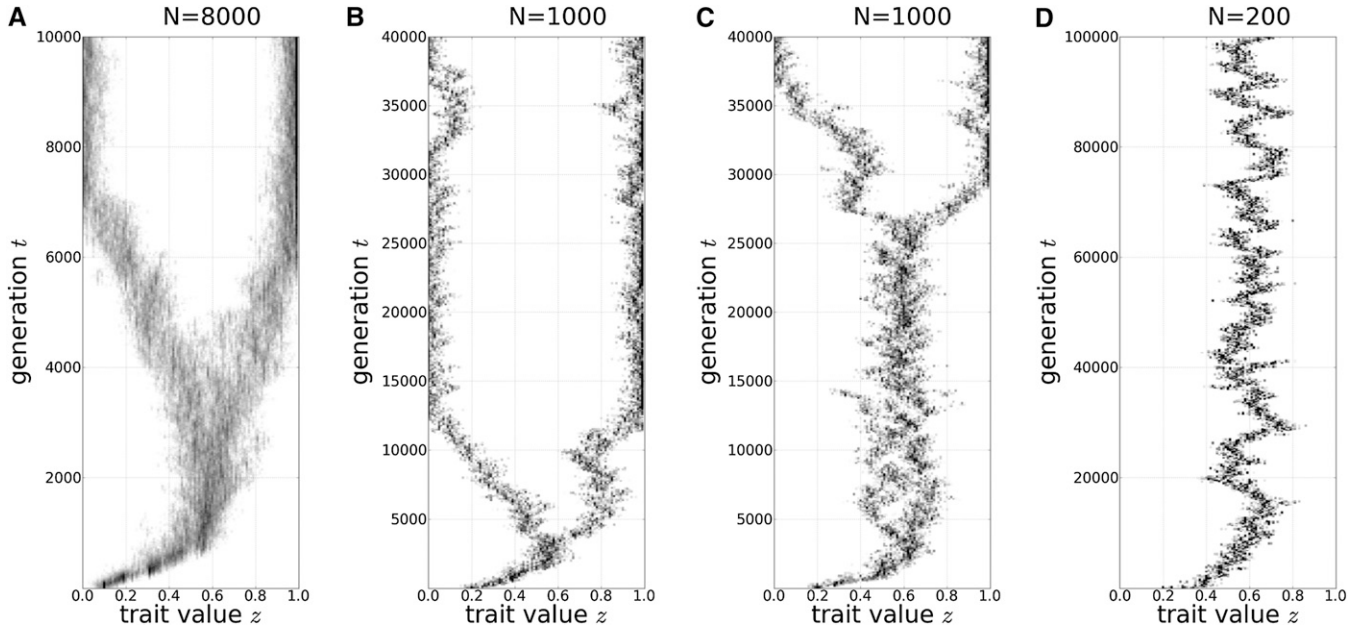


Figure 1 Evolutionary dynamics. (A) A simulation run with $N = 8000$. (B and C) Two different simulation runs with $N = 1000$. (D) A simulation run with $N = 200$. Parameters: $\mu = 0.01$, $\sigma = 0.02$, $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

point). Thus, evolutionary branching is expected when the population size is infinitely large.

When the population size is large ($N = 8000$), evolutionary branching occurs as soon as the trait evolves to reach the CS value z^* (Figure 1A). For an intermediate population size ($N = 1000$), the population lingers around the CS value, z^* , before it finally shows evolutionary branching. The waiting time until the evolutionary branching varies considerably among different simulation runs with the same parameter values (Figure 1, B and C). For a small population size ($N = 200$), we do not observe evolutionary branching and the population seems to stay around z^* , even though z^* is not evolutionarily stable (Figure 1D) (see also Wakano and Lehmann 2012).

The evolutionary branching dynamics are characterized by an abrupt increase or explosion of the variance of the focal phenotype (the trait variance). If all the parameters for the interactions are the same, a small population does not branch, an intermediately large population shows delayed branching, and a large population branches immediately. The intuitive reason for the smallness of the population size to suppress evolutionary branching is that genetic diversity is reduced by genetic drift every generation. In the following sections, we provide approximate dynamics for the trait variance to understand these simulation results.

Dynamics of Trait Variance

The full (individual-based) model is analytically intractable. Therefore, we derive a reduced simplified dynamical system here, which is tractable and gives predictions on the behavior

of the full model. One generation in the reduced model consists of two substeps.

Substep 1: Natural selection and mutation

Let $\varphi(z)$ denote a distribution of trait z , which is normalized to satisfy $\int \varphi(z) dz = 1$. When a fitness function $w(z)$ is given and selection and mutation are sufficiently weak, the trait distribution in the offspring pool after mutation, denoted by $\chi(z)$, can be approximated by the following replicator equation plus a mutation term,

$$\chi(z) = \varphi(z) + [w(z) - \bar{w}] \varphi(z) + \frac{\mu \sigma^2}{2} \frac{\partial^2}{\partial z^2} \varphi(z), \quad (4)$$

where \bar{w} is the population average of fitness values. The second term on the right-hand side represents the effect of selection and the third term represents the effect of mutation. The Taylor expansion of the fitness function around the average trait value \bar{z} is denoted as

$$w(z) = w(\bar{z}) + w_1(z - \bar{z}) + \frac{w_2}{2}(z - \bar{z})^2, \quad (5)$$

where the coefficients w_1 and w_2 are determined by the current state \bar{z} . The coefficient w_1 represents the selection gradient. The coefficient w_2 indicates stabilizing selection if it is negative and indicates disruptive selection if it is positive. The derivation of w_2 from the payoff function $f(z_i, z_j)$ is explained in Appendix A. For parameter values used in our simulations, we have $w_2 \approx 0.1$, implying that disruptive selection is operating. The population average and the variance of trait values in the offspring trait distribution $\chi(z)$ are respectively denoted as

$$Av_{\chi}[z] = \bar{z} + w_1 m_2 + \frac{w_2}{2} m_3, \quad (6a)$$

$$Var_{\chi}[z] = m_2 - (w_1^2 + \frac{w_2}{2}) m_2^2 + w_1 m_3 + \frac{w_2}{2} m_4 + \mu \sigma^2, \quad (6b)$$

where m_k denotes the k th centralized moment of $\varphi(z)$. The above approximation is derived by neglecting m_5 and higher-order moments (see *Appendix B*).

Substep 2: Random sampling (or random genetic drift)

From the offspring pool $\chi(z)$, N individuals are independently sampled to become adults. Hence, even for a fixed $\chi(z)$, different distributions of adult phenotypes are realized according to a probability distribution (probability measure) over all possible trait distributions. After sampling N individuals independently from $\chi(z)$, the expectation of the average trait value is identical to the average trait value in $\chi(z)$. In contrast, the expectation of the trait variance is decreased by a factor $(N - 1)/N$ (Ewens 2004). Hence, we have the following:

$$\begin{aligned} E[Av[Z]] &= Av_{\chi}[z] \\ E[Var[Z]] &= \frac{N-1}{N} Var_{\chi}[z]. \end{aligned} \quad (7)$$

Combining the two substeps, the expected changes in the population average and the variance in one generation are given by

$$E[\Delta Av[Z]] = w_1 m_2 + \frac{w_2}{2} m_3, \quad (8a)$$

$$E[\Delta Var[Z]] = \frac{N-1}{N} \left[m_2 + w_1 m_3 - w_1^2 m_2^2 + \frac{w_2}{2} (m_4 - m_2^2) + \mu \sigma^2 \right] - m_2. \quad (8b)$$

We are particularly interested in the dynamics of the variance of the trait distribution. To analyze Equation 8 further, we here assume that the trait distribution is close to a normal distribution whose mean and variance obey the dynamics given by Equation 8. For normal distributions, $m_3 = 0$ and $m_4 = 3m_2^2$. Using these relationships, we have the following recursive equations for the mean and variance of the trait.:

$$\Delta \bar{z} = w_1 m_2 \quad (9a)$$

$$\Delta m_2 = \frac{N-1}{N} (w_2 - w_1^2) m_2^2 - \frac{1}{N} m_2 + \frac{N-1}{N} \mu \sigma^2. \quad (9b)$$

The first, second, and third terms on the right-hand side of Equation 9b represent the effect of natural selection, the loss of diversity due to genetic drift, and the increase of diversity due to mutation, respectively. Without selection, the variance m_2 converges to $(N - 1)\mu\sigma^2$.

Condition for the Explosion of Trait Variance

As long as the selection gradient does not vanish ($w_1 \neq 0$), the mean trait value changes according to Equation 9a, which is a standard result of population genetics (Ewens 2004) and corresponds to the canonical equation in adaptive dynamics theory (Dieckmann and Law 1996; Champagnat *et al.* 2006). If a singular point z^* , where the selection gradient vanishes ($w_1 = 0$), is convergence stable, then the system approaches the singular point with time (Geritz *et al.* 1997). Once the mean trait (*i.e.*, the first moment of trait distribution) reaches the singularity, it will not change any longer. However, the trait variance (*i.e.*, the second centralized moment of the distribution) may change with time. See Figure 2 for a schematic illustration. Here we focus on the recursion for the trait variance,

$$\Delta m_2 = \frac{N-1}{N} w_2 m_2^2 - \frac{1}{N} m_2 + \frac{N-1}{N} \mu \sigma^2, \quad (10)$$

which indicates that the one-generational change in the trait variance is equal to the sum of three terms: the natural selection term, the genetic drift term, and the mutation term. The right-hand side of Equation 10 represents the “force” acting on the trait variance (Figure 3). Note that the orders of the three terms in Equation 10 are m_2^2 , m_2 , and 1, respectively. When the variance is small, the mutation (the third term) is dominant, and it increases the trait variance. As the variance increases, the genetic drift (the second term) increases its magnitude and it reduces the trait variance. The balance between these two processes would result in the equilibrium $m_2^* \approx N\mu\sigma^2$, which represents the variance determined by mutation–drift balance. When the trait variance is large, selection (the first term) becomes dominant. Selection decreases the trait variance when $w_2 < 0$ (stabilizing selection), but it increases the trait variance when $w_2 > 0$ (disruptive selection).

Combining these, we have three different cases, depending on the parameters.

Case i: When $w_2 < 0$, stabilizing selection is at work. Then the dynamics of the trait variance (Equation 10) have a single positive equilibrium

$$m_2^* = \frac{1 + \sqrt{1 - 4(N-1)^2 w_2 \mu \sigma^2}}{2(N-1)w_2},$$

which is globally stable, and the trait variance converges to this value for any initial value. At this equilibrium, the trait variance is maintained by mutation, which increases variance, and by genetic drift and stabilizing selection, which decrease variance. Especially when the stabilizing selection is weak, trait variance converges to the level maintained by mutation–drift balance ($m_2^* \approx (N - 1)\mu\sigma^2$) when $|w_2|$ is small.

Case ii: When $0 < w_2 < 1/4\mu\sigma^2(N-1)^2$, disruptive selection is at work, but it is not very strong. Then the dynamics of the trait variance have two positive equilibria: the smaller

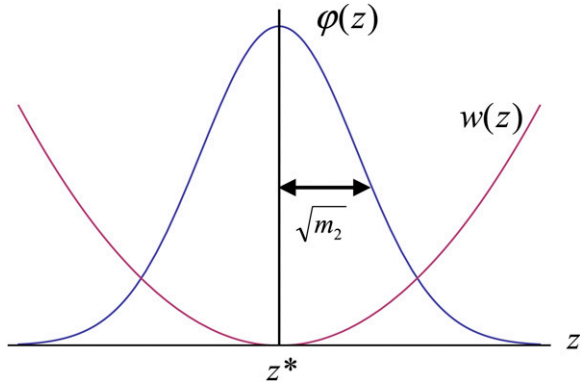


Figure 2 A schematic illustration of phenotype distribution $\varphi(z)$ and fitness function $w(z)$ when the mean trait value \bar{z} has reached the singular strategy z^* . The trait variance is $\text{Var}[\varphi(z)] = m_2$ so the standard deviation is $\sqrt{m_2}$.

one is stable, but the larger one is unstable. If the trait variance starts from a small value, it converges to the smaller equilibrium, where the trait variance is maintained mainly by the mutation–drift balance. If the trait variance starts from a larger value, then disruptive selection dominates, leading to the unlimited growth of the trait variance.

Case iii: When $w_2 > 1/4\mu\sigma^2(N-1)^2$, disruptive selection is strong. The dynamics of the trait variance no longer have a stable equilibrium. Starting from any initial value, the trait variance keeps increasing without limit.

Figure 4 shows a bifurcation diagram of the variance dynamics given by Equation 10, with the stabilizing/disruptive selection coefficient (w_2) as a bifurcation parameter. Figure 5 shows a bifurcation diagram in which the population size (N) is a bifurcation parameter. We have a similar bifurcation diagram where mutational effects ($\mu\sigma^2$) serve as a bifurcation parameter (not shown). We regard the explosion of the trait variance in the dynamics of Equation 10 as evolutionary branching. Then, these bifurcation diagrams show that evolutionary branching occurs as a result of a saddle-node bifurcation. Evolutionary branching is predicted to occur at a non-ES singular point when disruptive selection intensity, population size, mutation rate, and mutation step size are sufficiently large. Specifically, the deterministic dynamics of Equation 10 predict that evolutionary branching occurs when $4\mu\sigma^2(N-1)^2w_2 > 1$ holds.

Stochastic Calculation of the Explosion of Variance

The deterministic dynamics of the trait variance represented by Equation 10 indicate only the expectation of one-generational change. There must be stochasticity caused by the genetic drift. Results of an individual-based model, such as those illustrated in Figure 1, clearly show stochasticity. To account for stochasticity, the deterministic difference of Equation 10 needs to be modified by an additional term. In this section, we derive the stochastic differential equation

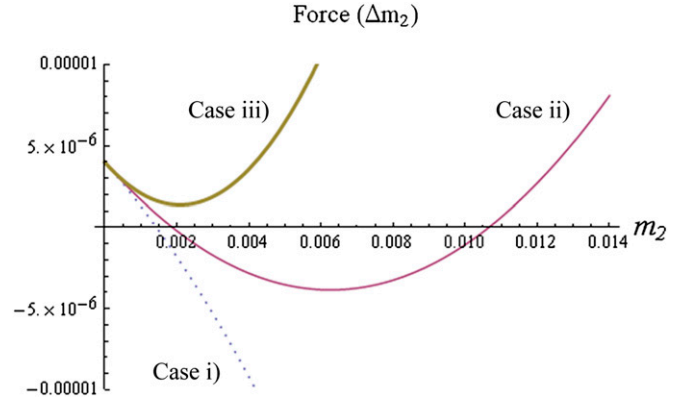


Figure 3 “Force” plotted against the trait variance based on Equation 10. Dotted, solid, and thick solid curves represent stabilizing selection ($w_2 = -0.2$), weak disruptive selection ($w_2 = 0.2$), and strong disruptive selection ($w_2 = 0.6$), respectively. They also correspond to cases i, ii, and iii in the main text. Parameters: $N = 400$, $\mu = 0.01$, $\sigma = 0.02$.

tion model considering the stochasticity of trait variance dynamics.

The deterministic dynamics of Equation 10 can be approximated by the ordinary differential equation (ODE)

$$\frac{d}{dt}m_2 = f(m_2), \quad (11a)$$

where

$$f(x) = \frac{(N-1)w_2x^2 - x + (N-1)\mu\sigma^2}{N}. \quad (11b)$$

If stochasticity has no correlation over generations, and if the paths can be approximated as continuous, then we can incorporate the Brownian motion term, and the differential Equation 11 becomes the stochastic differential equation (SDE)

$$dM_2 = f(M_2)dt + \sqrt{\frac{2}{N}}M_2dB, \quad (12)$$

where dB denotes Brownian motion and M_2 denotes a random variable representing the trait variance. $f(M_2)$ is the systematic change per generation in variable M_2 and is given by Equation 11b.

The second term on the right-hand side of Equation 12 represents stochasticity in trait variance caused by random genetic drift (see *Appendix C* for derivation). The stochastic fluctuation term depends on the trait variance itself (geometric Brownian motion), because the amplitude of fluctuation in the trait variance in each generation of offspring is large when the trait variance in each parental generation is large. The stochastic term also depends on population size, because fluctuation is averaged out in large populations. Hence, random genetic drift becomes less important in larger populations. The system is similar to but different from the

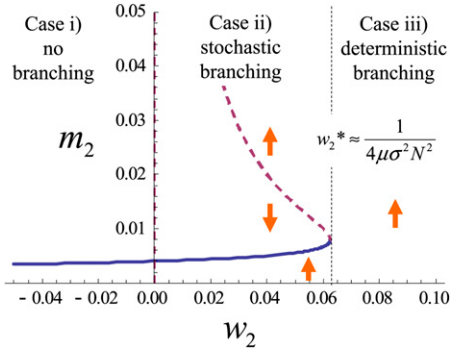


Figure 4 Stable (solid) and unstable (dashed) branches of the trait variance m_2 with disruptive selection intensity w_2 as a bifurcation parameter. Arrows indicate the direction of systematic change in m_2 . Parameters: $N = 1000$, $\mu = 0.01$, $\sigma = 0.02$.

motion of a particle moved by force shown in Figure 3 under heat noise.

Stationary distribution

The corresponding Fokker–Planck equation (or forward Kolmogorov equation) to Equation 12 is given by

$$\frac{\partial}{\partial t} p(m_2; t) = -\frac{\partial}{\partial m_2} [f(m_2) p(m_2; t)] + \frac{1}{2} \frac{\partial^2}{\partial m_2^2} \left[\frac{2}{N} m_2^2 p(m_2; t) \right] \quad (13)$$

(Feller 1971), where $p(m_2; t) dm_2 = \Pr[m_2 \leq M_2(t) < m_2 + dm_2]$ is the probability density for the trait variance m_2 . The stationary distribution of the trait variance is

$$\tilde{p}(x) = C x^{-3} e^{(N-1)a(x)}, \quad (14a)$$

where C is a normalizing constant and

$$a(x) = \frac{w_2 x^2 - \mu \sigma^2}{x}. \quad (14b)$$

Figure 6 illustrates an example of stationary distribution of the trait variance, showing that the system tends to show slightly smaller trait variance than the locally stable value in Equation 11, because the fluctuation is weaker when the variance is smaller (geometric Brownian motion).

Comparison with the results of simulations: We compared the stationary distribution of the trait variance between the theoretical prediction given by Equation 14 and the observed distribution obtained from the individual-based simulation (*Evolutionary Branching in a Game in a Finite Population* section). Figure 7 shows good agreement between the diffusion approximation (*i.e.*, the SDE model) and the corresponding simulation result with a small population size. In particular, the approximation captures the shift of the mode—the peak of the distribution lies at the value smaller than $N\mu\sigma^2 = 0.0008$. The average trait variance in an interval shown in Figure 7 is close to $N\mu\sigma^2$ for both analytic prediction (Equation 14) and simulation results.

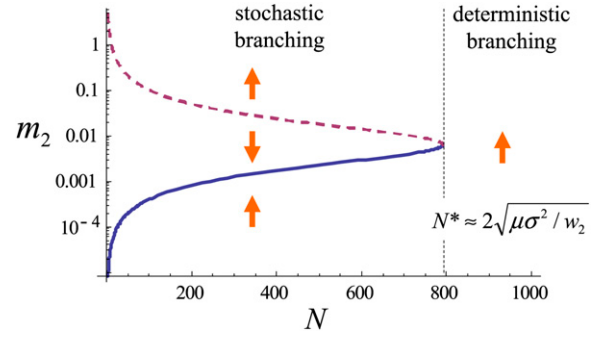


Figure 5 Stable (solid) and unstable (dashed) branches of the trait variance m_2 with population size N as a bifurcation parameter. A Log plot is shown. Arrows indicate the direction of systematic change in m_2 . Parameters: $\mu = 0.01$, $\sigma = 0.02$, $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

Waiting time for evolutionary branching

In case ii, there are stable and unstable equilibria for the trait variance, the latter being larger than the former. If the initial value of trait variance is smaller than the unstable equilibrium, the deterministic model Equation 11 predicts that variance would be maintained indefinitely. In contrast, if the initial value is larger than the unstable equilibrium, it quickly increases and grows without limit. Hence, whether trait variance eventually explodes depends on the initial condition. In the presence of stochasticity, however, the trait variance would not be maintained at smaller values than the unstable equilibrium; instead, it would fluctuate around the stable equilibrium for a period, but then eventually increase beyond the unstable equilibrium and diverge to infinity. Hence, under the stochastic differential equation, the trait variance is predicted to explode eventually even if the initial value is small. The timing of this explosion is controlled by the stochasticity.

The mean waiting time prior to explosion is a typical “first passage time” problem in diffusion theory. This has also been used in calculating the mean time of fixation in population genetics (Kimura and Ohta 1968) and the mean time to extinction in conservation biology (*e.g.*, Lande 1993; Hakoyama and Iwasa 2000) and physiology (*e.g.*, Ricciardi and Smith 1977).

To study the expectation of waiting time, we assume that a system has a reflecting boundary at $m_2 = l \approx 0$ and an absorbing boundary at $m_2 = r$ (note that the left boundary can never be achieved because the deterministic force is increasing the variance while stochastic fluctuation vanishes at $m_2 = 0$; *i.e.*, Equation 12 is singular at $m_2 = 0$). Let $T(\eta)$ be the expected waiting time until the system reaches the right absorbing boundary $m_2 = r$ when the initial value of the trait variance is $m_2(0) = \eta$. To obtain $T(\eta)$, we use the backward Kolmogorov equation

$$-1 = f(\eta) \frac{d}{d\eta} T(\eta) + \frac{\eta^2}{N} \frac{d^2}{d\eta^2} T(\eta), \quad (15)$$

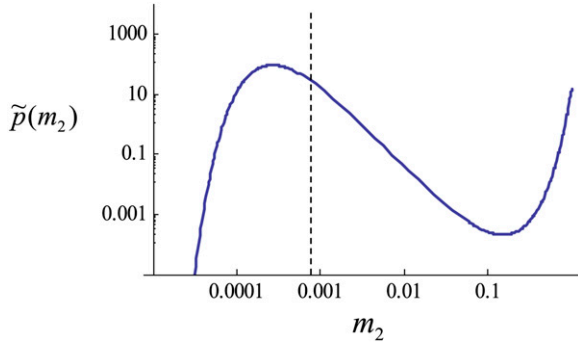


Figure 6 The stationary distribution $\tilde{p}(m_2)$ of the trait variance m_2 (Equation 14). A Log-Log plot is shown, and thus the probability density is much higher at the peak than it appears. The locally stable value of the variance according to Equation 11 is $m_2^* \approx 0.00081$, which is close to $N\mu\sigma^2 = 0.0008$ (dashed line). Note that a local peak lies at a value < 0.0008 because of geometric Brownian motion. Parameters: $N = 200$, $\mu = 0.01$, $\sigma = 0.02$, $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

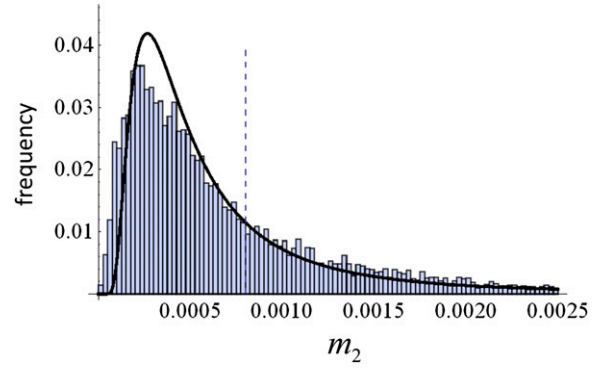


Figure 7 The stationary distribution of the trait variance m_2 . Curves represent an analytic result (Equation 14), while bars are a histogram of realized trait variances in a simulation run for 10^6 time steps. The dashed line represents $N\mu\sigma^2 = 0.0008$. In the simulation, the initial trait distribution was set as monomorphic at $z = 0.2$. The transition from this initial state to the stationary state was negligible as it took only ~ 5000 time steps. Parameters: $N = 200$, $\mu = 0.01$, $\sigma = 0.02$, $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

with the boundary conditions $T(r) = 0$ and $T'(l) = 0$ (see Appendix E). The solution is

$$T(\eta) = N \int_{\eta}^r \int_l^y x^{-3} y e^{(N-1)(a(x)-a(y))} dx dy, \quad (16)$$

where $a(x)$ is given by Equation 14b. The expected waiting time is smaller when the population size is larger, disruptive selection is stronger, or mutational effects are larger (see Appendix D).

Comparison with the results of simulations: We compared the mean waiting times observed in the computer simulations of the individual-based model and the prediction by the approximate stochastic differential equation model given by Equation 16. Once the system reaches a convergence stable point, the trait variance m_2 is expected to be close to the locally stable value $N\mu\sigma^2$ (see also our results on stationary distribution). Thus, we numerically calculate $T(N\mu\sigma^2)$ and compare the corresponding result in individual-based simulation (Figure 8). We also derive the probability distribution of the first passage time by using Kolmogorov's backward equation in Appendix E, but the fit of the shape of the probability distribution of waiting times to the distribution observed in individual-based simulations is less than perfect (Figure 9).

As we decrease N , the waiting time increases exponentially, and it eventually becomes impossible to escape from the locally stable unimodal state in a reasonable time (Figure 8). This implies that evolutionary branching is unlikely to occur in a small population.

When N is very large, genetic drift can be neglected and the system is approximated by the ODE: $dm_2/dt = w_2 m_2^2 + \mu\sigma^2$. Thus, the waiting time can be approximated by the time required to move from $m_2 = N\mu\sigma^2$ to the right boundary $m_2 = r$. This time is ~ 1500 for $N = 1000$ and ~ 960 for $N = 2000$. This calculation, in conjunction with

Figure 8, shows that the stochastic fluctuation plays a dominant role even in a population of 1000 individuals.

Our reduced model measures genetic diversity in trait distribution. An alternative measure of genetic diversity is the number of segregating alleles, which depends only on μ and is predicted by Ewens' sampling formula when selection is absent (Ewens 2004). Figure 10 shows the effect of mutation rate μ and mutation step size σ^2 . Simulations confirmed our prediction that only the value of $\mu\sigma^2$ matters, even though the numbers of segregating alleles in the simulation with $N = 1800$ were ~ 50 and ~ 150 for $\mu = 0.0025$ and 0.01 , respectively. It is also confirmed that smaller mutational effects require a longer waiting time until evolutionary branching.

Discussion

In this article, we have studied the effect of stochastic fluctuation due to finite population size on evolutionary branching. We have derived an ordinary differential equation and then a stochastic differential equation for the dynamics of trait variance. The analysis has revealed that there are two different manners in which evolutionary branching occurs: deterministic branching and stochastic branching. In the case of deterministic branching, evolutionary branching occurs relatively soon after the mean trait reaches the branching point. This occurs when disruptive selection intensity w_2 exceeds a threshold value, $w_2 > 1/(4\mu\sigma^2(N-1)^2)$, which can be approximated by $1/N < 2\sqrt{w_2\mu\sigma^2}$. This condition for deterministic branching implies that the intensity of random genetic drift $1/N$ is smaller than disruptive selection intensity w_2 and the mutational effects $\mu\sigma^2$ (see later in article). If the opposite inequality holds, *i.e.*, when disruptive selection is weaker than the threshold, then a unimodal distribution is locally stable in the deterministic model and

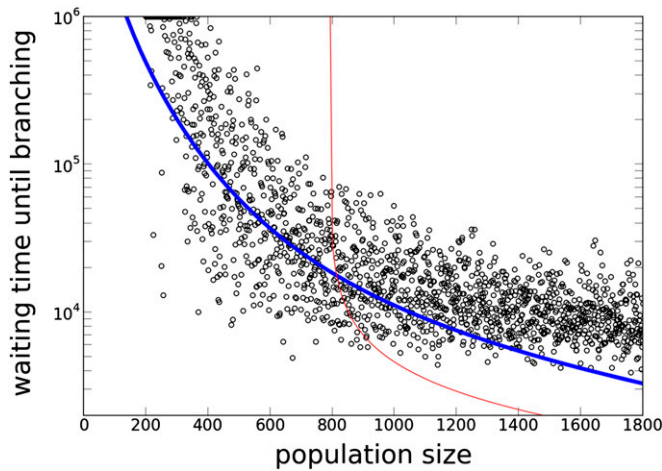


Figure 8 A Log plot of waiting times for evolutionary branching for different population sizes (N). The thick and thin curves represent analytic results based on a SDE model (Equation 16) and those based on an ODE model (Equation 11), respectively. Each open circle represents the result of a single run in individual-based simulation. The first time the trait variance exceeded 0.1 is shown. In the simulation, the initial trait distribution was set as monomorphic at $z = 0.2$ and each run consisted of 10^6 time steps. Simulation runs in which the variance did not explode are plotted at 10^6 . Time required to reach a singular point in the simulation was ~ 5000 when $N = 200$ and ~ 2000 when $N = 1800$, which is included to show open circles. Thus, open circles are expected to lie slightly above the curve. Parameters: $\mu = 0.01$, $\sigma = 0.02$, $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

stochastic fluctuation is necessary for the population to undergo evolutionary branching. We call this latter type of evolutionary branching “stochastic branching.” We observed a relatively long waiting time until evolutionary branching occurs in simulations. The distribution and expectation of the waiting time were mathematically calculated and these predictions were consistent with computer simulations of the individual-based model. Considering the several assumptions we have adopted, the consistency is better than expected, although not perfect. As population size or mutational effect decreases, the waiting time exponentially increases. Mathematically, this is because these factors appear in the exponent in the expression of the expected waiting time (Equation 16). Although the increase of the waiting time is a continuous phenomenon resulting from the underlying stochasticity in individual-based simulations, we can understand evolutionary branching by classifying it as either deterministic branching in which stochastic fluctuation is unnecessary or stochastic branching in which stochastic fluctuation is necessary.

Our analysis has revealed that evolutionary branching is almost impossible in some biological situations in which the effective population size or mutational effect is sufficiently small. As a population size increases, our condition for deterministic evolutionary branching approaches and converges to the standard branching-point condition. The condition can be written as $4w_2N^2\mu\sigma^2 > 1$, which can be used as an approximate indicator; for example, a population with size $N = 10^8$

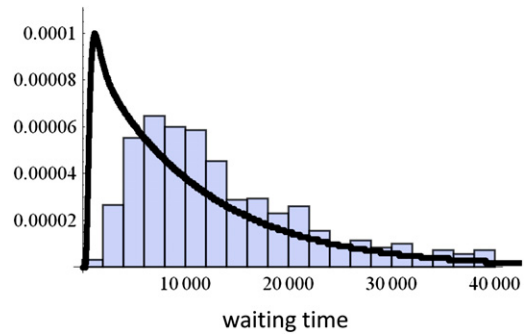


Figure 9 The distribution of waiting times. Curves represent an analytic result (see Appendix E), while bars represent a histogram of the waiting time obtained in 1000 different runs of individual-based simulation. Parameters: $N = 1000$, $\mu = 0.01$, $\sigma = 0.02$, $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

with mutational effects $\mu\sigma^2 = 10^{-14}$ is predicted to show deterministic branching when disruptive selection intensity is $w_2 = O(1)$, although such individual-based simulation is almost impossible.

The condition separating deterministic and stochastic branching is understood intuitively by considering the force suppressing variance (genetic drift) and the two forces promoting the variance (disruptive selection and mutation). The balance between them determines which type of branching occurs. The condition for deterministic branching, $1/N < 2\sqrt{w_2\mu\sigma^2}$ can be rewritten as $(1/N)^2 < 4w_2 \cdot \mu\sigma^2$, which means

$$(\text{genetic drift})^2 < 4 \times (\text{selection effect}) \times (\text{mutation effect}).$$

This implies that deterministic branching occurs when the intensity of genetic drift is smaller than double the geometric average of the selection and mutation intensities. It is important to note that the condition is determined by N^2 instead of N .

To mimic the full individual-based model by using the deterministic and stochastic equations of the trait variance, we have adopted several simplifying assumptions in deriving the dynamics: *i.e.*, trait distribution is assumed to be close to normal, and the trait variance is small enough to justify series expansion. From simulation results, it is clear that the trait distribution finally becomes multimodal when evolutionary branching occurs. Thus, our analytic results are valid only for the initial growth of the trait variance where unimodal distribution (approximated by a normal distribution) becomes broader at the singular point. This initial explosion is followed by highly nonlinear and complex dynamics that realize a multimodal distribution. Nevertheless, the agreement of our analytic prediction and simulation results suggests that our reduced model captures the onset of evolutionary branching.

Based on individual-based simulations, Claessen *et al.* (2007) reported delayed evolutionary branching and proposed two reasons: random genetic drift and extinction of

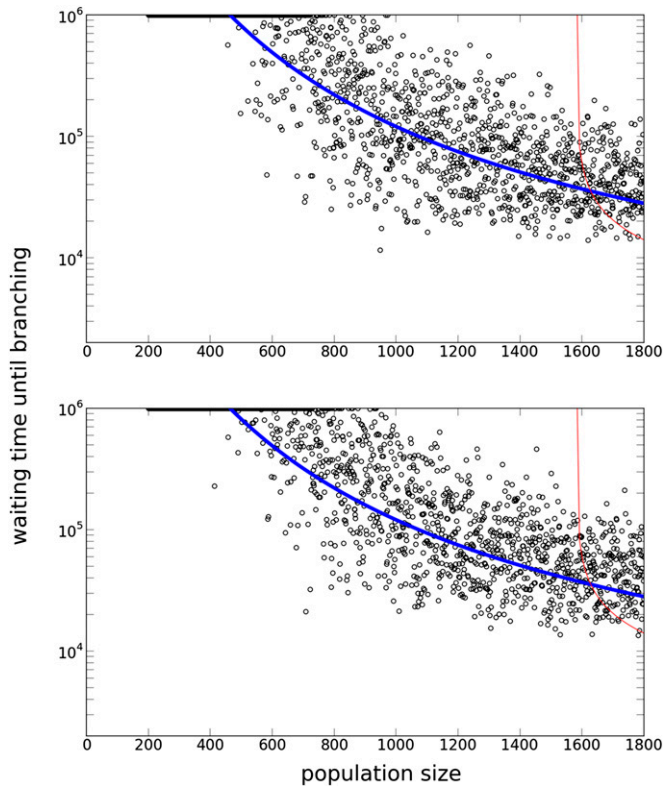


Figure 10 A Log plot of waiting times for evolutionary branching for different population sizes and mutation parameters. The top panel ($\mu = 0.0025, \sigma = 0.02$) depicts a result for a smaller mutation rate than in Figure 8, while the bottom panel ($\mu = 0.01, \sigma = 0.01$) depicts a result for a smaller mutation step size. As $\mu\sigma^2 = 10^{-6}$ is the same, the analytic prediction curves are the same. The two simulation results are also undistinguishable. Parameters: $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$.

incipient branches. In contrast to Claessen *et al.* (2007), our model has no fluctuation in the total population size. Thus, the delayed or absent evolutionary branching is caused by demographic stochasticity—the fluctuation in the number of offspring each individual produces. Our simple model has enabled us to derive the condition for evolutionary branching mathematically, with random genetic drift taken into consideration, which has been only verbally proposed before. In addition, our mathematical analysis and simulation results have revealed that the mutation rate and step size are as important as population size.

In this article, to illustrate the effect of genetic drift on the evolutionary branching, we used as an example a pairwise interaction game that incorporated a Wright–Fisher update. We conjecture that the extension of our model to include a Moran update or to a multiplayer game is readily possible. In the case of including a Moran update, the loss of genetic diversity (Equation 7) should be appropriately modified and also the unit of time should be rescaled since N Moran steps correspond to a single Wright–Fisher step. The result of our model should be easily extended to resource-competition models, which have been intensively studied in

adaptive dynamics (e.g., Dieckmann and Doebeli 1999; Sasaki and Dieckmann 2011; Mirrahimi *et al.* 2012).

We have focused on the dynamic processes in which the population escapes from the locally stable value of the trait variance, while keeping its trait distribution as a unimodal distribution. Conceptually, if we allow all possible trait distributions, we might have two locally stable states, one of which corresponds to a unimodal state and a second that corresponds to a fully branched state. Stochastic branching in the present model considers only the case in which a system located at the unimodal state escapes from this local optimum by stochastic fluctuation (random genetic drift). However, a branched state can also evolve back to the unimodal state by genetic drift; thus, a stochastic analysis including both escaping from and evolving back to the unimodal distribution would give us an even deeper understanding of evolutionary branching.

Acknowledgments

We thank the following people for their very helpful comments: C. Hauert, Y. Kobayashi, C. Miura, W. Nakahashi, H. Ohtsuki, and A. Sasaki. This work was supported by Japan Science and Technology Precursory Research for Embryonic Science and Technology (J.Y.W.), by a Grant-in-Aid for Scientific Research (B) from Japan Society for Promotion of Science (to Y.I.), and also by the support of the Environment Research and Technology Development Fund (S9) of the Ministry of the Environment, Japan.

Literature Cited

- Barton, N. H., and M. Turelli, 1991 Natural and sexual selection on many loci. *Genetics* 127: 229–255.
- Champagnat, N., R. Ferriere, and S. Meleard, 2006 Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theor. Popul. Biol.* 69: 297–321.
- Claessen, D., J. Andersson, L. Persson, and A. M. de Roos, 2007 Delayed evolutionary branching in small populations. *Evol. Ecol. Res.* 9: 51–69.
- Dieckmann, U., and R. Law, 1996 The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* 34: 579–612.
- Dieckmann, U., and M. Doebeli, 1999 On the origin of species by sympatric speciation. *Nature* 400: 354–357.
- Doebeli, M., and U. Dieckmann, 2000 Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am. Nat.* 156: S77–S101.
- Doebeli, M., C. Hauert, and T. Killingback, 2004 The evolutionary origin of cooperators and defectors. *Science* 306: 859–862.
- Ewens, W. J., 2004 *Mathematical Population Genetics*. Springer-Verlag, New York.
- Feller, W., 1971 *Introduction to Probability Theory and Its Application*, Vol. 2, Ed. 3. John Wiley & Sons, New York.
- Geritz, S. A. H., J. A. J. Metz, E. Kisdi, and G. Meszina, 1997 Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.* 78: 2024–2027.
- Geritz, S. A. H., E. Kisdi, G. Meszina, and J. A. J. Metz, 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 12: 35–57.

- Hakoyama, H., and Y. Iwasa, 2000 Extinction risk of a density-dependent population estimated from a time series of population size. *J. Theor. Biol.* 204: 337–359.
- Iwasa, Y., A. Pomiankowski, and S. Nee, 1991 The evolution of costly mate preferences. 2. The handicap principle. *Evolution* 45: 1431–1442.
- Karlin, S., and H. M. Taylor, 1981 *A Second Course in Stochastic Processes*. Academic Press, London.
- Kimura, M., 1957 Some problems of stochastic processes in genetics. *Ann. Math. Stat.* 28: 882–901.
- Kimura, M., and T. Ohta, 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61: 763–771.
- Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314–334.
- Lande, R., 1981 Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. USA* 78: 3721–3725.
- Lande, R., 1993 Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* 142: 911–927.
- Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz, 1992 How should we define “fitness” for general ecological scenarios? *Trends Ecol. Evol.* 7: 198–202.
- Metz, J. A. J., S. A. H. Geritz, G. Meszner, F. J. A. Jacobs, and J. S. van Heerwaarden, 1996 Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction, pp. 183–231 in *Stochastic and Spatial Structures of Dynamical Systems*, edited by S. J. van Strien and S. M. Verduyn Lunel. North-Holland Publishing, Amsterdam.
- Mirrahimi, S., B. Perthame, and J. Y. Wakano, 2012 Evolution of species trait through resource competition. *J. Math. Biol.* 64: 1189–1223.
- Pomiankowski, A., Y. Iwasa, and S. Nee, 1991 The evolution of costly mate preferences. 1. Fisher and biased mutation. *Evolution* 45: 1422–1430.
- Ricciardi, L. M., and C. E. Smith, 1977 Diffusion processes and related topics in biology, in *Lecture Notes in Biomathematics*, Vol. 14. Springer-Verlag, Berlin.
- Sasaki, A., and U. Dieckmann, 2011 Oligomorphic dynamics for analyzing the quantitative genetics of adaptive speciation. *J. Math. Biol.* 63: 601–635.
- Tazzyman, S. J., and Y. Iwasa, 2000 Sexual selection can increase the effect of random genetic drift: a quantitative genetic model of polymorphism in *Oophaga pumilio*, the strawberry poison-dart frog. *Evolution* 64: 1719–1728.
- Wakano, J. Y., and L. Lehmann, 2012 Evolutionary and convergence stability for continuous phenotypes in finite populations derived from two-allele models. *J. Theor. Biol.* 310: 206–215.
- Waxman, D., and S. Gavrillets, 2005 Twenty questions on adaptive dynamics. *J. Evol. Biol.* 18: 1139–1154.

Communicating editor: W. Stephan

Appendices

Appendix A: Fitness Function and the Coefficients

Because individuals are playing a game, their payoff and thus their fitness depend not only on an individual’s own trait but also on their opponents’ traits. In principle, it is impossible to determine a fitness function for an individual with trait z by an average trait \bar{z} because more information is needed [*i.e.*, the distribution of the trait $\varphi(z)$]. However, when the trait variance $\text{Var}[\varphi(z)] = m_2$ is small, fitness is mainly determined by the deviation of the trait value from the average, $z - \bar{z}$ and can be approximated by

$$w(z) = w(\bar{z}) + w_1(z - \bar{z}) + \frac{w_2}{2}(z - \bar{z})^2, \quad (\text{A1})$$

where the derivatives w_1 and w_2 are evaluated at a situation for which all individuals (except self) carry the average trait \bar{z} . In other words, Equation A1 represents invasion fitness. The coefficients are functions of \bar{z} , so that $w_1 = w_1(\bar{z})$ and $w_2 = w_2(\bar{z})$. According to adaptive dynamics, a singular point z^* satisfies $w_1(z^*) = 0$ and it is convergence stable if $(d/d\bar{z})w_1(\bar{z})|_{\bar{z}=z^*} < 0$. It is evolutionarily stable if $w_2(z^*) < 0$. When a singular point is convergence stable but not evolutionarily stable, evolutionary branching is known to occur.

Here we mainly focus on a case in which the average trait remains at $\bar{z} = z^*$, and thus we treat w_1 and w_2 as if they were constant coefficients. They can be explicitly calculated for finite N , but when $N > 100$, it is well approximated by the value obtained by an infinite population limit. In this limit, fitness function is denoted as

$$w(z) = \frac{f(z, z^*)}{f(z^*, z^*)}, \quad (\text{A2})$$

and thus

$$w_1 = \frac{1}{f(z^*, z^*)} \left(\frac{\partial}{\partial z} f(z, z^*) \Big|_{z=z^*} \right) \quad (\text{A3a})$$

and

$$w_2 = \frac{1}{f(z^*, z^*)} \left(\frac{\partial^2}{\partial z^2} f(z, z^*) \Big|_{z=z^*} \right) \quad (\text{A3b})$$

hold. To calculate the intensity of disruptive selection, not only the sign but also the magnitude is important. For the model of Doebeli *et al.* (2004), this becomes

$$w_2 = \frac{2(b_2 - c_2)}{f(z^*, z^*)}, \quad (\text{A4})$$

where $z^* = (c_1 - b_1)/(4b_2 - 2c_2)$. We numerically confirmed that the value of w_2 can be well approximated by $w_2 \approx 0.1$ for the parameter values $(b_1, b_2, c_1, c_2) = (6.0, -1.4, 4.56, -1.6)$ and for the population sizes $N > 100$. The magnitude of disruptive selection depends on b_1 and c_1 , as well as on b_2 and c_2 . In the vicinity of our choice of parameter values used in all figures, disruptive selection is stronger for smaller values of b_1 and larger values of c_1 .

Appendix B: Derivation of Deterministic Approximation

When both selection and mutation are weak, their effects can be approximated as additive and we obtain the replicator Equation 4. By integrating Equation 5, which is the Taylor expansion of $w(z)$ around

$$\bar{z} \equiv \int z \varphi(z) dz, \quad (\text{B1})$$

we have

$$\bar{w} \equiv \int \varphi(z) w(z) dz = w(\bar{z}) + \frac{w_2}{2} m_2, \quad (\text{B2})$$

where

$$m_k \equiv \int (z - \bar{z})^k \varphi(z) dz \quad (\text{B3})$$

is the k th centralized moment around \bar{z} . Putting this into Equation 4, we have

$$\chi(z) = \varphi(z) + \left[-\frac{w_2}{2} m_2 + w_1(z - \bar{z}) + \frac{w_2}{2}(z - \bar{z})^2 \right] \varphi(z, t) + \frac{\mu \sigma^2}{2} \frac{\partial^2}{\partial z^2} \varphi(z). \quad (\text{B4})$$

Imposing zero-flux boundary conditions $\partial \varphi / \partial z = 0$, we have

$$\begin{aligned} \int \chi(z) dz &= 1 \\ \int \chi(z)(z - \bar{z}) dz &= w_1 m_2 + \frac{w_2}{2} m_3 \\ \int \chi(z)(z - \bar{z})^2 dz &= \left(1 - \frac{w_2}{2} m_2\right) m_2 + w_1 m_3 + \frac{w_2}{2} m_4 + \mu \sigma^2. \end{aligned} \quad (\text{B5})$$

We denote statistical values of the offspring distribution by

$$\begin{aligned} \text{Av}_\chi[z] &\equiv \int \chi(z) z dz \\ \text{Var}_\chi[z] &\equiv \int \chi(z) (z - \text{Av}_\chi[z])^2 dz. \end{aligned} \quad (\text{B6})$$

Then we have $\text{Av}_\chi[z] = \bar{z} + w_1 m_2 + (w_2/2) m_3$, which is Equation 6a.

To calculate the variance in the offspring pool, we use

$$\begin{aligned} \text{Var}_\chi[z] &= \int \chi(z) ((z - \bar{z}) - (\text{Av}_\chi[z] - \bar{z}))^2 dz \\ &= \int \chi(z) (z - \bar{z})^2 dz - (w_1 m_2 + \frac{w_2}{2} m_3)^2 \end{aligned} \quad (\text{B7})$$

and we have

$$\text{Var}_\chi[z] = m_2 - \frac{w_2}{2} m_2^2 + w_1 m_3 + \frac{w_2}{2} m_4 + \mu \sigma^2 - (w_1 m_2 + \frac{w_2}{2} m_3)^2. \quad (\text{B8})$$

When we neglect the fifth and higher-order terms, we have $\text{Var}_\chi[z] = m_2 - (w_1^2 + w_2/2) m_2^2 + w_1 m_3 + (w_2/2) m_4 + \mu \sigma^2$, which is Equation 6b.

Appendix C: Derivation of SDE Approximation

We want to approximate the dynamics by an SDE with the form

$$dM_2 = f(M_2)dt + \sqrt{g(M_2)}dB, \quad (C1)$$

where dB denotes Brownian motion and M_2 denotes a random variable representing the variance in trait distribution. To obtain $g(x)$, we use an established result in statistics that a random variable

$$X = \sum_{i=1}^N \left(\frac{Z_i - z}{v} \right)^2 \quad (C2)$$

obeys the chi-square distribution where Z_i 's are N independent random variables each of which is sampled from a normal distribution with mean z and variance v^2 . The variance of the chi-square distribution is $2N$. When N is large, we can approximate z in Equation C2 by \bar{Z} . The trait variance M_2 can be written by

$$M_2 \equiv \frac{1}{N} \sum_{i=1}^N (Z_i - \bar{Z})^2 = \frac{v^2}{N} X, \quad (C3)$$

where M_2 is a random variable, due to random genetic drift. Its variance over probability distribution is given by

$$\text{Var}[M_2] = 2N \left(\frac{v^2}{N} \right)^2 = \frac{2v^4}{N}. \quad (C4)$$

According to our model assumptions, we approximate the variance v^2 by the trait variance in offspring distribution $\chi(z)$. Then, we have the approximation

$$\text{Var}[M_2] = \frac{2m_2^2}{N}, \quad (C5)$$

which yields $g(x) = 2x^2/N$ and Equation 12.

By a more detailed calculation, we can show

$$\text{Var}[M_2] = \frac{2(N-1)m_2^2}{N^2}, \quad (C6)$$

which results in $\tilde{p}(x) = x^{-(3N-2)/(N-1)}e^{Na(x)}$, but for our purpose Equation C5 is precise enough.

Appendix D: Properties of Waiting Times

Here, we show that the expectation of waiting time

$$T = N \int_{\eta}^r \int_l^y x^{-3} y e^{(N-1)(a(x)-a(y))} dx dy \quad (D1)$$

is a decreasing function of w_2 , N , and $\mu\sigma^2$. First, T is an increasing function of $a(x) - a(y)$, which satisfies

$$a(x) - a(y) = -w_2(y-x) - \mu\sigma^2 \left(\frac{1}{x} - \frac{1}{y} \right). \quad (D2)$$

Since $y > x$ holds in the region of integration in Equation D1, the waiting time T is a decreasing function of w_2 and $\mu\sigma^2$. To show the dependence on N , we use

$$\frac{d}{dN} T = \int_{\eta}^r \int_l^y x^{-3} e^{(N-1)(a(x)-a(y))} B(x,y) dx dy, \quad (D3a)$$

where

$$B(x, y) = y + N \frac{(x - y)(wxy + \mu\sigma^2)}{x}. \quad (\text{D3b})$$

When N is large, $B(x, y)$ has the same sign as $x - y$, which is negative. Thus, T is a decreasing function of N for large N . In addition, since $a(x) - a(y)$ is a negative constant independent of N , we have

$$\lim_{N \rightarrow \infty} T = 0, \quad (\text{D4})$$

with all the other parameters fixed at positive values. This is true even for the vanishing mutation rate (*i.e.*, $N\mu\sigma^2 = \text{const.}$ when $N \rightarrow \infty$). However, it does not necessarily go to zero if $N(a(x) - a(y))$ does not go to $-\infty$. For example, if Nw_2 and $N\mu\sigma^2$ remain constant as $N \rightarrow \infty$, the waiting time T can be an increasing function of N .

Appendix E: Distribution of Waiting Times

Let $\psi(x)$ be an indicator function such that $\psi(r) = 1$ and $\psi(x) = 0$ for $l \leq x < r$. Then a random variable $\psi(M_2(t))$ is unity if absorption at $m_2 = r$ has occurred until time t , and it is zero otherwise. Let $u(\eta, t) = E^\eta[\psi(M_2(t))]$, where E^η denotes the expectation in the stochastic process with the initial value $M_2(0) = \eta$. Then the following Kolmogorov backward equation holds (Karlin and Taylor 1981):

$$\frac{\partial u}{\partial t} = f(\eta) \frac{\partial u}{\partial \eta} + \frac{1}{2} g(\eta) \frac{\partial^2 u}{\partial \eta^2}. \quad (\text{E1})$$

Let Q be a random variable representing the first passage time; then it is intuitively clear that

$$u(t, \eta) = \Pr[Q \leq t | m_2(0) = \eta] \quad (\text{E2})$$

is the probability that absorption at $m_2 = r$ has occurred on or before time t . The probability density that absorption occurs exactly at time t is given by $\partial u / \partial t$. A similar logic and calculation were adopted in a classical article studying the fixation probability in population genetics (Kimura 1957). The expected first passage time is

$$T(\eta) = \int_0^\infty t \frac{\partial u}{\partial t} dt. \quad (\text{E3})$$

After some calculations, we obtain

$$-1 = f(\eta) \frac{d}{d\eta} T(\eta) + \frac{\eta^2}{N} \frac{d^2}{d\eta^2} T(\eta),$$

which is Equation 15. To obtain the distribution of waiting time, we numerically solve the partial differential Equation E1 with domain $\eta \in [l, r]$. We impose a reflecting boundary at left and Dirichlet boundary at right, $u(t, r) = 1$. In the numerical algorithm, the initial value is set as $u(0, \eta) = 0$ except $u(0, r) = 1$. The Forward-Time Central-Space method with small dt and $d\eta$ values gives the numerical solution $\tilde{u}(t, \eta)$. To check the validity of this result, we numerically put it into Equation E3 and compared the result with that of the numerically obtained value of Equation 16. The expected waiting times obtained by the two different approaches were almost the same. The distribution of the waiting time t from the initial trait variance $N\mu\sigma^2$ is given by $\tilde{u}(t, N\mu\sigma^2)$, which is shown in Figure 9.