



International Institute for  
Applied Systems Analysis  
Schlossplatz 1  
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342  
Fax: +43 2236 71313  
E-mail: [publications@iiasa.ac.at](mailto:publications@iiasa.ac.at)  
Web: [www.iiasa.ac.at](http://www.iiasa.ac.at)

---

## **Interim Report**

**IR-13-039**

### **Evolutionary branching under slow directional evolution**

Hiroshi C. Ito  
Ulf Dieckmann ([dieckmann@iiasa.ac.at](mailto:dieckmann@iiasa.ac.at))

---

#### **Approved by**

Pavel Kabat  
Director General and Chief Executive Officer

June 2015

---

*Interim Reports* on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

# 1 **Evolutionary branching under slow directional evolution**

2 Hiroshi C. Ito<sup>1,2</sup> and Ulf Dieckmann<sup>1</sup>

3

4 <sup>1</sup>*Evolution and Ecology Program, International Institute for Applied Systems Analysis,*  
5 *Schlossplatz 1, A-2361 Laxenburg, Austria*

6 <sup>2</sup>*Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced-*  
7 *Studies (Sokendai), Hayama 240-0193, Kanagawa, Japan*

8

9 Corresponding author: Hiroshi C. Ito

10 Phone +81 29 856 1720

11 Fax +81 29 856 1720

12 Email addresses: H. C. Ito, [hiroshibeetle@gmail.com](mailto:hiroshibeetle@gmail.com), U. Dieckmann, [dieckmann@iiasa.ac.at](mailto:dieckmann@iiasa.ac.at)

## 13 **Highlights**

14 + We derive conditions for evolutionary branching in directionally evolving populations.

15 + The derived conditions extend those for univariate trait spaces to bivariate trait spaces.

16 + Numerical analyses demonstrate their robustness.

17 + Our conditions are further extended to multivariate trait spaces.

18

## 1 **Abstract**

2 Evolutionary branching is the process by which ecological interactions induce evolutionary  
3 diversification. In asexual populations with sufficiently rare mutations, evolutionary  
4 branching occurs through trait-substitution sequences caused by the sequential invasion of  
5 successful mutants. A necessary and sufficient condition for evolutionary branching of  
6 univariate traits is the existence of a convergence stable trait value at which selection is  
7 locally disruptive. Real populations, however, undergo simultaneous evolution in multiple  
8 traits. Here we extend conditions for evolutionary branching to bivariate trait spaces in which  
9 the response to disruptive selection on one trait can be suppressed by directional selection on  
10 another trait. To obtain analytical results, we study trait-substitution sequences formed by  
11 invasions that possess maximum likelihood. By deriving a sufficient condition for  
12 evolutionary branching of bivariate traits along such maximum-likelihood-invasion paths  
13 (MLIPs), we demonstrate the existence of a threshold ratio specifying how much disruptive  
14 selection in one trait direction is needed to overcome the obstruction of evolutionary  
15 branching caused by directional selection in the other trait direction. Generalizing this finding,  
16 we show that evolutionary branching of bivariate traits can occur along evolutionary-  
17 branching lines on which residual directional selection is sufficiently weak. We then present  
18 numerical analyses showing that our generalized condition for evolutionary branching is a  
19 good indicator of branching likelihood even when trait-substitution sequences do not follow  
20 MLIPs and when mutations are not rare. Finally, we extend the derived conditions for  
21 evolutionary branching to multivariate trait spaces.

## 22 **Keywords**

23 frequency-dependent selection, speciation, adaptive dynamics, two-dimensional traits, multi-  
24 dimensional traits

25

# 1 Introduction

2 Real populations have undergone evolution in many quantitative traits. Even when such  
3 populations experience contemporary selection pressures, their selection response will usually  
4 be highly multivariate. However, not all responding adaptive traits evolve at the same speed:  
5 in nature, such evolutionary speeds exhibit a large variation (Hendry and Kinnison, 1999;  
6 Kinnison and Hendry, 2001)<sup>†</sup>. Past speciation processes may have been driven mainly by  
7 traits undergoing fast evolution (Schluter, 1996), while gradual evolutionary differentiation  
8 among species, genera, and families may derive from traits undergoing slow evolution. These  
9 differences in evolutionary speed can have two fundamentally different causes. First, they  
10 may be due to less genetic variation being available for evolution to act on: in asexual  
11 populations this occurs when mutation rates and/or magnitudes are smaller in some traits than  
12 in others, while in sexual populations this occurs when standing genetic variation is smaller in  
13 some traits than in others. Second, differences in evolutionary speed are also expected when  
14 fitness is much less sensitive to changes in some traits than to changes in others.

15 For ~~convenience~~ brevity, we refer to the slowly evolving and the ~~fast~~ rapidly evolving traits  
16 as slow ~~traits~~ and fast traits, respectively. If the slow traits are sufficiently slow, it is tempting  
17 to neglect their effects on the evolution of the fast traits. As far as evolutionary responses to  
18 directional selection are concerned, this simplification is usually unproblematic: ~~directional~~  
19 ~~evolution, i.e., the~~ directional trend of evolution (Rice, 2011) ~~resulting from such selection,~~  
20 ~~which we can refer to as directional evolution,~~ is described effectively by ordinary differen-

---

<sup>†</sup>~~To facilitate the reviewing process, we adopt the author-date style for citations. Naturally, we will immediately change these citations to numbers when it is suggested or once our manuscript is accepted.~~

1 tial equations or difference equations focusing only on those fast traits (Price 1970; Lande,  
2 1979; Dieckmann and Law 1996; Rice, 2011<sup>2</sup>).

3 On the other hand, such a simplification may not be safe where more complex evolutionary  
4 dynamics are involved. A typical example is adaptive speciation, i.e., evolutionary diversifi-  
5 cation driven by ecological interactions (Dieckmann et al., 2004; Rundle and Nosil, 2004). Ito  
6 and Dieckmann (2007) have shown numerically that when populations undergo disruptive  
7 selection in ~~the~~ fast trait, their evolutionary diversification can be suppressed by directional  
8 evolution of ~~a slow~~ another trait, even if the latter is slow. Conversely, if the slow directional  
9 evolution is sufficiently slow, disruptive selection in the fast trait can drive evolutionary di-  
10 versification, both in asexual populations and in sexual populations (Ito and Dieckmann,  
11 2007). The suppression of evolutionary diversification can occur even when the slow and fast  
12 traits are mutationally and ecologically independent of each other. Thus, in a multivariate trait  
13 space, evolutionary diversification in one trait can be suppressed by slow directional evolu-  
14 tion in just one of the many other traits. Moreover, such slow directional evolution may never  
15 cease, as the environments of populations are always changing, at least slowly, due to changes  
16 in abiotic components (e.g., climatic change) or biotic components (i.e., evolution in other  
17 species of the considered biological community). It is therefore important to improve the the-  
18 oretical understanding of this phenomenon by deriving conditions for evolutionary diversifi-  
19 cation under slow directional evolution.

20 As a starting point to this end, we can consider the special situation in which there is only a  
21 single fast trait, while all other traits of the considered population are evolving extremely  
22 slowly, such that they are completely negligible. In this case, the question whether the  
23 selection on the fast trait favors its evolutionary diversification can be examined through

---

1 conditions that have been derived for the evolutionary branching of univariate traits (Metz et  
2 al., 1992; Geritz et al., 1997, 1998). In general, evolutionary branching is the process through  
3 which a unimodal phenotype distribution of a population becomes bimodal in response to  
4 frequency-dependent disruptive selection (Metz et al., 1992; Geritz et al., 1997, 1998;  
5 Dieckmann et al., 2004), which can occur through all fundamental types of ecological  
6 interaction, including competition, ~~predator-prey interaction~~ [exploitation](#), mutualism, and  
7 cooperation (Doebeli and Dieckmann, 2000; Doebeli et al., 2004). This kind of diversifying  
8 evolution provides ecological underpinning for the sympatric or parapatric speciation of  
9 sexual populations (e.g., Doebeli, 1996; Dieckmann and Doebeli, 1999; Kisdi and Geritz,  
10 1999; Doebeli and Dieckmann, 2003; Dieckmann et al., 2004; Claessen et al., 2008; Durinx  
11 and Van Dooren, 2009; Heinz et al., 2009; Payne et al., 2011). Moreover, evolutionary  
12 branching may lead to selection pressures that favor further evolutionary branching, inducing  
13 recurrent adaptive radiations and extinctions (e.g., Ito and Dieckmann, 2007), and thus  
14 community assembly (e.g., Jansen and Mulder, 1999; Bonsall et al., 2004; Johansson and  
15 Dieckmann, 2009; [Brännström et al., 2012](#)) and food-web formation (e.g., Loeuille and  
16 Loreau, 2005; Ito et al., 2009; Brännström et al., 2010; [Takahashi et al., in press](#)). Therefore,  
17 evolutionary branching may be one of the important mechanisms underlying the evolutionary  
18 diversification of biological communities.

19 Conditions for the evolutionary branching of univariate traits can be extended to bivariate  
20 trait spaces, if all traits considered evolve at comparable speeds (Bolnick and Doebeli, 2003;  
21 Vukics et al., 2003; Ackermann and Doebeli, 2004; Van Dooren et al., 2004; Egas et al.,  
22 2005; Leimar, 2005; Van Dooren, 2006; Ito and Shimada, 2007; Ravigné et al., 2009). How-  
23 ever, if their evolutionary speeds are significantly different, the resultant conditions for biva-  
24 riate traits can fail to predict evolutionary branching observed in numerical analyses (Ito and  
25 Dieckmann, 2007; Ito et al., 2009, Ito and Dieckmann, 2012). In the present study, we there-  
26 fore derive conditions for a population's evolutionary branching in a fast trait when, at the  
27 same time, such a population is directionally evolving in one or more slow traits. The result-  
28 ant conditions reveal when slow directional evolution either prevents or permits evolutionary  
29 branching.

1 This paper is structured as follows. Section 2 explains heuristically how the likelihood of  
2 evolutionary branching in asexual populations depends on selection pressures and mutational  
3 step sizes. Section 3 derives a normal form for strong disruptive selection and weak  
4 directional selection in a bivariate trait space and explains when arbitrary bivariate fitness  
5 functions can be mapped onto this normal form. Section 4 introduces the concept of  
6 maximum-likelihood-invasion paths, formed by mutants with maximum likelihood of  
7 invasion. On that basis, Section 5 derives sufficient conditions for evolutionary branching.  
8 Section 6 numerically examines the robustness of these conditions when the simplifying  
9 assumptions underlying our derivation are relaxed. Section 7 summarizes all conditions  
10 needed for identifying evolutionary-branching lines and extends these conditions to  
11 multivariate trait spaces. Section 8 discusses how our results generalize previously derived  
12 conditions for evolutionary branching that ignored slow directional evolution, and how our  
13 maximum-likelihood-invasion paths are related to existing methods for determining  
14 evolutionary dynamics or reconstructing evolutionary histories.

## 15 **2 Heuristics**

16 We start by describing, in a heuristic way, how disruptive selection in one direction,  
17 directional selection in ~~the other~~ another direction, and mutational step sizes may affect the  
18 likelihood of evolutionary branching. We then explain the analyses required for deriving the  
19 conditions for evolutionary branching, which are conducted in the subsequent sections.

20 When a population undergoes disruptive selection in trait  $x$ , as well as directional selec-  
21 tion in trait  $y$ , its fitness landscape resembles that illustrated in Fig. 1a. The strength of dis-  
22 ruptive selection in  $x$  is given by the fitness landscape's curvature (i.e., second derivative)  
23 along  $x$ , denoted by  $D_{xx}$ , while the strength of directional selection in  $y$  is given by the  
24 fitness landscape's ~~steepness~~ slope (i.e., first derivative) along  $y$ , denoted by  $G_y$ . For sim-  
25 plicity, we assume that the population is monomorphic with a resident phenotype  $(x, y)$ , in-  
26 dicated by a small black circle in Fig. 1a, and that mutational step sizes are identical in all  
27 directions. In this case, possible mutants are located on a circle around the resident phenotype,

1 as shown in Fig. 1b-g. Then, small  $G_y$  means slow evolution in  $y$ . Roughly speaking, the  
2 direction of evolution favored by selection is indicated by the mutants possessing maximum  
3 fitness (small white circles in Fig. 1b-g). These mutants are located where the circle of con-  
4 sidered mutants is tangential to the fitness contours.

5 From this simple setting, we can already draw the following **geometrically evident**  
6 conclusions. If  $G_y$  is large compared to  $D_{xx}$ , which results in low curvatures for the fitness  
7 contours (Fig. 1b), the mutant having the maximum  $y$  has maximum fitness, in which case  
8 directional evolution along  $y$  is expected. On the other hand, if  $G_y$  is sufficiently small  
9 compared to  $D_{xx}$  (Fig. 1c, d), the high curvatures mean that two different mutants are  
10 sharing the same maximum fitness. In this case, evolutionary diversification in  $x$  may be  
11 expected. In addition, we can easily see (Fig. 1e-g) that the smaller the mutational step size,  
12 the smaller the  $G_y$  and/or larger the  $D_{xx}$  required for two different mutants jointly having  
13 maximum fitness (Fig. 1e-g).

14 It turns out that these qualitative and heuristic insights can be corroborated by formal  
15 analysis (Sections 3-6). For this, two things have to be done properly. First, we have to clarify  
16 the conditions under which a population undergoes disruptive selection in one direction and  
17 sufficiently weak directional selection in the other direction. To compare the strengths of  
18 selection among different directions, trait spaces have to be normalized so that mutation  
19 becomes isotropic in all directions, as in Fig. 1b-g. Second, because the existence of  
20 disruptive selection is a necessary but not a sufficient condition for evolutionary branching  
21 (Metz et al., 1996; Geritz et al., 1997, 1998), the emergence of an initial dimorphism and the  
22 subsequent process of divergent evolution have to be analyzed. Conducting these analyses in  
23 the subsequent sections, we end up being able quantitatively to predict the likelihood of  
24 evolutionary branching in terms of  $G_y$ ,  $D_{xx}$ , and mutational step sizes.



## 1 **3 Normal form for bivariate invasion-fitness functions** 2 **causing slow directional evolution**

3 In this section, we first derive a normal form that applies when evolution is slow in one  
4 direction. As mentioned before, this may occur when mutational steps or fitness sensitivities  
5 are strongly asymmetrical. Second, we explain the evolutionary dynamics that are expected  
6 under this normal form. Third, we outline the fundamental ideas underlying our subsequent  
7 analyses.

### 8 **3.1 Invasion Bivariate invasion-fitness functions causing slow directional** 9 **evolution**

10 We start by considering arbitrary bivariate trait spaces; accordingly, each phenotype  
11  $\mathbf{S} = (X, Y)^T$  comprises two scalar traits  $X$  and  $Y$ . We assume an asexual population with a  
12 large population size and sufficiently small mutation rates. The latter assumption has two con-  
13 sequences. First, the population dynamics have sufficient time to relax toward their equilibri-  
14 um after a new mutant emerges. Second, as long as the population experiences directional  
15 selection, only the phenotype with the highest fitness among the existing phenotypes survives  
16 as a result of selection. Thus, the population is essentially always close to equilibrium and  
17 monomorphic. This allows its directional evolution to be translated into a trait-substitution  
18 sequence based on the invasion fitness of a mutant phenotype  $\mathbf{S}'$  arising from a resident phe-  
19 notype  $\mathbf{S}$  (Metz et al., 1992, 1996; Dieckmann and Law, 1996).

20 The invasion fitness of  $\mathbf{S}'$  under  $\mathbf{S}$ , denoted by  $F(\mathbf{S}'; \mathbf{S})$ , is defined as the initial per capi-  
21 ta growth rate of  $\mathbf{S}'$  in the monomorphic population of  $\mathbf{S}$  at its equilibrium population size.  
22 The function  $F(\mathbf{S}'; \mathbf{S})$  can be treated as a fitness landscape in  $\mathbf{S}'$ , whose shape depends on  
23  $\mathbf{S}$ . When a mutant emerges, which occurs with probability  $\mu$  per birth, we assume that its  
24 phenotype follows a mutation probability distribution denoted by  $M(\delta\mathbf{S})$ , where  $\delta\mathbf{S} = \mathbf{S}' - \mathbf{S}$ .  
25 The distribution is assumed to be symmetric, unimodal, and smooth. As long as the mutation-  
26 al step sizes are sufficiently small, such that  $M(\delta\mathbf{S})$  has sufficiently narrow width, the distri-  
27 bution is well characterized by its variance-covariance matrix  $\Lambda$ ,

$$\Lambda = \iint \delta \mathbf{S} \delta \mathbf{S}^T M(\delta \mathbf{S}) d\delta X d\delta Y = \begin{pmatrix} V_{XX} & V_{XY} \\ V_{XY} & V_{YY} \end{pmatrix}, \quad (1)$$

2 where  $V_{XX} = \iint \delta X^2 M(\delta \mathbf{S}) d\delta X d\delta Y$ ,  $V_{YY} = \iint \delta Y^2 M(\delta \mathbf{S}) d\delta X d\delta Y$ , and  
 3  $V_{XY} = \iint \delta X \delta Y M(\delta \mathbf{S}) d\delta X d\delta Y$ . The standard deviation of mutational step sizes along each  
 4 direction is thus described by an ellipse,  $\delta \mathbf{S}^T \Lambda^{-1} \delta \mathbf{S} = 1$ , as shown in Fig. 2a. Since  $\Lambda$  is  
 5 symmetric, its two eigenvectors are orthogonal, and these vectors determine the directions of  
 6 the long and short principal axes of this ellipse. Through a coordinate rotation, we align the  
 7 axes of the coordinate system with the eigenvectors of  $\Lambda$ . In the rotated coordinate system,  
 8  $\Lambda$  is diagonal,  $V_{XY} = 0$ , and we can choose the axes such that  $\sigma_X > \sigma_Y$ , where  $\sigma_X = \sqrt{V_{XX}}$   
 9 and  $\sigma_Y = \sqrt{V_{YY}}$  (Fig. 2b). Then, stretching the trait space in the  $Y$ -direction by  $\sigma_X / \sigma_Y$   
 10 gives an isotropic mutation distribution with standard deviation  $\sigma = \sigma_X$  in all directions (Fig.  
 11 2c). This is achieved by introducing a new coordinate system  $\mathbf{s} = (x, y)^T$ , where  $x = X$  and  
 12  $y = (\sigma_X / \sigma_Y) Y$ . We assume that  $\sigma$  is small, such that  $\sigma \ll 1$ . In this normalized trait space,  
 13 the invasion-fitness function can be expanded in  $\mathbf{s}'$  and  $\mathbf{s}$  around a base point  $\mathbf{s}_0 = (x_0, y_0)^T$   
 14 as

$$f(\mathbf{s}'; \mathbf{s}) = \mathbf{G} \delta \mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C} \delta \mathbf{s} + \frac{1}{2} \delta \mathbf{s}^T \mathbf{D} \delta \mathbf{s} + O(\sigma^3), \quad (2a)$$

16 where  $\delta \mathbf{s} = \mathbf{s}' - \mathbf{s} = O(\sigma)$  and  $\mathbf{s} - \mathbf{s}_0 = O(\sigma)$  are assumed. The  $\mathbf{G}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  are given  
 17 by

$$\mathbf{G} = \begin{pmatrix} G_x & G_y \end{pmatrix} = \mathbf{f}_m, \quad \mathbf{C} = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} = \mathbf{f}_{mm} + \mathbf{f}_{rm}, \quad \mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} = \mathbf{f}_{mm}, \quad (2b)$$

$$\mathbf{f}_m = \begin{pmatrix} f_{x'} & f_{y'} \end{pmatrix}_{\mathbf{s}' = \mathbf{s} = \mathbf{s}_0}, \quad \mathbf{f}_{mm} = \begin{pmatrix} f_{x'x'} & f_{x'y'} \\ f_{x'y'} & f_{y'y'} \end{pmatrix}_{\mathbf{s}' = \mathbf{s} = \mathbf{s}_0}, \quad \mathbf{f}_{rr} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}_{\mathbf{s}' = \mathbf{s} = \mathbf{s}_0}, \quad \mathbf{f}_{rm} = \begin{pmatrix} f_{xx'} & f_{xy'} \\ f_{yx'} & f_{yy'} \end{pmatrix}_{\mathbf{s}' = \mathbf{s} = \mathbf{s}_0}$$

19 where the subscripts 'm' and 'r' refer to mutants and residents, respectively, and where  $f_\alpha$   
 20 for  $\alpha = x', y', x, y$  and  $f_{\alpha\beta}$  for  $\alpha, \beta = x', y', x, y$  denote the first and second derivatives of  
 21  $f(\mathbf{s}'; \mathbf{s})$ , respectively. See Appendix A for the derivation of Eqs. (2). Notice that  $\mathbf{G}$ ,  $\mathbf{C}$ , and  
 22  $\mathbf{D}$  are functions of the base point  $\mathbf{s}_0$ . The vector  $\mathbf{G}$  describes the fitness gradient at  $\mathbf{s}$   
 23 when  $\mathbf{s} = \mathbf{s}_0$ . The matrix  $\mathbf{C}$  describes how the fitness gradient at  $\mathbf{s}$  changes as  $\mathbf{s}$  deviates

1 from  $\mathbf{s}_0$ . The matrix  $\mathbf{D}$  describes the curvature (i.e., second derivative) of the fitness  
 2 landscape, which is approximately constant as long as the third-order terms can be neglected.

3 If  $\sigma_Y$  is much smaller than  $\sigma_X$ , such that  $\sigma_Y = O(\sigma_X^2)$ , the stretching of the trait space in  
 4 the  $Y$ -direction for the normalization makes derivatives with respect to  $y$  very small, in the  
 5 sense  $G_y = O(\sigma)$ ,  $C_{xy} = O(\sigma)$ ,  $C_{yx} = O(\sigma)$ ,  $D_{xy} = O(\sigma)$ ,  $C_{yy} = O(\sigma^2)$ , and  $D_{yy} = O(\sigma^2)$ .  
 6 This simplifies Eq. (2a),

$$7 \quad f(\mathbf{s}'; \mathbf{s}) = G_x \delta x + C_{xx} (x - x_0) \delta x + \frac{1}{2} D_{xx} \delta x^2 + G_y \delta y + O(\sigma^3), \quad (3)$$

8 where terms with  $C_{xy}$ ,  $C_{yx}$ ,  $C_{yy}$ ,  $D_{xy}$ , and  $D_{yy}$  are subsumed in the higher-order terms  
 9  $O(\sigma^3)$ . For a derivation of Eq. (3), see Appendix B. Notice that on the right-hand side only  
 10 the first term is of order  $O(\sigma)$ , while the other terms, including  $G_y \delta y$ , are of order  $O(\sigma^2)$ .  
 11 This means that this normal form describes fitness functions with significantly weak  
 12 directional selection along  $y$  compared to that along  $x$  as long as  $G_x$  and  $G_y$  in the  
 13 original trait space have similar magnitudes. We call Eq. (3) the normal form for invasion-  
 14 fitness functions with significant sensitivity difference.

15 A comparison of Eqs. (2a) and (3) shows that the latter can be obtained even for  $\sigma_Y = \sigma_X$ ,  
 16 provided the sensitivity of the fitness function to variation in trait  $Y$  is significantly weaker  
 17 than that in  $X$ , so that  $G_y$ ,  $C_{xy}$ ,  $C_{yx}$ ,  $C_{yy}$ ,  $D_{xy}$ , and  $D_{yy}$  are all relatively small,  
 18 satisfying

$$19 \quad \frac{|G_y| + |C_{xy}| + |C_{yx}| + |C_{yy}| + |D_{xy}| + |D_{yy}|}{|G_x| + |C_{xx}| + |D_{xx}|} = O(\sigma). \quad (4)$$

20 For a derivation of Eq. (4), see Appendix B.

21 Notice that the assumption needed for the derivation of Eq. (3), i.e.,  $\sigma_Y = O(\sigma_X^2)$ , also  
 22 satisfies Eq. (4). Thus, the condition for a significant difference between mutational step sizes  
 23 in the original trait space  $\mathbf{S}$  can naturally be integrated with the condition for a significant  
 24 sensitivity difference of the invasion-fitness function in the normalized trait space  $\mathbf{s}$ . Based  
 25 on Eq. (4), we therefore define the condition for significant sensitivity difference as follows.

1           **Significant sensitivity difference:** After normalization to make mutation isotropic, the  
 2           invasion-fitness function can be made to satisfy Eq. (4) by rotating the  $x$ - and  $y$ -  
 3           axes.

### 4   **3.2 Evolutionary dynamics expected under normal form**

5   We now consider the expected evolutionary dynamics induced by the normal form in Eq. (3).  
 6   For this purpose we first recap expectations for the simpler case in which  $G_y$  is so small that  
 7    $G_y = O(\sigma^2)$ . In that case,  $G_y \delta y = O(\sigma^3)$  is negligible, so that  $y$  vanishes from Eq. (3),  
 8   and the evolutionary dynamics therefore become univariate in  $x$ , so that phenotypes are  
 9   characterized by that trait value alone. In this simpler case, conditions for evolutionary  
 10   branching are easier to understand (Metz et al., 1996; Geritz et al., 1997, 1998), as follows.

11   Suppose that the base point  $x_0$  of the expansion of  $f(\mathbf{s}'; \mathbf{s})$  can be chosen such that  
 12    $G_x = 0$ . Such a point, which we denote by  $x_b$ , is called an evolutionarily singular point (or  
 13   simply a singular point or an evolutionary singularity) because a resident located at  $x_b$   
 14   experiences no directional selection. In contrast, a resident located close to  $x_b$  experiences  
 15   directional selection along  $x$ ,

$$16 \quad \left. \frac{\partial f(\mathbf{s}'; \mathbf{s})}{\partial x'} \right|_{x'=x} = C_{xx} (x - x_b). \quad (5)$$

17   If  $C_{xx}$  is negative, the fitness gradient is positive for  $x < x_b$  and negative for  $x > x_b$ , which  
 18   means that it attracts a monomorphic population through directional evolution. In other  
 19   words, the singular point is then convergence stable (Christiansen, 1991).

20   When a population comes close to  $x_b$ , it may become possible for a mutant  $\mathbf{s}'$  to coexist  
 21   with a resident  $\mathbf{s}$ . Mutual invasibility between  $\mathbf{s}'$  and  $\mathbf{s}$ , which gives rise to protected di-  
 22   morphism (Prout, 1968), is defined by  $f(\mathbf{s}'; \mathbf{s}) > 0$  and  $f(\mathbf{s}; \mathbf{s}') > 0$ , which requires  
 23    $D_{xx} - C_{xx} > 0$ . Following the emergence of a protected dimorphism of trait values denoted by  
 24    $\mathbf{s}_1$  and  $\mathbf{s}_2$ , the resultant fitness landscape  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  maintains approximately the same cur-  
 25   vature (i.e., second derivative)  $D_{xx}$  along  $x$ . If this curvature is positive, i.e., this point is  
 26   not a local **evolutionarily stable strategy, or ESS** (Maynard Smith and Price, 1973), the two

1 subpopulations evolve in opposite directions, keeping their coexistence, in a process called  
 2 dimorphic divergence. When dimorphic divergence occurs in univariate trait spaces, it can  
 3 never collapse (if it is assumed that mutual invasibility among phenotypes ensures their coex-  
 4 istence). This is because only mutants outside of the interval between the two residents can  
 5 invade, and such an invading mutant then always excludes the closer resident and is mutually  
 6 invisable with the other more distant resident, resulting in a new protected dimorphism with a  
 7 larger phenotypic distance (Geritz et al., 1998). As we will see below, such collapses, howev-  
 8 er, become crucial when analyzing dimorphic divergence in bivariate trait spaces.

9 The evolutionary process described above is called evolutionary branching. It requires  
 10 monomorphic convergence ( $C_{xx} < 0$ ), mutual invasibility ( $D_{xx} - C_{xx} > 0$ ), and dimorphic  
 11 divergence ( $D_{xx} > 0$ ). We therefore see that the necessary and sufficient conditions for  
 12 univariate evolutionary branching are given by the existence of a point  $x = x_b$  satisfying

13 
$$G_x = 0, \tag{6a}$$

14 
$$C_{xx} < 0, \tag{6b}$$

15 
$$D_{xx} > 0, \tag{6c}$$

16 with  $x_0 = x_b$ . When trait  $y$  is also taken into account, the point  $x = x_b$  forms a line in the  
 17 bivariate trait space. Thus, the aforementioned conditions for univariate evolutionary  
 18 branching can be translated into the following statement.

19 **Bivariate translation of conditions for univariate evolutionary branching:** When  
 20 directional selection in  $y$  is very weak, such that  $G_y = O(\sigma^2)$ , monomorphic  
 21 populations around a line  $x = x_b$  converge to that line and bring about evolutionary  
 22 branching, if and only if Eqs. (6) are all satisfied.

23 Using these simple results as a baseline for comparison, we now consider the case that  $G_y$   
 24 is of order  $\sigma^1$ , which, according to Eq. (3), implies that the evolution in  $y$  can affect the  
 25 evolution in  $x$ . When the population is not close to the singular line  $x = x_b$ , directional  
 26 evolution in  $x$  dominates the effects of  $y$ . Thus, the singular line still attracts monomorphic

1 populations if and only if  $C_{xx} < 0$ . We thus call such a singular line a convergence-stable line.  
2 When a population is in the neighborhood of a convergence-stable line and evolves toward it,  
3 directional selection in  $x$  inevitably becomes small, such that  $G_y$  may affect the  
4 evolutionary dynamics.

5 When  $x$  is close to  $x_b$ , a dimorphism in  $x$  may emerge, but invasion by a mutant in  $y$   
6 with higher fitness may exclude both of the coexisting resident phenotypes and thereby abort  
7 the incipient evolutionary branching. Such an abortion is especially likely when  $G_y$  is large.  
8 Thus, the larger  $G_y$  becomes, the more difficult evolutionary branching is expected to be.  
9 Below, we examine how the resultant likelihood of evolutionary branching can be estimated.

### 10 **3.3 Motivation for further analyses**

11 In principle, bivariate evolutionary branching is possible even for very large  $G_y$ , as long as  
12 trait-substitution sequences comprise invasions only in  $x$  for an adequately large number of  
13 substitutions after the inception of dimorphism. However, for large  $G_y$ , sequential invasions  
14 of this kind are unlikely, because the fitness advantage of mutants in  $y$  then is large, which  
15 favors their invasion, which in turn easily destroys any initial dimorphisms. Thus, the average  
16 number of invasions required for evolutionary branching is expected to be quite large in this  
17 case. We can thus measure the likelihood of evolutionary branching as the probability of its  
18 successful completion within a given number of invasions.

19 It is difficult to calculate this probability directly, and thus to determine its dependence on  
20 the parameters  $G_y$ ,  $C_{xx}$ , and  $D_{xx}$  of the normal form. To avoid this difficulty, we focus on  
21 invasions that individually have maximum likelihood for a given composition of residents.  
22 We can loosely interpret the successions of residents formed by such invasions as describing  
23 typical evolutionary paths. Because of their special construction, it is possible analytically to  
24 derive sufficient conditions for evolutionary branching along these paths. It is expected that  
25 the conditions thus obtained can serve as useful indicators for the probability of evolutionary  
26 branching along the more general evolutionary paths formed by arbitrary stochastic invasions.

27 Notice that when we refer to stochastic invasions, we refer to the stochasticity of mutations  
28 and to the stochasticity of the initial survival of rare mutants, but not to the effects resulting

1 from small resident population size, which occasionally allow mutants to invade even when  
2 they have negative fitness. In formal terms, these clarifications are implied by our assumption  
3 of sufficiently large resident population size.

4 In our analyses below, we assume that the conditions for evolutionary branching in  
5 univariate trait spaces, Eqs. (6), are satisfied. Our goal is to determine how conditions for  
6 bivariate evolutionary branching in  $x$  under weak directional selection in  $y$  differ from  
7 Eqs. (6).

## 8 **4 Maximum-likelihood-invasion paths**

9 In this section, we define evolutionary paths formed by sequential invasions each of which  
10 has maximum likelihood. Among all possible evolutionary paths formed by arbitrary  
11 stochastic invasions, these paths have high likelihood and may therefore be regarded as  
12 typical. Our reason for introducing these maximum-likelihood-invasion paths (MLIPs) is that  
13 we can derive, in Section 5, conditions for evolutionary branching along those typical paths ~~in~~  
14 [Section 5](#).

### 15 **4.1 Definition of oligomorphic stochastic invasion paths**

16 We start by explaining how probabilities of invasion events are formally defined. We consider  
17 a monomorphic population with phenotype  $\mathbf{s}$ , as a trait vector with an arbitrary dimension, at  
18 equilibrium population size  $\hat{n}$  that is uniquely determined by  $\mathbf{s}$ . The birth and death rates  
19 (i.e., the number of birth and death events per individual per unit time, respectively) of a rare  
20 mutant phenotype  $\mathbf{s}'$  are denoted by  $b(\mathbf{s}';\mathbf{s})$  and  $d(\mathbf{s}';\mathbf{s})$ , where  $b(\mathbf{s};\mathbf{s})$  and  $d(\mathbf{s};\mathbf{s})$   
21 denote the birth and death rates of the resident  $\mathbf{s}$ , which must satisfy  
22  $f(\mathbf{s};\mathbf{s}) = b(\mathbf{s};\mathbf{s}) - d(\mathbf{s};\mathbf{s}) = 0$ , because the resident is at population dynamical equilibrium. The  
23 invasion fitness of the mutant  $\mathbf{s}'$  in the environment determined by the resident  $\mathbf{s}$  is given  
24 by

$$25 \quad f(\mathbf{s}';\mathbf{s}) = b(\mathbf{s}';\mathbf{s}) - d(\mathbf{s}';\mathbf{s}). \quad (7)$$

1 Once a mutant  $s'$  has arisen, the probability of its successful invasion in a population of  
 2 resident  $s$  is approximately given by  $f(s';s)_+/b(s';s)$  (Dieckmann and Law, 1996). Here,  
 3 the subscript “+” denotes conversion of negative values to zero. The probability (density per  
 4 unit time) for the emergence of a successfully invading mutant  $s'$  in a population of  
 5 residents  $s$  is given by multiplying the number density  $\mu\hat{n}b(s;s)M(s'-s)$  of mutants  
 6 emerging per unit time with their probability of successful invasion,

$$7 \quad E(s';s) = \mu\hat{n}b(s;s)M(s'-s)\frac{f(s';s)_+}{b(s';s)} \quad (8a)$$

$$\approx \mu\hat{n}M(s'-s)f(s';s)_+,$$

8 where  $\mu$  is the mutation probability per birth event, and  $M(s'-s)$  is the mutation  
 9 probability distribution. The above approximation applies in the leading order of  $s'-s$ , when  
 10  $\sigma$  is sufficiently small such that  $b(s';s)-b(s;s)$  is much smaller than  $b(s';s)$  (i.e.,  
 11  $[b(s';s)-b(s;s)]/b(s';s) = O(\sigma)$ ) (see Appendix C for the derivation). The expected waiting  
 12 time for the next invasion event is given by  $T = 1/\int E(s';s)ds'$ . When an invasion event  
 13 occurs, the successfully invading mutants  $s'$  follow the invasion-event probability density

$$14 \quad P(s';s) = TE(s';s). \quad (8b)$$

15 For a polymorphism with resident phenotypes  $\mathbf{s}_R = (s_1, \dots, s_N)$ , the probability density per  
 16 unit time of successful invasion by a mutant phenotype  $s'$  originating from the resident  $s_i$  is  
 17 given by an approximation analogous to the monomorphic case, Eq. (8a),

$$18 \quad E_i(s';s_1, \dots, s_N) \approx \mu\hat{n}_i M(s'-s_i) f(s';s_1, \dots, s_N)_+, \quad (9a)$$

19 where  $\hat{n}_i$  is the equilibrium population size of the resident  $s_i$ . As in Eq. (8b), we can also  
 20 define an invasion-event probability density

$$21 \quad P_i(s';s_1, \dots, s_N) = TE_i(s';s_1, \dots, s_N), \quad (9b)$$

22 where  $T = 1/\int \sum_{i=1}^N E_i(s';s_1, \dots, s_N) ds'$  is the expected waiting time for the next invasion  
 23 event. Notice that the invasion event is identified by the combination  $(s', i)$  of the mutant  
 24 phenotype  $s'$  and its parental resident  $s_i$ . Consequently, the invasion event probability den-



1 sities are normalized according to  $\int \sum_{i=1}^N P_i(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N) d\mathbf{s}' = 1$ . When only a single resident  
 2 exists, Eqs. (9a) and (9b) are identical to Eqs. (8a) and (8b), respectively.

3 The invasion by a mutant leads the community to a new population dynamical equilibrium.  
 4 In most cases, the mutant replaces only its parental resident, while under certain conditions  
 5 the coexistence of both, extinction of both, or extinction of other residents may occur. A se-  
 6 quence of such invasions specifies a succession dynamics of resident phenotypes, which is  
 7 called a trait-substitution sequence (Metz et al., 1996).

8 If the invasion event is calculated stochastically according to Eqs. (8b) and (9b), the  
 9 resultant trait substitution is called an oligomorphic stochastic process (Ito and Dieckmann,  
 10 2007). When considering an initial monomorphic resident  $\mathbf{s}_a$  and a mutant-invasion  
 11 sequence

$$12 \quad \mathbf{I} = (\mathbf{s}'(1), \dots, \mathbf{s}'(k), \dots, \mathbf{s}'(K)), \quad (10a)$$

13 where  $\mathbf{s}'(k)$  is the mutant that invades in the  $k$ th invasion event, the trait-substitution  
 14 sequence is denoted by

$$15 \quad \mathbf{R}(\mathbf{I}; \mathbf{s}_a) = (\mathbf{s}_R(1), \dots, \mathbf{s}_R(k), \dots, \mathbf{s}_R(K)), \quad (10b)$$

16 where  $\mathbf{s}_R(k) = (\mathbf{s}_1(k), \dots, \mathbf{s}_{N(k)}(k))$  is an  $N(k)$ -dimensional vector composed of the  $N(k)$   
 17 resident phenotypes that coexist after the invasion of  $\mathbf{s}'(k)$ . This kind of trait-substitution  
 18 sequence constitutes an evolutionary path, which we call an oligomorphic stochastic invasion  
 19 path (OSIP).

20 If the  $k$ th invasion event leads to the extinction of the entire community, no further inva-  
 21 sions occur. In this case, the lengths of  $\mathbf{I}$  and  $\mathbf{R}(\mathbf{I}; \mathbf{s}_a)$  are limited by  $k$ . In this study, we  
 22 condition all analyses on the absence of complete community extinction.

## 23 **4.2 Definition of maximum-likelihood-invasion paths**

24 We now introduce the concept of maximum-likelihood invasion. Specifically, we define a  
 25 maximum-likelihood-invasion event as the combination of the mutant  $\mathbf{s}'_{\text{MLI}}$  and its parental  
 26 resident  $\mathbf{s}_i$  with  $i = i_{\text{MLI}}$  that maximizes the invasion-event probability density, Eq. (9b),  
 27 across all  $\mathbf{s}'$  and  $i$  for a given set of residents,

$$1 \quad (\mathbf{s}'_{\text{MLI}}, i_{\text{MLI}}) = \operatorname{argmax}_{(s', i)} P_i(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N), \quad (11)$$

2 where we refer to  $\mathbf{s}'_{\text{MLI}}$  and  $\mathbf{s}_{i_{\text{MLI}}}$  as the MLI mutant and MLI resident, respectively, and  
 3 denote  $\mathbf{s}_{i_{\text{MLI}}}$  by  $\mathbf{s}_{\text{MLI}}$  for convenience. A maximum-likelihood-invasion path (MLIP) is a  
 4 trait-substitution sequence formed by MLI events, denoted by

5  $\mathbf{I}_{\text{MLI}} = (\mathbf{s}'_{\text{MLI}}(0), \dots, \mathbf{s}'_{\text{MLI}}(k), \dots, \mathbf{s}'_{\text{MLI}}(K))$ . The MLIP, which is expressed as  $\mathbf{R}(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a)$  ~~with the~~  
 6 ~~for an~~ initial monomorphic resident  $\mathbf{s}_a$ , is included in the set of all corresponding possible  
 7 OSIPs  $\mathbf{R}(\mathbf{I}; \mathbf{s}_a)$ .

8 Note that the MLI mutational steps  $|\mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}}|$  are bounded by  $\sqrt{2}\sigma$ , if invasion-fitness  
 9 functions are approximated by quadratic forms of  $\mathbf{s}'$  (e.g., Eqs. 2) and if mutation probability  
 10 distributions are approximated by multivariate Gaussian functions (Appendix F).

11 Also note that the MLIP does not give the maximum-likelihood OSIP, which would require  
 12 maximization of the likelihood at the level of the mutant-invasion sequence rather than at the  
 13 level of individual ~~mutant~~-invasion events. Although such sequence-level maximization  
 14 would be more appropriate for our purpose, it seems analytically intractable. On the other  
 15 hand, the event-level maximization defined by MLIPs is analytically tractable, and the MLIP  
 16 is still expected to have a relatively large likelihood among corresponding OSIPs. Likewise,  
 17 as illustrated by our numerical results in Section 6, when an MLIP  $\mathbf{R}(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a)$  exhibits evolu-  
 18 tionary branching, then a large fraction of the corresponding OSIPs  $\mathbf{R}(\mathbf{I}; \mathbf{s}_a)$  also exhibit  
 19 evolutionary branching.

## 20 **5 Conditions for evolutionary branching along MLIPs**

21 In this section, we derive sufficient conditions for evolutionary branching along MLIPs, in  
 22 terms of the properties of the normal form for invasion-fitness functions with significant  
 23 sensitivity difference, Eq. (3).

### 24 **5.1 Further rescaling**

25 Here we assume that the base point of expansion  $\mathbf{s}_0 = (x_0, y_0)^T$  is on a convergence-stable  
 26 line  $x = x_b$  that satisfies univariate conditions for evolutionary branching, Eqs. (6). To

1 simplify the analysis, we adjust the trait space as follows, without loss of generality. First, we  
 2 shift the origin of the trait space to the base point  $\mathbf{s}_0$  so that  $\mathbf{s}_0 = (0,0)^T$  and  $x_b = 0$ .  
 3 Second, we rescale the trait space so that  $\sigma = 1$ . (In this case, magnitude differences among  
 4  $|\delta\mathbf{s}| = O(\sigma)$ ,  $|\delta\mathbf{s}|^2 = O(\sigma^2)$  with  $\sigma \ll 1$  are transformed into those among the corresponding  
 5 derivative coefficients  $\mathbf{G}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$ , while the magnitudes of  $|\delta\mathbf{s}|$ ,  $|\delta\mathbf{s}|^2$  themselves  
 6 become similar to each other.) Third, we rescale time and potentially flip the direction of the  
 7  $y$ -axis so that  $G_y = 1$ . For simplicity, we consider the first- and second-order terms only.  
 8 Consequently,  $f(\mathbf{s}';\mathbf{s})$  is given by

$$9 \quad f(\mathbf{s}';\mathbf{s}) = \delta y + D\delta x^2 + Cx\delta x, \quad (12a)$$

10 where

$$11 \quad C = \frac{\sigma C_{xx}}{|G_y|} < 0 \quad (12b)$$

12 and

$$13 \quad D = \frac{\sigma D_{xx}}{2|G_y|} > 0, \quad (12c)$$

14 with  $\sigma$  before the rescaling (i.e.,  $\sigma \ll 1$ ). In the simplified normal form in Eq. (12a), only  
 15 two dimensionless parameters  $D$  and  $C$  determine the geometry of the fitness landscape.  
 16 This geometry not only determines the fitness landscape's shapes ( $D$ ), but also how the  
 17 landscape changes when the resident phenotype is varied ( $C$ ). Eq. (12a) then shows that any  
 18 possible fitness landscape  $f(\mathbf{s}';\mathbf{s})$  can be obtained from  $f(\mathbf{s}';(0,0)^T) = \delta y + D\delta x^2$  by a  
 19 parallel shift, i.e.,  $f(\mathbf{s}';\mathbf{s}) = f(\mathbf{s}' - \mathbf{s}_w; (0,0)^T)$  with  $\mathbf{s}_w = (x_w, y_w)^T = (\frac{1}{4}C^2x^2 / D^2, \frac{1}{2}Cx / D)^T$ .  
 20 This means that the contour curve  $f(\mathbf{s}';\mathbf{s}) = 0$ , given by  $y' = -D(x' - x - x_w)^2 + y + y_w$ ,  
 21 always has a constant parabolic shape specified by  $D$ , so that the position of this curve  
 22 determines the fitness landscape (Fig. 3a).

23 In the next two subsections, we derive conditions on  $D$  and  $C$  for evolutionary  
 24 branching along MLIPs. We first obtain conditions on MLI mutants,  $\mathbf{s}'_{MLI}$ , for evolutionary

1 branching. Then we analyze these conditions considering the dependence of  $s'_{\text{MLI}}$  on  $D$  and  
 2  $C$ , which provides conditions on these two parameters for evolutionary branching.

### 3 **5.2 Conditions on MLI mutants for evolutionary branching**

4 Here we obtain conditions on  $s'_{\text{MLI}}$  for evolutionary branching. The process of evolutionary  
 5 branching can be decomposed into two steps: emergence of protected dimorphism (dimorphic  
 6 emergence) and directional evolution of these two morphs in opposite directions (dimorphic  
 7 divergence). First, sufficient conditions for dimorphic emergence and specific evolutionary  
 8 dynamics ensured by these conditions are expressed as follows (see Appendix G for the  
 9 derivation).

10 **Lemma 1:** Suppose the conditions for dimorphic emergence below hold. Then, for an  
 11 arbitrary initial resident  $s_a$ , repeated invasions by  $s'_{\text{MLI}}$  first induce directional  
 12 evolution of the population toward the convergence-stable line  $x = x_b = 0$ , and then  
 13 bring about protected dimorphism after sufficient convergence.

14 **Conditions for dimorphic emergence:** Any  $s'_{\text{MLI}} = (x'_{\text{MLI}}, y'_{\text{MLI}})^T$  under an arbitrary  
 15 monomorphic resident  $s$  satisfies

$$16 \quad \begin{cases} x'_{\text{MLI}} > x & \text{for } x < 0 \\ x'_{\text{MLI}} < x & \text{for } x > 0, \end{cases} \quad (13a)$$

17 and

$$18 \quad |y'_{\text{MLI}} - y| < D(x'_{\text{MLI}} - x)^2. \quad (13b)$$

19 The set of mutants satisfying inequalities (13) ~~are illustrated as ais~~ indicated by the white  
 20 region in Fig. 3c. Clearly, inequalities (13a) ensure that the MLI mutant is always closer than  
 21 the resident to the convergence-stable line, resulting in directional evolution toward this line  
 22 as long as the mutant replaces the resident. Inequality (13b) restricts the deviation of the MLI  
 23 mutant from the resident along the  $y$ -axis, and thus ensures that a protected dimorphism  
 24 (with  $f(s';s) > 0$  and  $f(s;s') > 0$ ) emerges after sufficient convergence to the line.

1 After emergence of an initial protected dimorphism, we denote the coexisting phenotypes  
 2 by  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , with  $x_1 < x_2$ , without loss of generality (Fig. 3b). A sufficient condition for  
 3 dimorphic divergence and specific evolutionary dynamics ensured by these conditions are  
 4 expressed as follows (see Appendix H for the derivation).

5 **Lemma 2:** Suppose the conditions for dimorphic divergence below hold. Then, for  
 6 any initial protected dimorphism of  $\mathbf{s}_1$  and  $\mathbf{s}_2$  emerged under the conditions for  
 7 dimorphic emergence, subsequent invasions by  $\mathbf{s}'_{\text{MLI}}$  continue directional evolution of  
 8 the two morphs in opposite directions in  $x$  without collapse.

9 **Conditions for dimorphic divergence:** Any  $\mathbf{s}'_{\text{MLI}}$  satisfies

$$10 \quad x'_{\text{MLI}} < x_1 \quad \text{and} \quad |y'_{\text{MLI}} - y_1| < D(x'_{\text{MLI}} - x_1)^2 \quad (14a)$$

11 or

$$12 \quad x'_{\text{MLI}} > x_2 \quad \text{and} \quad |y'_{\text{MLI}} - y_2| < D(x'_{\text{MLI}} - x_2)^2, \quad (14b)$$

13 where  $x_1 < x_2$  is assumed without loss of generality.

14 The set mutants satisfying inequalities (14) ~~are illustrated as~~ indicated by the white  
 15 regions in Fig. 3d. In each invasion step,  $\mathbf{s}'_{\text{MLI}}$  replaces only its parental resident, so the  
 16 divergence of the new dimorphism in  $x$  is larger than that of  $\mathbf{s}_1$  and  $\mathbf{s}_2$ .

17 Clearly, if conditions for dimorphic emergence and ~~that~~ for dimorphic divergence both  
 18 hold, evolutionary branching along MLIPs inevitably occurs for an arbitrary initial resident  
 19  $\mathbf{s}_a$ .

### 20 5.3 MLIP condition

21 As  $\mathbf{s}'_{\text{MLI}}$  is a function of  $D$  and  $C$ , substituting this function into the conditions for  
 22 dimorphic emergence and divergence above and solving those for  $D$  and  $C$  gives  
 23 conditions on these parameters for evolutionary branching.

1 To derive  $\mathbf{s}'_{\text{MLI}}$  as a function of  $D$  and  $C$ , we explicitly define the mutation distribution.  
 2 For analytical tractability, we assume that the mutation distribution is approximated by a two-  
 3 dimensional Gaussian distribution, which is expressed in the normalized and rescaled trait  
 4 space  $\mathbf{s} = (x, y)^{\text{T}}$  as

$$5 \quad M(\delta\mathbf{s}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}|\delta\mathbf{s}|^2\right), \quad (15)$$

6 where the standard deviation of mutational step sizes is scaled to 1.

7 Under monomorphism with phenotype  $\mathbf{s}$ , the MLI mutant  $\mathbf{s}'_{\text{MLI}}$ , which maximizes the  
 8 invasion-event probability density, is given by the  $\mathbf{s}'$  that maximizes Eq. (8b),

$$9 \quad \begin{aligned} P(\mathbf{s}'; \mathbf{s}) &= T\mu\hat{n}M(\delta\mathbf{s})f(\mathbf{s}'; \mathbf{s})_+ \\ &= \frac{T\mu\hat{n}}{2\pi} \exp\left(-\frac{1}{2}[\delta x^2 + \delta y^2]\right) [\delta y + D\delta x^2 + Cx\delta x]_+. \end{aligned} \quad (16)$$

10 We first focus on the special case that  $\mathbf{s}$  is located exactly on the convergence-stable line,  
 11 i.e.,  $\mathbf{s} = (0, y)^{\text{T}}$  with arbitrary  $y$ . In this case,  $\mathbf{s}'_{\text{MLI}}$  is given by

$$12 \quad \mathbf{s}'_{\text{MLI}} = \begin{cases} (0, 1)^{\text{T}} + \mathbf{s} & \text{for } D < \frac{1}{2} \\ \left(\pm \frac{\sqrt{4D^2 - 1}}{\sqrt{2D}}, \frac{1}{2D}\right)^{\text{T}} + \mathbf{s} & \text{for } D \geq \frac{1}{2}. \end{cases} \quad (17)$$

13 Substitution of Eq. (17) and  $\mathbf{s} = (0, y)^{\text{T}}$  into the inequality condition (13b) for dimorphic  
 14 emergence yields

$$15 \quad D > \frac{1}{\sqrt{2}}. \quad (18)$$

16 We do not need to examine inequalities (13a), which are of interest only for  $\mathbf{s} = (x, y)^{\text{T}}$  with  
 17  $x \neq 0$ . Even if the resident is not located on the convergence-stable line, i.e.,  $\mathbf{s} = (x, y)^{\text{T}}$  with  
 18  $x \neq 0$ , inequalities (13) still hold under  $D > 1/\sqrt{2}$ , as shown in Appendix D. Therefore,  
 19 conditions for dimorphic emergence are satisfied if  $D > 1/\sqrt{2}$  holds. In this case,  $\mathbf{s}'_{\text{MLI}}$  is  
 20 always located inside of the dark gray rectangle in Fig. 3e.

21 Moreover, under  $D > 1/\sqrt{2}$ , the derivation in Appendix D ~~derives~~ shows that  $\mathbf{s}'_{\text{MLI}}$  always  
 22 satisfies the condition for dimorphic divergence. Therefore, inequality (18),  $D > 1/\sqrt{2}$ , is a

1 sufficient condition for evolutionary branching along MLIPs starting from an arbitrary initial  
2 monomorphic resident. More specifically, we have

3 **Theorem 1:** Suppose a normalized and rescaled invasion-fitness function having  
4 significant sensitivity difference, ~~which is~~ expressed in ~~a~~ form of Eq. (3) with  $\sigma = 1$ ,  
5 satisfies both  $C < 0$  and inequality (18),  $D > 1/\sqrt{2}$ . Then, any MLIP starting from  
6 an arbitrary initial monomorphic resident monotonically converges toward the  
7 convergence-stable line,  $x = 0$ , and brings about protected dimorphism, which leads  
8 to dimorphic divergence without collapse, i.e., evolutionary branching.

9 We thus call inequality (18),  $D > 1/\sqrt{2}$ , the MLIP condition for evolutionary branching, and  
10 refer to a convergence-stable line satisfying this condition as an evolutionary-branching line.

#### 11 **5.4 Directional evolution sufficient for evolutionary branching**

12 Under the MLIP condition, dimorphism with  $|x_2 - x_1| \geq \Delta x^*$  for arbitrary  $\Delta x^* > 0$  emerges  
13 before the population directionally has evolved by

$$14 \quad \bar{y} - y_a = L_{y_0}(|x_a|, \Delta x^*) = (|x_a| + \Delta x^*) / \sqrt{2}, \quad (19)$$

15 where the second equality defines the function  $L_{y_0}(|x_a|, \Delta x^*)$ , and  $\mathbf{s}_a = (x_a, y_a)^T$  is the initial  
16 monomorphic resident (see Appendix J for the derivation). The  $\bar{y}$  is the mean value of  $y$ ,  
17 given by  $\bar{y} = y$  for monomorphism or by  $\bar{y} = (\hat{n}_1 y_1 + \hat{n}_2 y_2) / (\hat{n}_1 + \hat{n}_2)$  for dimorphism, where  
18  $\hat{n}_1$  and  $\hat{n}_2$  are the equilibrium population sizes of  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , respectively.

### 19 **6 Numerical examination of MLIP condition**

20 In this section we investigate how the MLIP condition is related to the likelihood of  
21 evolutionary branching in numerically calculated MLIPs, OSIPs, and polymorphic stochastic  
22 invasion paths (PSIPs) in which mutation rates are not small. See Appendices K, M, and L for  
23 details on ~~the calculation~~ algorithms and initial ~~settings~~ conditions.

24 When deriving the MLIP condition, we assumed the ~~bivariate Gaussian~~ mutation  
25 distribution defined in Eq. (15), ~~called bivariate Gaussian here.~~ The resultant MLIP

1 condition may also be applicable to other types of mutation distributions. To examine this  
 2 kind of robustness, below we investigate ~~on an~~ additional three different mutation distributions  
 3 for the calculation of OSIPs and PSIPs. ~~A~~The bivariate fixed-step distribution has possible  
 4 mutations that are bounded on a circle (Fig. 5b). ~~A~~The univariate Gaussian distribution  
 5 applies when mutations in  $x$  and  $y$  occur separately, each following a one-dimensional  
 6 Gaussian distribution (Fig. 5c). ~~A~~The univariate fixed-step distribution also limits possible  
 7 mutations to affect either  $x$  or  $y$ , but with fixed step sizes (Fig. 5d). See Appendix L for  
 8 details on these mutation distributions.

9 The ~~cumulative~~ likelihood of evolutionary branching is measured as a probability  
 10  $p(L_y - \hat{L}_{y0})$ , where  $L_y$  is the length of directional evolution in  $y$  along MLIPs, OSIPs, or  
 11 PSIPs ~~until evolutionary branching has occurred~~, while  $\hat{L}_{y0}$ , calculated ~~with~~through Eq. (19),  
 12 is the length of directional evolution in  $y$  along MLIPs sufficient for the occurrence of evo-  
 13 lutionary branching (see Appendix K for details on  $\hat{L}_{y0}$ ). Thus,  $p(L_y - \hat{L}_{y0})$  ~~measures~~gives  
 14 ~~the cumulative probability of evolutionary branching before~~when the population has direc-  
 15 tionally evolved in  $y$  by  $L_y$ , ~~beyond what is implied by the MLIP condition ( $\hat{L}_{y0}$ )~~ $L_y - \hat{L}_{y0}$   
 16 ~~is the additionally needed directional evolution in  $y$ , relative to what is implied by the MLIP~~  
 17 ~~condition~~. In the case of MLIPs,  $p(L_y - \hat{L}_{y0}) = 1$  clearly holds for  $L_y - \hat{L}_{y0} = 0$ . In the case of  
 18 OSIPs and PSIPs, when values of  $p(L_y - \hat{L}_{y0})$  for  $L_y - \hat{L}_{y0} = 0$  are close to 1, this indicates  
 19 that the MLIP condition is working well also under such relaxed conditions. However,  
 20  $p(L_y - \hat{L}_{y0})$  never reaches 1 in OSIPs, differently from MLIPs. One reason is that even under  
 21 very large  $D$  there are non-zero probabilities for repeated mutant invasions only in the  $y$ -  
 22 direction, ~~providing~~causing directional evolution in the  $y$ -direction. Another reason is that  
 23 even after ~~the~~ emergence of a protected dimorphism, ~~the~~this dimorphism may collapse by  
 24 subsequent mutant invasions in the case of OSIPs. When a dimorphism has collapsed, leaving  
 25 behind a monomorphic resident, by the definition of OSIPs, the information about the col-  
 26 lapse itself is lost, and it is only the remaining resident that determines the likelihood of evo-  
 27 lutionary branching in the “next trial”. A sufficiently large  $D$  is expected to induce evolu-  
 28 tionary branching within a few trials, keeping the total directional evolution in the  $y$ -  
 29 direction short, which results in a high value of  $p(L_y - \hat{L}_{y0})$  for  $L_y - \hat{L}_{y0} = 0$  ~~and vice versa~~.



## 1 6.1 Sufficient vs. necessary conditions: MLIPs

2 Fig. 4a shows the branching likelihood in MLIPs under the bivariate Gaussian mutation  
3 distribution for varying  $C < 0$  and  $D > 0$ : the contour curves indicate where a 97%  
4 cumulative probability of ~~Fig. 4a shows the occurrence of~~ evolutionary branching ~~in MLIPs is~~  
5 reached for  $L_y - \hat{L}_{y0} = 0$ ,  $100 L_y - \hat{L}_{y0} = 100$ , and  $200 L_y - \hat{L}_{y0} = 200$  (i.e.,  $p(0) = 0.97$ ,  
6  $p(100) = 0.97$ , and  $p(200) = 0.97$ ), ~~at various values for  $C < 0$  and  $D > 0$  under bivariate~~  
7 ~~Gaussian mutation~~. For  $D > 1/\sqrt{2}$ , MLIPs quickly undergo evolutionary branching in the  
8 gray area in Fig. 4a, while they do not undergo evolutionary branching in the white area in  
9 Fig. 4a. Examples of branching and non-branching MLIPs are shown as gray curves in Fig. 4b  
10 and Fig. 4c,d, respectively. Importantly, the threshold  $D = 1/\sqrt{2}$  provided by the MLIP  
11 condition and indicated by the ~~black-dashed~~ line in Fig. 4a characterizes very well the area  
12 that ensures the occurrence of evolutionary branching. In particular, the MLIP condition  
13  $D > 1/\sqrt{2}$  seems to give a necessary and sufficient condition as  $C$  converges to 0.

## 14 6.2 Robustness of MLIP condition: OSIPs

15 When the MLIP condition  $D > 1/\sqrt{2}$  holds, OSIPs tend to undergo immediate evolutionary  
16 branching (black curves in Fig. 4b). On the other hand, even for  $D < 1/\sqrt{2}$ , OSIPs may still  
17 undergo evolutionary branching (black curves in Fig. 4c). In this case, however, the required  
18  $L_y - \hat{L}_{y0}$  becomes large as  $D$  is decreased. As  $D$  is decreased further, evolutionary  
19 branching may not be observed even for very large  $L_y - \hat{L}_{y0}$  (black curves in Fig. 4d).

20 Fig. 5a shows the branching likelihood in OSIPs under ~~the~~ bivariate Gaussian mutation dis-  
21 tribution for varying  $C < 0$  and  $D > 0$ : ~~t-~~The contour curves indicate ~~where a 97% likeli-~~  
22 ~~hood-cumulative probability of evolutionary branching is reached~~ for  $L_y - \hat{L}_{y0} = 0$ , 100, and  
23 200 (i.e.,  $p(0) = 0.97$ ,  $p(100) = 0.97$ , and  $p(200) = 0.97$ ). We see that more than 97%  
24 branching likelihood is attained for  $L_y - \hat{L}_{y0} = 0$ , as expected by the MLIP condition. Similar-  
25 ly, more than 90% branching likelihood is attained for  $L_y - \hat{L}_{y0} = 0$  for each of the three oth-  
26 er mutation distributions (Fig. 5b-d), as long as the mutation rate in  $y$  is not very small com-  
27 pared to that in  $x$  (i.e.,  $\mu_y / \mu_x \geq 0.05$ ) for the univariate Gaussian and univariate fixed-step

1 mutation distributions. Thus, for the examined OSIPs, the MLIP condition turns out to be ro-  
2 bust (at a likelihood level of 97%) as an almost sufficient condition for evolutionary branch-  
3 ing; it is also robust against variations in mutation distributions.

### 4 **6.3 Robustness of MLIP condition: PSIPs**

5 For PSIPs ~~assume that~~, mutation rates ~~are~~ need not be low. In this case, evolutionary dynamics  
6 are no longer, ~~in contrast with OSIPs~~, given by trait-substitution sequences (as for OSIPs), but  
7 by gradual changes of polymorphic phenotype distributions. Population dynamics of PSIPs  
8 are calculated based on the stochastic sequence of individual births and deaths (Dieckmann  
9 and Law, 1996). The stochastic effects become large when fitness gradients and curvatures  
10 are both weak and/or population sizes are small. In this case, the likelihood of evolutionary  
11 branching in PSIPs, in contrast with OSIPs, may be affected not only by  $C$  and  $D$ , but also  
12 by other parameters, such as the mutational step size  $\sigma$ , the mutation rate  $\mu$ , and the  
13 carrying capacity along the evolutionary-branching line,  $K_0$ . We have numerically confirmed  
14 that the MLIP condition is still useful for characterizing evolutionary branching in PSIPs  
15 across a certain range of parameter values. For example,  $D > 1/\sqrt{2}$  provides  $p(0) > 0.9$   
16 under all four mutation distributions for  $0.001 \leq \sigma \leq 0.01$ ,  $300 \leq K_0 \leq 10000$ , and  
17  $3.3 \times 10^{-5} \leq \mu \leq 1 \times 10^{-1}$ , with  $3 \times 10^{-3} \leq \mu \sigma K_0 \leq 3 \times 10^{-2}$  (results not shown).

18 Fig. 5e-h show the branching likelihood in PSIPs for varying  $C < 0$  and  $D > 0$ , with  
19  $\sigma = 0.01$ ,  $K_0 = 600$ , and  $\mu = 5.1 \times 10^{-3}$ . The contour curves indicate where a 95% likelihood  
20 cumulative probability of evolutionary branching is reached for  $L_y - \hat{L}_{y0} = 0, 40, \text{ and } 80$   
21 (i.e.,  $p(0) = 0.95$ ,  $p(40) = 0.95$ , and  $p(80) = 0.95$ ). We see that more than 95% branching  
22 likelihood is attained for  $L_y - \hat{L}_{y0} = 0$  under all four mutation distributions, as long as the  
23 mutation rate in  $y$  is not very small compared to that in  $x$  (i.e.,  $\mu_y / \mu_x \geq 0.05$ ) for the uni-  
24 variate Gaussian and univariate fixed-step mutation distributions. Thus, for the examined  
25 PSIPs, the MLIP condition turns out to be robust as a good indicator for evolutionary branch-  
26 ing, even when mutation rates are not small and/or mutation distributions other than bivariate  
27 Gaussian are considered.

# 1 7 Conditions for evolutionary-branching lines

2 In this section, we first summarize the conditions for evolutionary-branching lines in bivariate  
3 trait spaces. Second, we extend these conditions to multivariate trait spaces. Third, we explain  
4 how to find evolutionary-branching lines or manifolds in arbitrary trait spaces with arbitrary  
5 dimensionality.

## 6 7.1 Conditions for evolutionary-branching lines in bivariate trait spaces

7 By an appropriate affine transformation, arbitrary trait spaces  $\mathbf{S} = (X, Y)^T$  can be normalized  
8 into a representation  $\mathbf{s} = (x, y)^T$  with isotropic mutation with standard deviation  $\sigma$ . In this  
9 normalized trait space, the invasion-fitness function can be expanded around  $\mathbf{s}_0$  as shown in  
10 Eq. (2a),

$$11 \quad f(\mathbf{s}'; \mathbf{s}) = \mathbf{G}\delta\mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C}\delta\mathbf{s} + \frac{1}{2} \delta\mathbf{s}^T \mathbf{D}\delta\mathbf{s} + O(\sigma^3),$$

12 where  $\delta\mathbf{s} = (\delta x, \delta y)^T = \mathbf{s}' - \mathbf{s}$ . If the  $x$ - and  $y$ -axes can be adjusted such that Eq. (4) holds,

$$13 \quad \frac{|G_y| + |C_{xy}| + |C_{yx}| + |C_{yy}| + |D_{xy}| + |D_{yy}|}{|G_x| + |C_{xx}| + |D_{xx}|} = O(\sigma),$$

14 then  $f(\mathbf{s}'; \mathbf{s})$  is significantly less sensitive to trait  $y$  than to trait  $x$ . In this case, Eq. (2a) is  
15 transformed into Eq. (3),

$$16 \quad f(\mathbf{s}'; \mathbf{s}) = G_x \delta x + C_{xx} (x - x_0) \delta x + \frac{1}{2} D_{xx} \delta x^2 + G_y \delta y + O(\sigma^3).$$

17 If Eq. (6a) holds,

$$18 \quad G_x = 0,$$

19 then there exists a singular line  $x = x_0$  denoted by  $x_b$ . This line is convergence stable if Eq.  
20 (6b) holds,

$$21 \quad C_{xx} < 0$$

22 (i.e., if  $C < 0$ ). The convergence-stable line causes evolutionary branching along MLIPs  
23 (maximum-likelihood-invasion paths), if inequality (18), i.e., the MLIP condition, holds,

$$D = \frac{\sigma D_{xx}}{2|G_y|} > \frac{1}{\sqrt{2}}.$$

In this case, any MLIP starting from a monomorphic resident  $\mathbf{s}_a = (x_a, y_a)^T$  with  $|\mathbf{s}_a - \mathbf{s}_0| = O(\sigma)$  inevitably converges to the line and brings about evolutionary branching. Here we refer to Eqs. (4), (6a), (6b), and (18) as the conditions for evolutionary-branching lines, which is summarized as follows.

**Theorem 2:** Suppose that  $\mathbf{s}_0$  a point  $\mathbf{s}_0$  in a normalized trait space  $\mathbf{s}$  satisfies the conditions for evolutionary-branching lines below. Then, there exists an evolutionary-branching line passing through  $\mathbf{s}_0$  and parallel with the  $y$ -axis of the trait space. In this case, any MLIP starting from a monomorphic resident  $\mathbf{s}_a = (x_a, y_a)^T$  with  $|\mathbf{s}_a - \mathbf{s}_0| = O(\sigma)$  monotonically converges to the line and brings about a protected dimorphism, which leads to dimorphic divergence in  $x$  without collapse, as long as their deviations from  $\mathbf{s}_0$  are  $O(\sigma)$ .

**Conditions for evolutionary-branching lines along MLIPs in normalized bivariate trait spaces:** The  $x$ - and  $y$ -axes can be adjusted by rotation such that the first and second derivatives of the invasion-fitness function at  $\mathbf{s}_0$  satisfy all of Eqs. (4), (6a), (6b), and (18).

Rescaling trait spaces such that  $\sigma = 1$  and applying Theorem 1 proves this theorem.

If these conditions for evolutionary-branching lines hold, then evolutionary branching occurs with high likelihood in evolutionary paths even under relaxed assumptions (i.e., in OSIPs and PSIPs, as shown in Section 6). As the sensitivity difference goes to infinity, which means that the right-hand side of Eq. (4) converges to zero, the conditions for evolutionary-branching lines converge to the univariate conditions for evolutionary branching, given by Eqs. (6).

Notice that the MLIP condition requires that  $\sigma$  is not infinitesimally small, but finite; otherwise, satisfying this inequality is impossible. Thus, as long as the population is directionally evolving, its evolutionary branching requires finite mutational step sizes. Conversely,  $\sigma$  can have large magnitudes, as long as approximating the invasion-fitness func-

1 tions ~~with~~ by the normal form, in Eq. (3), is appropriate ~~in~~ at the scale of that  $\sigma$ . In this case,  
 2 ~~A~~ a single large mutational step may generate a mutant such that the mutant and resident to-  
 3 gether straddle an evolutionary-branching line, resulting in protected dimorphism with a rela-  
 4 tively large phenotypic difference. Although this sounds different from the process of evolu-  
 5 tionary branching with small phenotypic difference, the two cases are formally equivalent as  
 6 long as the invasion-fitness function is well approximated by Eq. (3), i.e., terms ~~more than be-~~  
 7 ~~yond~~ the second order are negligible. This becomes clear when the two trait spaces ~~is~~ are re-  
 8 scaled, and thus become comparable.

## 9 7.2 Conditions for evolutionary-branching lines in multivariate trait spaces

10 The conditions for evolutionary-branching lines explained above can be applied also to  
 11 multivariate trait spaces: for this we only have to extend the condition for significant  
 12 sensitivity difference, as explained below.

13 As before, an arbitrary  $L$ -variate trait space  $\mathbf{S} = (U_1, \dots, U_L)^T$  can be normalized by an ap-  
 14 propriate affine transformation into a representation  $\mathbf{s} = (u_1, \dots, u_L)^T$  with isotropic mutation  
 15 with standard deviation  $\sigma$  (see Appendix P). In this normalized trait space, the invasion-  
 16 fitness function can be expanded as in the bivariate case,

$$17 \quad f(\mathbf{s}'; \mathbf{s}) = \mathbf{G}\delta\mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C}\delta\mathbf{s} + \frac{1}{2} \delta\mathbf{s}^T \mathbf{D}\delta\mathbf{s} + O(\sigma^3), \quad (20)$$

18 where  $\delta\mathbf{s} = (\delta u_1, \dots, \delta u_L)^T = \mathbf{s}' - \mathbf{s}$ ,  $\mathbf{G}$  is a  $L$ -dimensional row vector, and  $\mathbf{C}$  and  $\mathbf{D}$  are  $L$ -  
 19 by- $L$  matrices.

20 In a manner similar to the bivariate case, the trait space can be decomposed into a  $\hat{L}$ -  
 21 variate sensitive subspace  $\mathbf{x} = (x_1, \dots, x_{\hat{L}})^T = (u_1, \dots, u_{\hat{L}})^T$  and an  $(L - \hat{L})$ -variate insensitive  
 22 subspace  $\mathbf{y} = (y_1, \dots, y_{L-\hat{L}})^T = (u_{\hat{L}+1}, \dots, u_L)^T$ , if trait axes can be adjusted such that

$$23 \quad \frac{|G_j| + |C_{ij}| + |C_{ji}| + |C_{jj}| + |D_{ij}| + |D_{jj}|}{|G_i| + |C_{ii}| + |D_{ii}|} = O(\sigma) \quad (21)$$

24 holds for all  $i = 1, \dots, \hat{L}$  and  $j = \hat{L} + 1, \dots, L$ , where  $G_j$  is the  $j$ th component of  $\mathbf{G}$ , and  
 25  $C_{ij}$  and  $D_{ij}$  are the  $(i, j)$ th components of  $\mathbf{C}$  and  $\mathbf{D}$ , respectively.

1 If Eq. (21) holds for  $\hat{L} = 1$ , i.e., the sensitive subspace is univariate, then Eq. (20)  
 2 simplifies to the normal form for invasion-fitness functions with significant sensitivity  
 3 difference,

$$4 \quad f(\mathbf{s}'; \mathbf{s}) = G_x \delta x + C_{xx} (x - x_0) \delta x + \frac{1}{2} D_{xx} \delta x^2 + G_y \delta y + O(\sigma^3), \quad (22)$$

5 where  $G_x = G_1$ ,  $C_{xx} = C_{11}$ ,  $D_{xx} = D_{11}$ ,  $G_y = |\mathbf{G}_y| = |(G_2, \dots, G_L)|$ , and  $\delta y = \mathbf{G}_y (\mathbf{y}' - \mathbf{y}) / G_y$ .  
 6 See Appendix O for derivations of Eqs. (21) and (22). Notice that the insensitive subspace  $\mathbf{y}$   
 7 contributes to the invasion fitness only through the element parallel to the fitness gradient  
 8  $\mathbf{G}_y$ . Thus, local evolutionary dynamics based on Eq. (22) can be contracted into a bivariate  
 9 trait space  $(x, y)^T$ . As Eq. (22) is identical to the bivariate invasion-fitness functions with  
 10 significant sensitivity difference, Eq. (3), the conditions for evolutionary-branching lines in  
 11 bivariate trait spaces, Eqs. (4), (6a), (6b), and (18), can be applied **as they are directly**. If  
 12  $\mathbf{s} = \mathbf{s}_0$  satisfies those conditions, it forms in the trait space  $\mathbf{s}$  an  $(L-1)$ -dimensional  
 13 evolutionary-branching manifold,  $x = x_0$ . Thus, **†Theorem 2** is translated as follows.

14 **Theorem 3:** Suppose that a point  $\mathbf{s}_0$  in a normalized  $L$ -dimensional trait space  $\mathbf{s}$   
 15 satisfies **the** conditions for evolutionary-branching manifolds below. Then, there exists  
 16 an  $(L-1)$ -dimensional evolutionary-branching manifold passing through  $\mathbf{s}_0$  and  
 17 vertical to the sensitive direction in the space, denoted by  $x$ . In this case, any MLIP  
 18 starting from a monomorphic resident  $\mathbf{s}_a$  with  $|\mathbf{s}_a - \mathbf{s}_0| = O(\sigma)$  monotonically  
 19 converges to the manifold and brings about **a** protected dimorphism, which leads to  
 20 dimorphic divergence in  $x$  without collapse, as long as **their** the deviations from  $\mathbf{s}_0$   
 21 are  $O(\sigma)$ .

22 **Conditions for evolutionary-branching manifolds along MLIPs in normalized**  
 23 **multivariate trait spaces:** Trait axes can be adjusted by rotation such that the first and  
 24 second derivatives of the invasion-fitness function at  $\mathbf{s}_0$  satisfy Eqs. (21) with  $\hat{L} = 1$ ,  
 25 and then the simplified invasion-fitness function, Eqs. (22), satisfies Eqs. (6a), (6b),  
 26 and (18).

1 Even if the sensitive subspace is more than univariate, conditions for evolutionary  
 2 branching might be constructed in a similar form, as explained in Appendix O.

### 3 **7.3 Finding evolutionary-branching lines without prior normalization**

4 For checking conditions for evolutionary-branching lines (or manifolds) in an arbitrary trait  
 5 space  $\mathbf{S}$  with arbitrary dimension  $L$ , the vector  $\mathbf{G}$  and the matrices  $\mathbf{C}$  and  $\mathbf{D}$  of the  
 6 invasion-fitness function  $F(\mathbf{S}';\mathbf{S})$  are all that is needed. These are given by

$$\begin{aligned}
 \mathbf{G} &= \mathbf{F}_m \mathbf{Q}^T \mathbf{B}, \\
 \mathbf{C} &= \mathbf{B}^T \mathbf{Q} (\mathbf{F}_{mm} + \mathbf{F}_{rm}) \mathbf{Q}^T \mathbf{B}, \\
 \mathbf{D} &= \mathbf{B}^T \mathbf{Q} \mathbf{F}_{mm} \mathbf{Q}^T \mathbf{B},
 \end{aligned}
 \tag{23}$$

8 where  $\mathbf{F}_m$  is the gradient vector of  $F(\mathbf{S}';\mathbf{S})$  (i.e., comprising first derivatives) with respect  
 9 to  $\mathbf{S}'$  at  $\mathbf{S}_0$ , while  $\mathbf{F}_{mm}$ ,  $\mathbf{F}_{rr}$ , and  $\mathbf{F}_{rm}$  are the Hessian matrices (i.e., comprising second  
 10 derivatives) there, where the subscripts  $m$  and  $r$  correspond to the mutant  $\mathbf{S}'$  and the  
 11 resident  $\mathbf{S}$ , respectively. The matrix  $\mathbf{Q}$ , which describes the normalization of the trait space  
 12 to attain isotropic mutation with standard deviation  $\sigma$  holding, fulfills  $\mathbf{\Lambda} = (\sigma \mathbf{Q})^T (\sigma \mathbf{Q})$ ,  
 13 while the matrix  $\mathbf{B}$ , which describes the adjustment of the axes by rotation, is given by

$$\mathbf{B} = (\tilde{\mathbf{v}}_{D1}, \dots, \tilde{\mathbf{v}}_{DL}), \tag{24}$$

15 where  $\tilde{\mathbf{v}}_{D1}, \dots, \tilde{\mathbf{v}}_{DL}$  are the eigenvectors of  $\mathbf{Q} \mathbf{F}_{mm} \mathbf{Q}^T$ , ordered such that the corresponding  
 16 eigenvalues satisfy  $\tilde{\lambda}_{D1} > \tilde{\lambda}_{Dj}$  for all  $j = 2, \dots, L$ . See Appendix P for the derivations of Eqs.  
 17 (23) and (24).

18 Notice that the conditions for evolutionary-branching lines (or manifolds) explained above  
 19 are based on locally approximated invasion-fitness functions. Thus, satisfying those condi-  
 20 tions at  $\mathbf{S}_0$  ensures the existence of an evolutionary-branching line (or manifold) only at the  
 21 local scale around this point. However, it is easily shown that if  $\mathbf{S}_0$  satisfies those conditions,  
 22 ~~some of~~ other points slightly deviated from  $\mathbf{S}_0$  are also expected to satisfy those conditions.  
 23 By connecting these points, evolutionary-branching lines (or manifolds) can be found at the  
 24 global scale (Ito and Dieckmann, 2012).

## 1 **8 Discussion**

2 In this paper, we have analytically obtained the conditions for evolutionary branching when  
3 the invasion-fitness function has significant sensitivity differences among directions in  
4 bivariate trait spaces, by focusing on evolutionary paths, called MLIPs, composed of  
5 invasions each of which has maximum likelihood. The result, called the MLIP condition, is  
6 numerically demonstrated to be a useful indicator for the likelihood of evolutionary branching  
7 in evolutionary paths calculated under relaxed assumptions of stochastic invasions (OSIPs)  
8 and of non-rare mutations (PSIPs). The obtained conditions have been extended to  
9 multivariate trait spaces.

10 The MLIP condition requires stronger disruptive selection than is needed for univariate  
11 branching along OSIPs (Metz et al., 1996; Geritz et al., 1997, 1998). The MLIP condition  
12 remains unchanged in multivariate trait spaces as long as the sensitive subspace is univariate,  
13 because directional evolution in the insensitive subspace can be contracted into a single  
14 dimension. Thus, the MLIP condition generalizes the univariate branching conditions to  
15 situations in which a population slowly evolves by weak directional selection in other traits.  
16 This generalization is important, as real populations feature many evolving traits with a large  
17 variation in evolutionary speeds, with the result that the slow traits are likely to keep evolving  
18 directionally after the fast traits have converged to an evolutionary singularity.

19 One of our main assumptions is that mutational step sizes are sufficiently small—~~mutational~~  
20 ~~step sizes~~, so that the first- and second-order terms of the invasion-fitness functions, ~~i.e.~~,  
21 (quantifying the strengths of directional and stabilizing/disruptive ~~selections~~selection  
22 pressures, respectively, ~~provide~~) capture the dominant selection pressures. ~~In this sense~~On the  
23 one hand, as explained above, mutational step sizes ~~are not necessary to~~must be  
24 ~~infinitesimally small, for approximate prediction. Rather, finite~~ for the MLIP condition to  
25 hold, ~~certain magnitudes of mutational step sizes are required.~~ On the other hand, the MLIP  
26 condition cannot be applied when the higher-order terms of invasion-fitness functions have  
27 ~~certain~~a non-negligible influence. ~~In this case, however, resulting~~, which implies selection  
28 pressures ~~become~~that are more complex than combinations of directional and



1 stabilizing/disruptive ~~selections~~selection. Therefore, as long as we try to understand  
2 ~~selections~~selection pressures as combinations of directional and stabilizing/disruptive  
3 ~~selections~~selection, our assumption of small mutational steps ~~seems a good one~~is appropriate.

4 The conditions for evolutionary-branching lines, which are a combination of the condition  
5 for significant sensitivity difference, the condition for convergence stability, and the MLIP  
6 condition, can be used to examine the likelihood of evolutionary branching that could not be  
7 treated by previous branching conditions requiring convergence-stable singular points  
8 (Ackermann and Doebeli, 2004; Ito and Shimada, 2007). For example, Leimar (2005) and Ito  
9 et al. (2009) have numerically shown that evolutionary branching occurs in bivariate trait  
10 spaces ~~which~~that do not contain any evolutionarily singular points that are convergence sta-  
11 ble. In these cases, there exists instead an evolutionarily singular point that is convergence  
12 stable only in one direction, but unstable in the other direction. By applying ~~our~~the conditions  
13 ~~we have presented here~~, evolutionary-branching lines can be identified in the trait spaces of  
14 those models (Ito and Dieckmann, 2012). In such applications, the ~~condition for significant~~  
15 ~~sensitivity- difference condition~~ might be relaxed further, or ~~be omitted~~ altogether, because  
16 this condition ~~partly~~ overlaps with the MLIP condition, ~~and~~; the non-overlapping parts of the  
17 sensitivity-difference condition may be required only for enabling ~~the~~ analytical derivation of  
18 the MLIP condition. In this sense, the MLIP condition may still ~~work~~work well even when  
19 the sensitivity-difference condition does not hold.

20 As the MLIP condition ~~tells~~describes how weak directional selection ~~should~~needs to be in  
21 comparison with disruptive selection ~~for evolutionary branching to occur~~, this information  
22 ~~can~~may also be useful for predicting evolutionary branching ~~induced by~~in the vicinity of evo-  
23 lutionary-branching points. ~~That is~~The MLIP condition then describes how close to an evolu-  
24 tionary-branching point a monomorphic population has to come, for ~~occurrence of~~ evolution-  
25 ary branching. ~~With~~ to occur when mutational steps are finite. Based on a heuristic modifica-  
26 tion of ~~the~~ conditions for evolutionary-branching lines, ~~the~~ areas ~~with~~possessing high likeli-  
27 hoods of evolutionary branching can ~~thus~~ be identified around evolutionary-branching points,  
28 ~~i.e.~~. The resultant evolutionary-branching areas (Ito and Dieckmann, 2012). ~~These areas~~ are  
29 important ~~not the least~~ because, in reality, invasion-fitness functions are ~~always~~ changing at

1 least slowly, due to environmental changes or ~~due to the~~ evolution of other species, ~~imply-~~  
2 ~~ing~~ implying slow shifts of ~~these~~ the locations of evolutionary branching points in trait spaces.  
3 Such shifts may prevent monomorphic ~~populations' sufficient convergence~~ populations from  
4 ~~sufficiently converging~~ to the points required for the emergence of dimorphism, or they may  
5 destroy ~~the initial dimorphism~~ dimorphisms even after those have emerged (Metz et al., 1996;  
6 Metz, 2011). In such cases, by examining whether environmental changes are sufficiently  
7 slow ~~such that~~ for monomorphic populations ~~can get inside of~~ to enter the evolutionary-  
8 branching areas, likelihoods of evolutionary branching may be estimated ~~also under such~~  
9 ~~more realistic circumstances~~.

10 A focus on MLIPs, treated as typical and deterministic paths among corresponding OSIPs,  
11 has enabled our analytical treatment of evolutionary branching in bivariate trait spaces. This  
12 analysis of MLIPs is adding ~~to the evolutionary literature~~ a second deterministic description  
13 of mutation-limited evolutionary dynamics. The more common alternative is the mean ~~evolu-~~  
14 ~~tionary~~ path defined by the canonical equation of ~~adaptive dynamics theory~~ (Dieckmann and  
15 Law, 1996). Roughly speaking, ~~such~~ a mean ~~evolutionary~~ path is formed by mutant invasions,  
16 each of which occurs by the mean mutant phenotype among all mutants ~~that are~~ able to in-  
17 vade, weighted ~~by~~ according to their invasion probabilities. It is therefore interesting to con-  
18 sider how these two deterministic descriptions of ~~mutation-limited evolutionary dynamics~~ are  
19 related. ~~An~~ In particular, an MLIP is identical to the corresponding mean ~~evolutionary~~ path  
20 given by the canonical equation, if directional evolution of a single population with a multi-  
21 variate Gaussian mutation distribution is considered, although the speed ~~of~~ along the MLIPs is  
22 just  $\sqrt{2/\pi} \approx 0.798$  times as fast as ~~along~~ the corresponding mean paths (see Appendix Q).

23 In general, however, MLIPs and mean ~~evolutionary~~ paths are different, because an MLIP is  
24 formed by mutants that are the modes of the invasion-event probability distribution at each  
25 invasion event, while a mean path is formed by mutants that are the means of this distribution.  
26 Thus, differences between the two descriptions can arise, especially when the mutation distri-  
27 bution is discrete, as ~~in~~, e.g., for the univariate fixed-step mutation distribution. As MLIPs are  
28 affected only by the global maximum of an invasion-event probability distribution, but not by  
29 any other of its features, and ~~also~~ as a ~~distribution's~~ global ~~and local~~ maximum may abruptly

1 change ~~their roles~~its role with a local maximum, the mean evolutionary paths may be deemed  
2 more robust than MLIPs for describing directional evolution. On the other hand, by construc-  
3 tion, the canonical equation is not capable of describing evolutionary branching, while MLIPs  
4 can do so. To our knowledge, MLIPs are the only way of deterministically describing evolu-  
5 tionary dynamics that include evolutionary diversifications, without loss of analytical tracta-  
6 bility. Therefore, MLIPs may be useful for analyzing other evolutionary phenomena in multi-  
7 variate trait spaces.

8 Our ~~analysis conducted with analyses~~ of invasion-event probabilities ~~is~~are related to phy-  
9 logeny reconstruction and ancestral–state reconstruction based on empirical data (Wiens,  
10 2000; Barton et al., 2007; Nunn, 2011). ~~In the~~), and may hint at worthwhile extensions of  
11 ~~such methods~~. The standard methods for ancestral–state reconstruction, ~~first reconstruct~~ phy-  
12 logenetic trees ~~are reconstructed~~ based on DNA sequences, and then ~~reconstruct the~~ ancestral  
13 states of ~~the~~ focal traits ~~are reconstructed~~ based on ~~the~~those trees, ~~with using~~ constraints alter-  
14 ~~natively given by~~ maximum parsimony, maximum likelihood, Bayesian methods, etc. ~~Alt-~~  
15 ~~hough~~ While our MLIPs maximize ~~not path-level~~ likelihoods ~~but their parts~~ (i.e., not at the  
16 level of invasion-~~event~~ sequences, but at the level ~~likelihoods~~), it is possible ~~with~~ of invasion  
17 events, numerical ~~calculation~~ methods could be devised to maximize ~~path-level~~ the likelihoods  
18 of OSIPs. When those ~~paths~~ OSIPs are calculated ~~backward~~ from a given present composition  
19 of residents ~~backward~~ to their common ancestor (e.g., ~~with the~~ using Markov Chain Monte  
20 Carlo methods), the past evolutionary dynamics can be reconstructed as a phylogeny in ~~the~~  
21 trait space. In this case, ~~the~~ phylogeny and ancestral states are reconstructed at once, based on  
22 a given fitness function ~~as a kind of~~ providing prior information. ~~Thus, this~~ This alternative  
23 kind of ~~phylogenetic~~ reconstruction might be useful for some genera or families, if ~~the~~ key  
24 ~~quantitative traits and associated~~ ecological settings are known sufficiently well for such ~~that~~  
25 ~~the~~ knowledge ~~can~~ be translated into fitness functions ~~on trait spaces~~, and if ~~the~~ exogenous  
26 changes of those functions ~~from past~~ are expected to be small. Comparing ~~the~~ results obtained  
27 ~~results by~~ from this ecology-based reconstruction method with ~~those by the~~ results from stand-  
28 ard reconstruction methods might provide new ~~understandings~~ insights.

1 In our ~~method~~ approach, it is important to identify the fast traits ~~(,~~ or fast phenotypic direc-  
2 tions), along which ~~the dominant parts of the~~ pasta considered evolutionary diversifications ~~of~~  
3 ~~the focal group may be explained. There are empirical data~~ diversification unfolds. Empirical  
4 evidence suggests that in some taxonomic groups the directions of observed trait differences  
5 among related populations are positively correlated with the trait directions ~~of greatest~~pos-  
6 sessed the largest additive genetic variance within the populations;. These directions, which  
7 have been called “~~the~~ lines of least resistance” (Schluter, 1996). ~~Thus,~~), closely resemble the  
8 fast traits ~~or directions might be given by the lines of least resistance.~~in our approach. If cor-  
9 responding slow traits ~~that may affect the fast evolutionary dynamics are~~ can also ~~found~~be  
10 identified, our conditions for evolutionary-branching lines ~~or~~and manifolds may be applied to  
11 understand the evolutionary ecology of the underlying diversifications.

## 12 **Acknowledgement**

13 The authors thank Hans Metz, Géza Meszéna, Michael Doebeli, Yoh Iwasa, Takashi Ikegami,  
14 Masakazu Shimada, ~~and~~ two anonymous reviewers, and an editor for valuable comments on  
15 earlier versions of this manuscript. H.I. also thanks David Munro for inventing a tool for nu-  
16 merical analysis and visualization, named Yorick, and distributing it for free-. Figures 4 and 5  
17 in this study were produced with Yorick. H.I. acknowledges support in the form of a Research  
18 Fellowship for Young Scientists by the Japan Society for the Promotion of Science (JSPS),  
19 and by IIASA’s Evolution and Ecology Program, facilitating his participation in the Young  
20 Scientists Summer Program (YSSP) of the International Institute for Applied Systems Analy-  
21 sis (IIASA). U.D. gratefully acknowledges financial support by the European Science Foun-  
22 dation, the Austrian Science Fund, the Austrian Ministry of Science and Research, and the  
23 Vienna Science and Technology Fund, as well as by the European Commission, through the  
24 Marie Curie Research Training Network FishACE and the Specific Targeted Research Project  
25 FinE.

## 1 **References**

- 2 Ackermann M, Doebeli M (2004) Evolution of niche width and adaptive diversification.  
3 Evolution 58: 2599-2612
- 4 Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH (2007) Evolution. Cold Spring  
5 Harbor Laboratory Press, New York
- 6 Bolnick DI, Doebeli M (2003) Sexual dimorphism and adaptive speciation: two sides of the  
7 same ecological coin. Evolution 57: 2433-2449
- 8 Bonsall MB, Jansen VAA, Hassell MP (2004) Life history trade-offs assemble ecological  
9 guilds. Science 306:111-114
- 10 Brännström Å, Johansson J, Loeuille N, Kristensen N, Troost T, HilleRisLambers R,  
11 Dieckmann U (2012) Modeling the ecology and evolution of communities: A review of  
12 past achievements, current efforts, and future promises. Evol Ecol Res 14: 601-625
- 13 Brännström Å, Loeuille N, Loreau M, Dieckmann U (2010) Emergence and maintenance of  
14 biodiversity in an evolutionary food-web model. Theor Ecol 4: 467-478
- 15 Claessen D, Andersson J, Persson L (2008) The effect of population size and recombination  
16 on delayed evolution of polymorphism and speciation in sexual populations. Am Nat 172:  
17 E18-E34
- 18 Christiansen FB (1991) On conditions for evolutionary stability for a continuously varying  
19 character. Am Nat 138:37-50
- 20 Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. Nature  
21 400: 354-357
- 22 Dieckmann U, Law R (1996) The dynamical theory of coevolution: a derivation from  
23 stochastic ecological processes. J Math Biol 34: 579-612
- 24 Dieckmann U, Metz JAJ, Doebeli M, Tautz D (2004) Adaptive speciation. Cambridge  
25 University Press, Cambridge
- 26 Doebeli M (1996) A quantitative genetic model for sympatric speciation. J Evol Biol 9: 893-  
27 909

- 1 Doebeli M, Dieckmann U (2000) Evolutionary branching and sympatric speciation caused by  
2 different types of ecological interactions. *Am Nat* 156: S77-S101
- 3 Doebeli M, Dieckmann U (2003) Speciation along environmental gradients. *Nature* 421: 259-  
4 264
- 5 Doebeli M, Hauert C, Killingback T (2004) The evolutionary origin of cooperators and  
6 defectors. *Science* 306: 859-862
- 7 Durinx M, Metz JAJ, Meszéna G (2008) Adaptive dynamics for physiologically structured  
8 population models. *J Math Biol* 56: 673-742
- 9 Durinx M, Van Dooren TJ (2009) Assortative mate choice and dominance modification:  
10 alternative ways of removing heterozygote disadvantage. *Evolution* 63: 334-352
- 11 Egas M, Sabelis MW, Dieckmann U (2005) Evolution of specialization and ecological  
12 character displacement of herbivores along a gradient of plant quality. *Evolution* 59: 507-  
13 520
- 14 Geritz SAH, Metz JAJ, Kisdi É, Meszéna G (1997) Dynamics of adaptation and evolutionary  
15 branching. *Phys Rev Let* 78: 2024-2027
- 16 Geritz SAH, Kisdi É, Meszéna G, Metz JAJ (1998) Evolutionarily singular strategies and the  
17 adaptive growth and branching of the evolutionary tree. *Evol Ecol* 12: 35-57
- 18 Heinz SK, Mazzucco R, Dieckmann U (2009) Speciation and the evolution of dispersal along  
19 environmental gradients. *Evol Ecol* 23: 53-70
- 20 Hendry AP, Kinnison MT (1999) The pace of modern life: measuring rates of contemporary  
21 microevolution. *Evolution* 53: 1637-1653
- 22 Ito HC, Dieckmann U (2007) A new mechanism for recurrent adaptive radiations. *Am Nat*  
23 170: E96-E111
- 24 Ito HC, Shimada M (2007) Niche expansion: coupled evolutionary branching of niche  
25 position and width. *Evol Ecol Res* 9: 675-695
- 26 Ito HC, Shimada M, Ikegami T (2009) Coevolutionary dynamics of adaptive radiation for  
27 food-web development. *Popul Ecol* 51: 65-81
- 28 Ito HC, Dieckmann U (2012) Evolutionary branching lines and areas in bivariate trait spaces.  
29 *Evol Ecol Res* 14: 555–582

- 1 Jansen VAA, Mulder GSEE (1999) Evolving biodiversity. *Ecol Let* 2: 379-386
- 2 Johansson J, Dieckmann U (2009) Evolutionary responses of communities to extinctions.  
3 *Evol Ecol Res* 11: 561-588
- 4 Kinnison MT, Hendry AP (2001) The pace of modern life II: from rates of contemporary  
5 microevolution to pattern and process. *Genetica* 112-113: 145-164
- 6 Kisdi É, Geritz SAH (1999) Adaptive dynamics in allele space: evolution of genetic  
7 polymorphism by small mutations in a heterogeneous environment. *Evolution* 53: 993-  
8 1008
- 9 Lande R (1979) Quantitative genetic analysis of multivariate evolution, applied to brain: body  
10 size allometry. *Evolution* 33: 402-416
- 11 Leimar O (2005) The evolution of phenotypic polymorphism: randomized strategies versus  
12 evolutionary branching. *Am Nat* 165: 669-681
- 13 Leimar O (2008) Multidimensional convergence stability and the canonical adaptive  
14 dynamics. In: Dieckmann U, Metz JAJ (eds) *Elements of Adaptive Dynamics*. Cambridge  
15 University Press, Cambridge (in press)
- 16 Loeuille N, Loreau M (2005) Evolutionary emergence of size-structured food webs. *PNAS*  
17 102: 5761-5766
- 18 MacArthur R (1972) *Geographical Ecology*. Harper and Row, New York
- 19 Maynard Smith J, Price GR (1973) The logic of animal conflict. *Nature* 246: 15-18
- 20 Metz JAJ, Geritz SAH, Meszéna G, Jacobs FJA, van Heerwaarden JS (1996) Adaptive  
21 dynamics: a geometrical study of the consequences of nearly faithful reproduction. In: van  
22 Strien SJ, Verduyn-Lunel SM (eds) *Stochastic and Spatial Structures of Dynamical*  
23 *Systems*. North Holland, Amsterdam, The Netherlands, pp 83-231
- 24 Metz JAJ, Nisbet RM, Geritz SAH (1992) How should we define 'fitness' for general  
25 ecological scenarios? *Trends Ecol Evol* 7: 198-202
- 26 Metz JAJ (2011) Thoughts on the geometry of meso-evolution: collecting mathematical  
27 elements for a postmodern synthesis. In: Chalub FACC, Rodrigues JF (eds) *The*  
28 *Mathematics of Darwin's Legacy*, Springer Verlag, pp. 193-232

- 1 Payne JL, Mazzucco R, Dieckmann U (2011) The evolution of conditional dispersal and  
2 reproductive isolation along environmental gradients. *J Theor Biol* 273: 147-155
- 3 Price GR (1970) Selection and covariance. *Nature* 227: 520-521
- 4 Prout T (1968) Sufficient conditions for multiple niche polymorphism. *Am Nat* 102: 493-496
- 5 Ravigné V, Dieckmann U, Olivieri I (2009) Live where you thrive: joint evolution of habitat  
6 choice and local adaptation facilitates specialization and promotes diversity. *Am Nat* 174:  
7 E141-E169
- 8 Rice SH, Papadopoulos A, Harting J (2011) Stochastic processes driving directional  
9 evolution. In: Pontarotti P (ed) *Evolutionary Biology – Concepts, Biodiversity,  
10 Macroevolution and Genome Evolution*, Springer, Verlag, pp. 21-33.
- 11 Rundle HD, Nosil R (2004) Ecological speciation. *Ecol Let* 8: 336-352
- 12 Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution* 50:  
13 1766-1774
- 14 Takahashi D, Brännström Å, Mazzucco R, Yamauchi A, Dieckmann U (in press) Abrupt  
15 community transitions and cyclic evolutionary dynamics in complex food webs. *J Theor  
16 Biol*, doi:10.1016/j.jtbi.2013.08.003
- 17 Van Dooren TJM (2006) Protected polymorphism and evolutionary stability in pleiotropic  
18 models with trait-specific dominance. *Evolution* 60: 1991-2003
- 19 Van Dooren TJM, Durinx M, Demon I (2004) Sexual dimorphism or evolutionary branching?  
20 *Evol Ecol Res* 6: 857-871
- 21 Vukics A, Asboth J, Meszéna G (2003). Speciation in multidimensional evolutionary space.  
22 *Phys Rev E* 68: 041903
- 23 Wiens JJ (2000) *Phylogenetic Analysis of Morphological Data*. Smithsonian Institution Press,  
24 Washington
- 25



## 1 **Figure captions**

2 **Figure 1.** Heuristic estimation of the likelihood of evolutionary branching. (a) Illustration of a  
3 fitness landscape around a population directionally evolving in trait  $y$  under disruptive  
4 selection in trait  $x$ . The population's resident phenotype is indicated with a small filled  
5 circle. The strength of disruptive selection in  $x$  corresponds to the curvature of the surface  
6 along  $x$ , denoted by  $D_{xx}$ , while the strength of directional selection in  $y$  corresponds to  
7 the gradient of the surface along  $y$ , denoted by  $G_y$ . (b-g) The small filled circles again  
8 indicate the resident phenotypes. The large circles indicate possible mutants, and their  
9 radiuses show the mutational step sizes. The dotted curves highlight the fitness contours that  
10 are tangential to these circles, with the tangential points (indicated with small white circles)  
11 corresponding to the mutants with maximum fitness.

12 **Figure 2.** Coordinate transformations for normalizing the mutation probability distribution.  
13 The first transformation, (a) to (b), is a rotation, while the others cause scaling.

14 **Figure 3.** Conditions for dimorphic emergence and dimorphic divergence. In panels (a) and  
15 (b), the white and light gray regions indicate positive and negative invasion fitnesses, respec-  
16 tively. The thick gray curves in (a) and (b) indicate zero-countours of the invasion fitnesses  
17 for monomorphism,  $f(\mathbf{s}';\mathbf{s})$ , and for dimorphism,  $f(\mathbf{s}';\mathbf{s}_1,\mathbf{s}_2)$ , respectively, which are para-  
18 bolic curves sharing the same shape. In panels (c) and (d), the white regions indicate mutants  
19 that satisfy the conditions for dimorphic emergence and those for dimorphic divergence, re-  
20 spectively. The thin parabolic curves giving the boundaries share the same shape with zero-  
21 contours of the invasion fitnesses (thick gray curves). In panels (e) and (f), the mutants of  
22 maximum-likelihood invasion are included in the dark gray rectangles. If the MLIP condition  
23  $D > 1/\sqrt{2}$  holds, the dark gray rectangles are included in the white regions that ensure evolu-  
24 tionary branching. The dark gray and white regions touch each other only when  $D = 1/\sqrt{2}$ .  
25 The trait space has been normalized and rescaled so that the standard deviation of mutational  
26 step sizes equals 1 in all directions.

27 **Figure 4.** Occurrence of evolutionary branching along MLIPs. (a) Occurrence of evolutionary  
28 branching when the population has directionally evolved in  $y$  by  $L_y$ , beyond what is

1 implied by the MLIP condition ( $\hat{L}_{y_0}$ ), for  $L_y - \hat{L}_{y_0} = 0, 100, 200$ . Results are shown for a  
 2 bivariate Gaussian mutation distribution for combinations of  $C(<0)$  and  $D$ . The ~~black-~~  
 3 ~~dashed~~ line indicates the threshold for  $D$  given by the MLIP condition,  $D = 1/\sqrt{2}$ . (b-d)  
 4 MLIPs (gray curves) and OSIPs (black curves) for different combinations of  $C$  and  $D$ ,  
 5 shown in panel (a):  $D = 1.0, 0.1, 0.05$  for (b), (c), (d), respectively, and  $C = -0.1$ . The trait  
 6 space has been normalized and rescaled so that the standard deviation of mutational step sizes  
 7 equals 1 in all directions.

8 **Figure 5.** Occurrence of evolutionary branching along OSIPs and PSIPs. Contour lines show  
 9 the combinations of  $C(<0)$  and  $D$  at which the cumulative probability of evolutionary  
 10 branching ~~occurs with a probability of more than~~ reaches 97% ~~for~~ (a), 90% ~~for~~ (b-d), and or  
 11 95% ~~for~~ (e-h), when the population has directionally evolved in  $y$  by  $L_y$ , beyond what is  
 12 implied by the MLIP condition ( $\hat{L}_{y_0}$ ), for  $L_y - \hat{L}_{y_0} = 0, 100, 200$  along OSIPs (a-d), and for  
 13  $L_y - \hat{L}_{y_0} = 0, 40, 80$  along PSIPs (e-h). Results are shown for mutation distributions that are  
 14 bivariate Gaussian (a,e), bivariate fixed-step (b,f), univariate Gaussian (c,g), and univariate  
 15 fixed-step (d,h). The ~~white-dashed~~ lines indicate the threshold for  $D$  given by the MLIP  
 16 condition,  $D = 1/\sqrt{2}$ . The trait spaces ~~h~~ have been normalized and rescaled so that the  
 17 standard deviation of mutational step sizes equals 1 in all directions. ~~Model~~ Parameters:  
 18  $b_1 = 1.0$ ,  $K_0 = 300$  (e-h),  $\mu = 5.1 \times 10^{-3}$  (e-f),  $\mu_x = 0.95$ , and  $\mu_y = 0.05$  (g,h).

# Evolutionary branching under slow directional evolution (Appendices)

Hiroshi C. Ito and Ulf Dieckmann

These appendices are structured as follows. Appendices A and B ~~are on~~ contain the derivation of the normal form of invasion-fitness functions with significant sensitivity difference. Appendices C to J ~~are on~~ cover maximum-likelihood-invasion paths (MLIPs) and ~~on~~ conditions for evolutionary branching along those paths under the normal form of invasion-fitness functions. Appendices K to N ~~are on~~ provide methods for the numerical examination of the obtained conditions for evolutionary-branching lines. Appendices O and P extend and apply these conditions for evolutionary-branching lines. Appendix Q analyzes directional evolution along MLIPs in comparison with another deterministic description of directional evolution, the canonical equation of adaptive dynamics theory (Dieckmann and Law, 1996).

## Appendix A: Derivation of quadratic form of invasion-fitness functions

Here we derive an approximate quadratic form of  $f(\mathbf{s}'; \mathbf{s})$ , Eqs. (2) in Section 3. For convenience, ~~and~~ without any loss of generality, we shift the origin close to the resident phenotype, so that  $|\mathbf{s}| = O(\sigma)$  and  $|\mathbf{s}'| = O(\sigma)$ . We then expand  $f(\mathbf{s}'; \mathbf{s})$  ~~around~~ about this origin  $\mathbf{s}_0 = (0, 0)^T$  as

$$f(\mathbf{s}'; \mathbf{s}) = \mathbf{f}_m \mathbf{s}' + \mathbf{f}_r \mathbf{s} + \frac{1}{2} \left[ \mathbf{s}'^T \mathbf{f}_{mm} \mathbf{s}' + \mathbf{s}^T \mathbf{f}_{rr} \mathbf{s} + \mathbf{s}^T \mathbf{f}_{rm} \mathbf{s}' + \mathbf{s}'^T \mathbf{f}_{mr} \mathbf{s} \right] + O(\sigma^3), \quad (\text{A.1a})$$

with

$$\begin{aligned} \mathbf{f}_m &= \begin{pmatrix} f_{x'} & f_{y'} \end{pmatrix}_{\mathbf{s}'=\mathbf{s}_0}, & \mathbf{f}_r &= \begin{pmatrix} f_x & f_y \end{pmatrix}_{\mathbf{s}=\mathbf{s}_0}, \\ \mathbf{f}_{mm} &= \begin{pmatrix} f_{x'x'} & f_{x'y'} \\ f_{x'y'} & f_{y'y'} \end{pmatrix}_{\mathbf{s}'=\mathbf{s}_0}, & \mathbf{f}_{rr} &= \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}_{\mathbf{s}=\mathbf{s}_0}, \\ \mathbf{f}_{rm} &= \begin{pmatrix} f_{xx'} & f_{xy'} \\ f_{yx'} & f_{yy'} \end{pmatrix}_{\mathbf{s}'=\mathbf{s}_0}, & \mathbf{f}_{mr} &= \begin{pmatrix} f_{x'x} & f_{x'y} \\ f_{y'x} & f_{y'y} \end{pmatrix}_{\mathbf{s}'=\mathbf{s}_0} = \mathbf{f}_{rm}^T, \end{aligned} \quad (\text{A.1b})$$

1 where the subscripts ‘m’ and ‘r’ refer to mutants and residents, respectively, and where  
 2  $f_\alpha = \partial f(\mathbf{s}'; \mathbf{s}) / \partial \alpha$  for  $\alpha = x', y', x, y$  and  $f_{\alpha\beta} = \partial^2 f(\mathbf{s}'; \mathbf{s}) / \partial \alpha \partial \beta$  for  $\alpha, \beta = x', y', x, y$   
 3 denote the first and second derivatives of  $f(\mathbf{s}'; \mathbf{s})$ , respectively. We transform Eq. (A.1a) into

$$\begin{aligned}
 f(\mathbf{s}'; \mathbf{s}) &= \mathbf{f}_m(\mathbf{s}' - \mathbf{s}) \\
 &+ \frac{1}{2}(\mathbf{s}' - \mathbf{s})^T \mathbf{f}_{mm}(\mathbf{s}' - \mathbf{s}) \\
 &+ \mathbf{s}^T [\mathbf{f}_{mm} + \mathbf{f}_{rm}] (\mathbf{s}' - \mathbf{s}) \\
 &+ (\mathbf{f}_m + \mathbf{f}_r) \mathbf{s} + \frac{1}{2} \mathbf{s}^T [\mathbf{f}_{mm} + \mathbf{f}_{rr} + \mathbf{f}_{rm} + \mathbf{f}_{mr}] \mathbf{s} + O(\sigma^3),
 \end{aligned} \tag{A.2}$$

5 where  $\mathbf{s}'^T \mathbf{f}_{mm} \mathbf{s} = [\mathbf{s}'^T \mathbf{f}_{mm} \mathbf{s}]^T = \mathbf{s}^T \mathbf{f}_{mm} \mathbf{s}'$ ,  $\mathbf{s}'^T \mathbf{f}_{mr} \mathbf{s} = [\mathbf{s}'^T \mathbf{f}_{mr} \mathbf{s}]^T = \mathbf{s}^T \mathbf{f}_{mr}^T \mathbf{s}' = \mathbf{s}^T \mathbf{f}_{rm} \mathbf{s}'$ , and  
 6  $\mathbf{s}^T (\mathbf{f}_{rm} - \mathbf{f}_{mr}) \mathbf{s} = 0$  are used. The attaining of a population dynamical equilibrium implies  
 7  $f(\mathbf{s}; \mathbf{s}) = 0$ . Thus, any  $\mathbf{s}$  has to satisfy

$$8 \quad f(\mathbf{s}; \mathbf{s}) = (\mathbf{f}_m + \mathbf{f}_r) \mathbf{s} + \frac{1}{2} \mathbf{s}^T (\mathbf{f}_{mm} + \mathbf{f}_{rr} + \mathbf{f}_{rm} + \mathbf{f}_{mr}) \mathbf{s} + O(\sigma^3) = 0. \tag{A.3}$$

9 Therefore, all terms in the fourth row of Eq. (A.2) are  $O(\sigma^3)$ , which gives

$$\begin{aligned}
 f(\mathbf{s}'; \mathbf{s}) &= \mathbf{f}_m(\mathbf{s}' - \mathbf{s}) + \mathbf{s}^T [\mathbf{f}_{mm} + \mathbf{f}_{rm}] (\mathbf{s}' - \mathbf{s}) + \frac{1}{2} (\mathbf{s}' - \mathbf{s})^T \mathbf{f}_{mm} (\mathbf{s}' - \mathbf{s}) + O(\sigma^3) \\
 &= \mathbf{G}(\mathbf{s}' - \mathbf{s}) + \mathbf{s}^T \mathbf{C}(\mathbf{s}' - \mathbf{s}) + \frac{1}{2} (\mathbf{s}' - \mathbf{s})^T \mathbf{D}(\mathbf{s}' - \mathbf{s}) + O(\sigma^3),
 \end{aligned} \tag{A.4}$$

11 where the second line uses the notation introduced in Eq. (2b). The substitution  $\delta \mathbf{s} = \mathbf{s}' - \mathbf{s}$   
 12 and considering  $\mathbf{s}_0 \neq (0, 0)^T$  give rise to Eq. (2a) in Section 3. Notice that the quadratic  
 13 approximation in Ito and Dieckmann (2012), where  $\mathbf{C}$  is multiplied by  $1/2$ , becomes  
 14 identical to Eq. (2a) by defining  $\mathbf{C}$  as  $\mathbf{C} = \mathbf{f}_{mm} + \mathbf{f}_{rm}$  (the convention used here) instead of  
 15 as  $\mathbf{C} = \frac{1}{2}(\mathbf{f}_{mm} + \mathbf{f}_{rm})$  (the convention used there).

## 16 **Appendix B: Condition for significant sensitivity difference**

17 Here we derive the condition for significant sensitivity difference of normalized invasion-  
 18 fitness functions, Eq. (4) in Section 3. First, we show how sensitivity difference can be caused  
 19 by the asymmetry of mutational step sizes in the original trait space. Second, we extend this  
 20 relationship into a general condition for significant sensitivity difference.

## 1 Sensitivity difference due to mutational asymmetry

2 We assume that the  $X$  - and  $Y$  -axes of the original trait space  $\mathbf{S}$  have been aligned as  
 3 shown in Fig. 2b, so that  $V_{XY} = 0$ . In this space, the invasion-fitness function  $F(\mathbf{S}'; \mathbf{S})$  is  
 4 expanded similarly to Eqs. (2) as

$$5 \quad F(\mathbf{S}'; \mathbf{S}) = \check{\mathbf{G}} \delta \mathbf{S} + (\mathbf{S} - \mathbf{S}_0)^T \check{\mathbf{C}} \delta \mathbf{S} + \frac{1}{2} \delta \mathbf{S}^T \check{\mathbf{D}} \delta \mathbf{S} + O(\sigma_X^3), \quad (\text{B.1a})$$

6 where

$$7 \quad \check{\mathbf{G}} = \begin{pmatrix} \check{G}_X & \check{G}_Y \end{pmatrix}, \quad \check{\mathbf{C}} = \begin{pmatrix} \check{C}_{XX} & \check{C}_{XY} \\ \check{C}_{YX} & \check{C}_{YY} \end{pmatrix}, \quad \check{\mathbf{D}} = \begin{pmatrix} \check{D}_{XX} & \check{D}_{XY} \\ \check{D}_{XY} & \check{D}_{YY} \end{pmatrix}. \quad (\text{B.1b})$$

8 We now consider the case that  $\sigma_Y$  is much smaller than  $\sigma_X$ , such that  $\sigma_Y = O(\sigma_X^2)$ . We  
 9 introduce a new coordinate system  $\mathbf{s} = (x, y)^T$ , where  $x = X$  and  $y = (\sigma_X / \sigma_Y) Y$ , which  
 10 results in isotropic mutation with standard deviation  $\sigma = \sigma_X$  (Fig. 2c). Substituting  $X = x$   
 11 and  $Y = (\sigma_Y / \sigma_X) y$  into Eq. (B.1a) yields the following normalized invasion-fitness  
 12 function,

$$13 \quad \begin{aligned} f(\mathbf{s}'; \mathbf{s}) &= F((x', (\sigma_Y / \sigma_X) y')^T; (x, (\sigma_Y / \sigma_X) y)^T) \\ &= G_x \delta x + C_{xx} (x - x_0) \delta x + \frac{1}{2} D_{xx} \delta x^2 + G_y \delta y \\ &\quad + C_{xy} (x - x_0) \delta y + C_{yx} (y - y_0) \delta x + C_{yy} (y - y_0) \delta y \\ &\quad + D_{xy} \delta x \delta y + \frac{1}{2} D_{yy} \delta y^2 + O(\sigma^3), \end{aligned} \quad (\text{B.2a})$$

14 where

$$15 \quad \begin{aligned} G_x &= \check{G}_X, \quad C_{xx} = \check{C}_{XX}, \quad D_{xx} = \check{D}_{XX}, \quad x - x_0 = O(\sigma), \quad y - y_0 = O(\sigma), \\ G_y &= (\sigma_Y / \sigma_X) \check{G}_Y = O(\sigma), \quad C_{xy} = (\sigma_Y / \sigma_X) \check{C}_{XY} = O(\sigma), \\ C_{yx} &= (\sigma_Y / \sigma_X) \check{C}_{YX} = O(\sigma), \quad C_{yy} = (\sigma_Y / \sigma_X)^2 \check{C}_{YY} = O(\sigma^2), \\ D_{xy} &= (\sigma_Y / \sigma_X) \check{D}_{XY} = O(\sigma), \quad D_{yy} = (\sigma_Y / \sigma_X)^2 \check{D}_{YY} = O(\sigma^2). \end{aligned} \quad (\text{B.2b})$$

16 Thus, by including all applicable terms in  $O(\sigma^3)$ , we see that Eq. (B.2a) yields the normal  
 17 form of invasion-fitness functions with significant sensitivity difference, Eq. (3) in Section 3.

## 1 Generalization of sensitivity difference

2 The normal form in Eq. (3) can also be obtained when the sensitivity of a fitness function to  
 3 variation in trait  $y$  is weak, so that  $G_y$ ,  $D_{xy}$ ,  $D_{yy}$ ,  $C_{xy}$ ,  $C_{yx}$ , and  $C_{yy}$  are all relatively  
 4 small. To make this notion precise, we proceed as follows. We suppose an arbitrary invasion-  
 5 fitness function defined in a normalized trait space  $\mathbf{s} = (x, y)^T$ , in which mutation is isotropic  
 6 with standard deviation  $\sigma \ll 1$ . This function is then given by Eq. (2),

$$7 \quad f(\mathbf{s}'; \mathbf{s}) = \mathbf{G}\delta\mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C}\delta\mathbf{s} + \frac{1}{2} \delta\mathbf{s}^T \mathbf{D}\delta\mathbf{s} + O(\sigma^3). \quad (\text{B.3})$$

8 Here we assume that the value of invasion fitness is scaled such that the sensitivity of the  
 9 function to trait  $x$  is not small, i.e.,  $|G_x| + |C_{xx}| + |D_{xx}|$  is of order  $\sigma^0$ . On that basis, we can  
 10 define a significant sensitivity difference as follows.

### 11 Definition of significant sensitivity difference of invasion-fitness functions:

12 Suppose that for a normalized invasion-fitness function the  $x$ - and  $y$ -axes can be ad-  
 13 justed by coordinate rotation, such that the function can be decomposed into a function  
 14 depending only on  $x'$  and  $x$ , and into a residual of sufficiently small magnitude  
 15  $\varepsilon = O(\sigma)$ ,

$$16 \quad f(\mathbf{s}'; \mathbf{s}) = g(x'; x) + \varepsilon h(\mathbf{s}'; \mathbf{s}), \quad (\text{B.4})$$

17 where  $h(\mathbf{s}'; \mathbf{s}) = [f(\mathbf{s}'; \mathbf{s}) - g(x'; x)] / \varepsilon$  is kept smooth and finite, i.e., its first and sec-  
 18 ond derivatives are  $O(\varepsilon^0)$ . Then, it is said that the function  $f$  has significant sensi-  
 19 tivity difference with respect to the  $x$ - and  $y$ -directions.

20 If  $f$  in Eq. (B.3) has significant sensitivity difference, then substituting Eq. (B.3) into Eq.  
 21 (B.4) and assuming  $\varepsilon = \sigma$  yields

$$22 \quad \begin{aligned} g(x'; x) &= G_x \delta x + C_{xx} (x - x_0) \delta x + \frac{1}{2} D_{xx} \delta x^2 + O(\sigma^3), \\ h(\mathbf{s}'; \mathbf{s}) &= \frac{G_y}{\sigma} \delta y \\ &\quad + \frac{C_{xy}}{\sigma} (x - x_0) \delta y + \frac{C_{yx}}{\sigma} (y - y_0) \delta x + \frac{C_{yy}}{\sigma} (y - y_0) \delta y \\ &\quad + \frac{D_{yy}}{2\sigma} \delta y^2 + \frac{D_{xy}}{\sigma} \delta x \delta y + O(\sigma^3). \end{aligned} \quad (\text{B.5})$$

1 To keep the first and second derivatives of  $h(\mathbf{s}';\mathbf{s})$  at  $O(\sigma^0)$ ,  $G_y$ ,  $C_{xy}$ ,  $C_{yy}$ ,  $D_{xy}$ , and  
 2  $D_{yy}$  must all be  $O(\sigma)$ , in which case Eq. (2a) in Section 3 can be transformed into Eq. (3).  
 3 Thus, a sufficient condition for the normalized invasion-fitness function in Eq. (2a) to be  
 4 transformed into the normal form in Eq. (3) is expressed as Eq. (4),

$$5 \frac{|G_y| + |C_{xy}| + |C_{yx}| + |C_{yy}| + |D_{xy}| + |D_{yy}|}{|G_x| + |C_{xx}| + |D_{xx}|} = O(\sigma), \quad (\text{B.6})$$

6 where the denominator enables application to normalized invasion-fitness functions that are  
 7 not yet suitably scaled.

## 8 **Appendix C: Approximation of invasion-event rate density**

9 Here we explain how the invasion-event rate density is approximated in Eq. (8a) in Section 4.  
 10 The first row of Eq. (8a) is transformed into

$$11 \begin{aligned} E(\mathbf{s}';\mathbf{s}) &= \mu \hat{n} b(\mathbf{s};\mathbf{s}) M(\mathbf{s}' - \mathbf{s}) \frac{f(\mathbf{s}';\mathbf{s})_+}{b(\mathbf{s}';\mathbf{s})} \\ &= \mu \hat{n} M(\mathbf{s}' - \mathbf{s}) f(\mathbf{s}';\mathbf{s})_+ (1 - \phi(\mathbf{s}';\mathbf{s})), \end{aligned} \quad (\text{C.1a})$$

12 where

$$13 \phi(\mathbf{s}';\mathbf{s}) = \frac{b(\mathbf{s}';\mathbf{s}) - b(\mathbf{s};\mathbf{s})}{b(\mathbf{s}';\mathbf{s})} = O(\sigma), \quad (\text{C.1b})$$

14 because  $b(\mathbf{s}';\mathbf{s})$  converges to  $b(\mathbf{s};\mathbf{s})$  as  $\sigma \rightarrow 0$ , which gives  $\mathbf{s}' \rightarrow \mathbf{s}$ . Thus, the  
 15 approximation in Eq. (8a) applies in the leading order of  $\mathbf{s}' - \mathbf{s}$ , as long as  $\sigma$  is sufficiently  
 16 small such that  $b(\mathbf{s}';\mathbf{s}) - b(\mathbf{s};\mathbf{s})$  is much smaller than  $b(\mathbf{s}';\mathbf{s})$ , i.e.,  
 17  $[b(\mathbf{s}';\mathbf{s}) - b(\mathbf{s};\mathbf{s})] / b(\mathbf{s}';\mathbf{s}) = O(\sigma)$ .

## 18 **Appendix D: MLIP condition**

19 Here we prove that the conditions for dimorphic emergence and those for dimorphic  
 20 divergence are both satisfied if and only if the MLIP condition  $D > 1/\sqrt{2}$  holds, provided  
 21 that  $C < 0$ .

## 1 Main proof

2 When the resident is located on the convergence-stable line, MLI mutants  $\mathbf{s}'_{\text{MLI}}$ , given by Eq.  
3 (17) in Section 5, do not satisfy the second condition for dimorphic emergence, inequality  
4 (13b), for  $D \leq 1/\sqrt{2}$ . Thus,  $D > 1/\sqrt{2}$  is necessary for dimorphic emergence. Therefore,  
5 we only have to examine whether  $D > 1/\sqrt{2}$  satisfies the conditions for dimorphic  
6 emergence ~~as well as those for~~ and dimorphic divergence, as follows.

7 First, we examine the conditions for dimorphic emergence, i.e., inequalities (13). For  
8 convenience, instead of the MLI mutant  $\mathbf{s}'_{\text{MLI}}$  itself, we use its deviation from its parental  
9 resident,  $\delta\mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}$ . For this  $\delta\mathbf{s}_{\text{MLI}}$ , the following lemma holds (see the next subsection  
10 for the proof).

### 11 Lemma D.1:

12 If the MLIP condition  $D > 1/\sqrt{2}$  holds, then, for any monomorphic resident  $\mathbf{s}$ , the  
13 MLI ~~mutation~~  $\delta\mathbf{s}_{\text{MLI}}$  satisfies

$$14 \quad \begin{cases} \delta x_{\text{MLI}} > 0 & \text{for } x < 0 \\ \delta x_{\text{MLI}} < 0 & \text{for } x > 0, \end{cases} \quad (\text{D.1a})$$

$$15 \quad 0 < \delta y_{\text{MLI}} < D\delta x_{\text{MLI}}^2, \quad (\text{D.1b})$$

$$16 \quad 1 < |\delta x_{\text{MLI}}| < \sqrt{2}. \quad (\text{D.1c})$$

17 Inequality (D.1a) is identical to inequality (13a). Inequality (D.1b) is a sufficient condition for  
18 inequality (13b). Thus, Lemma D.1 ensures that conditions for dimorphic emergence hold  
19 under the MLIP condition.

20 Next, we examine the conditions for dimorphic divergence, i.e., inequalities (14). Under  
21 dimorphism of  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , the MLI mutation is ~~written as~~ given by  $\delta\mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}}$  with  
22  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_1$  or  $\mathbf{s}_2$ . For this  $\delta\mathbf{s}_{\text{MLI}}$ , the following lemma holds (see the last subsection for the  
23 proof).



1 **Lemma D.2:**

2 If the MLIP condition  $D > 1/\sqrt{2}$  holds, then, for any dimorphic residents  $\mathbf{s}_1$  and  $\mathbf{s}_2$   
 3 satisfying  $|y_2 - y_1| < D(x_2 - x_1)^2$ , the MLI mutation  $\delta\mathbf{s}_{\text{MLI}}$  satisfies

$$4 \quad \begin{cases} \delta x_{\text{MLI}} > 0 & \text{for } \tilde{x} < 0 \\ \delta x_{\text{MLI}} < 0 & \text{for } \tilde{x} > 0, \end{cases} \quad (\text{D.2a})$$

$$5 \quad 0 < \delta y_{\text{MLI}} < D\delta x_{\text{MLI}}^2, \quad (\text{D.2b})$$

$$6 \quad 1 < |\delta x_{\text{MLI}}| < \sqrt{2}, \quad (\text{D.2c})$$

7 where  $\tilde{\mathbf{s}} = (\tilde{x}, \tilde{y})^T = ([D(2x_{\text{MLI}} - x_2 - x_1) - (y_2 - y_1)] / (x_2 - x_1) / C, 0)^T$ .

8 Below we check whether Lemma D.2 ensures that the conditions for dimorphic divergence,  
 9 inequalities (14), hold under  $D > 1/\sqrt{2}$ . As in inequalities (14),  $x_1 < x_2$  is assumed without  
 10 loss of generality. First, we suppose  $\mathbf{s}'_{\text{MLI}}$  originates from  $\mathbf{s}_1$  (i.e.,  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_1$  and  
 11  $\delta\mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_1$ ), in which case inequality (14a) ~~is expected to hold~~, i.e.,  
 12  $x'_{\text{MLI}} < x_1$  and  $|y'_{\text{MLI}} - y_1| < D(x'_{\text{MLI}} - x_1)^2$ , ~~hold as explained below~~. Clearly, the second  
 13 inequality,  $|y'_{\text{MLI}} - y_1| < D(x'_{\text{MLI}} - x_1)^2$ , is satisfied by inequality (D.2b). The first inequality,  
 14  $x'_{\text{MLI}} < x_1$ , also holds under inequalities (D.2a), because  $\delta x_{\text{MLI}} = x'_{\text{MLI}} - x_1 < 0$  follows from  
 15  $\tilde{x} > 0$ :

$$16 \quad \begin{aligned} \tilde{x} &= \frac{1}{C} \left[ D(x_1 - x_2) - \frac{y_2 - y_1}{x_2 - x_1} \right] \\ &= \frac{1}{C(x_1 - x_2)} \left[ D(x_1 - x_2)^2 + (y_2 - y_1) \right] \\ &\geq \frac{1}{C(x_1 - x_2)} \left[ D(x_1 - x_2)^2 - |y_2 - y_1| \right] \\ &> \frac{1}{C(x_1 - x_2)} \left[ D(x_1 - x_2)^2 - D(x_2 - x_1)^2 \right] = 0, \end{aligned} \quad (\text{D.3})$$

17 where  $C(x_1 - x_2)$  is positive because  $C < 0$  and  $x_1 - x_2 < 0$ .

18 Second, we suppose  $\mathbf{s}'_{\text{MLI}}$  originates from  $\mathbf{s}_2$  (i.e.,  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_2$  and  
 19  $\delta\mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_2$ ), in which case inequality (14b) ~~is expected to hold~~, i.e.,  
 20  $x'_{\text{MLI}} > x_2$  and  $|y'_{\text{MLI}} - y_2| < D(x'_{\text{MLI}} - x_2)^2$ , ~~again hold as explained below~~. Clearly, the second

1 inequality,  $|y'_{\text{MLI}} - y_2| < D(x'_{\text{MLI}} - x_2)^2$ , is satisfied by inequality (D.2b). The first inequality,  
 2  $x'_{\text{MLI}} > x_2$ , also holds under inequalities (D.2a), because  $\delta x_{\text{MLI}} = x'_{\text{MLI}} - x_2 > 0$  follows from  
 3  $\tilde{x} < 0$ :

$$\begin{aligned}
 \tilde{x} &= \frac{1}{C} \left[ D(x_2 - x_1) - \frac{y_2 - y_1}{x_2 - x_1} \right] \\
 &= \frac{1}{C(x_1 - x_2)} \left[ -D(x_2 - x_1)^2 + (y_2 - y_1) \right] \\
 &\leq \frac{1}{C(x_1 - x_2)} \left[ -D(x_1 - x_2)^2 + |y_2 - y_1| \right] \\
 &< \frac{1}{C(x_1 - x_2)} \left[ -D(x_1 - x_2)^2 + D(x_2 - x_1)^2 \right] = 0.
 \end{aligned}
 \tag{D.4}$$

5 Thus, the conditions for dimorphic divergence, inequalities (14), as well as those for  
 6 dimorphic emergence, inequalities (13), hold under  $D > 1/\sqrt{2}$ . Therefore, conditions for  
 7 dimorphic emergence and those for dimorphic divergence both hold if and only if the MLIP  
 8 condition  $D > 1/\sqrt{2}$  holds. This completes the proof.

## 9 Proof of Lemma D.1

10 Among the three inequalities (D.1) of Lemma D.1, we first analyze inequalities (D.1a). MLI  
 11 mutants, which maximize  $P(\mathbf{s}'; \mathbf{s})$  in Eq. (8b), always have positive invasion fitnesses. Thus,  
 12 the subscript “+” is ~~needed~~ not needed here. For arbitrary  $\mathbf{s}'$  and  $\mathbf{s}$ ,  $P(\mathbf{s}'; \mathbf{s})$  satisfies

$$13 \quad P(\delta \mathbf{s} + \mathbf{s}; \mathbf{s}) - P(\delta \mathbf{s}_{x_-} + \mathbf{s}; \mathbf{s}) = \frac{T\mu\hat{n}}{\pi} \exp\left(-\frac{1}{2}[\delta x^2 + \delta y^2]\right) Cx\delta x, \tag{D.5}$$

14 where  $\delta \mathbf{s}_{x_-} = (-\delta x, \delta y)^T$  is identical to  $\delta \mathbf{s} = (\delta x, \delta y)^T$  except that the sign of  $\delta x$  is  
 15 reversed. When  $x < 0$ , Eq. (D.5) is negative for  $\delta x < 0$ , because  $C < 0$ . This means that  
 16 when  $x < 0$ , for every mutational step in the negative  $x$ -direction, there exists a step in the  
 17 positive  $x$ -direction that has a higher probability density. Thus, the global maximum is  
 18 reached for some  $\delta x > 0$ . When  $x > 0$ , on the other hand, Eq. (D.5) is negative for  $\delta x > 0$ ,  
 19 which implies that  $\delta x_{\text{MLI}}$  must be negative for  $x > 0$ . Therefore, inequalities (D.1a) hold.

20 Second, we examine inequalities (D.1b) and (D.1c). We analyze the extremal conditions

$$\frac{\partial P(\delta \mathbf{s} - \mathbf{s}; \mathbf{s})}{\partial \delta x} = -\frac{T\mu\hat{n}}{2\pi} \exp\left(-\frac{1}{2}[\delta x^2 + \delta y^2]\right) \delta x \left[ \delta y + D\delta x^2 + Cx\delta x - \frac{2D\delta x + Cx}{\delta x} \right], \quad (\text{D.6a})$$

$$= 0$$

$$\frac{\partial P(\delta \mathbf{s} - \mathbf{s}; \mathbf{s})}{\partial \delta y} = -\frac{T\mu\hat{n}}{2\pi} \exp\left(-\frac{1}{2}[\delta x^2 + \delta y^2]\right) \delta y \left[ \delta y + D\delta x^2 + Cx\delta x - \frac{1}{\delta y} \right], \quad (\text{D.6b})$$

$$= 0$$

which hold for  $\delta \mathbf{s} = \delta \mathbf{s}_{\text{MLI}}$ . Eq. (D.6a) is transformed into

$$\delta y = -D(\delta x^2 - 2) - \frac{Cx(\delta x^2 - 1)}{\delta x} = g(\delta x), \quad (\text{D.7})$$

from where we define the right hand side as second equation defines  $g(\delta x)$ . Taking the difference of Eqs. (D.6a) and (D.6b),

$$\frac{1}{\delta x} \frac{\partial E(\delta \mathbf{s} - \mathbf{s}; \mathbf{s})}{\partial \delta x} - \frac{1}{\delta y} \frac{\partial E(\delta \mathbf{s} - \mathbf{s}; \mathbf{s})}{\partial \delta y} = 0, \quad (\text{D.8})$$

we find

$$\delta y = \frac{\delta x}{2D\delta x + Cx} = h(\delta x), \quad (\text{D.9})$$

from where we define the right hand side as second equation defines  $h(\delta x)$ .

When  $x < 0$ , in which case  $\delta x_{\text{MLI}}$  must be positive,  $\delta \mathbf{s}_{\text{MLI}}$  is the crossing point of the two curves given by Eqs. (D.7) and (D.9). As  $\partial g(\delta x)/\partial \delta x < 0$  and  $\partial h(\delta x)/\partial \delta x > 0$  hold for positive  $\delta x$  and  $D > 0$ ,  $g(\delta x)$  and  $h(\delta x)$  are monotonically decreasing and increasing functions of  $\delta x$ , respectively, for positive  $\delta x$  and  $D > 0$ . Suppose that  $D > 1/\sqrt{2}$ . Then  $h(1) - g(1) < 0$  and  $h(\sqrt{2}) - g(\sqrt{2}) > 0$ . Since  $h(\delta x) - g(\delta x)$  is a strictly increasing function of  $\delta x$  and is zero for  $\delta x = \delta x_{\text{MLI}}$ , it follows that  $1 < \delta x_{\text{MLI}} < \sqrt{2}$  and hence that  $\delta y_{\text{MLI}} < \min\{g(1), h(\sqrt{2})\} = g(1) = D < D\delta x_{\text{MLI}}^2$ . In addition, as  $h(\delta x)$  is always positive for positive  $\delta x$ ,  $\delta y_{\text{MLI}} > 0$  holds. Thus, inequalities (D.1b) and (D.1c) both hold for  $x < 0$ .

When  $x > 0$ , reversing the direction of the  $x$ -axis (i.e., multiplying  $x$  and  $\delta x$  by  $-1$ ) yields a situation identical to the case  $x < 0$ , without loss of generality. Thus, inequalities (D.1b) and (D.1c) both hold also for  $x > 0$ .

1 When  $x = 0$ , the MLI mutant is explicitly obtained from Eq. (17) in Section 5. Clearly,  
 2 inequalities (D.1b) and (D.1c) both hold for  $D > 1/\sqrt{2}$ . This completes the proof.

### 3 **Proof of Lemma D.2**

4 For the proof in this subsection, we denote MLI mutants as functions of the resident  
 5 phenotypes, i.e.,  $\mathbf{s}'_{\text{MLI}}(\mathbf{s})$  for monomorphism and  $\mathbf{s}'_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2)$  for dimorphism. Then, the  
 6 MLI mutations in these cases can be expressed as  $\delta\mathbf{s}_{\text{MLI}}(\mathbf{s}) = \mathbf{s}'_{\text{MLI}}(\mathbf{s}) - \mathbf{s}$  and  
 7  $\delta\mathbf{s}_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{s}'_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2) - \mathbf{s}_{\text{MLI}}$ , respectively, where  $\mathbf{s}_{\text{MLI}}$  with  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_1$  or  $\mathbf{s}_2$  is the  
 8 parental resident of the MLI mutant  $\mathbf{s}'_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2)$ .

9 We prove Lemma D.2 by demonstrating that the MLI mutation under a dimorphism of  $\mathbf{s}_1$   
 10 and  $\mathbf{s}_2$  is identical to that under monomorphism of an appropriately chosen  $\tilde{\mathbf{s}}$ ,

$$11 \quad \tilde{\mathbf{s}} = ([D(2x_{\text{MLI}} - x_2 - x_1) - (y_2 - y_1)/(x_2 - x_1)]/C, 0)^T, \quad (\text{D.10})$$

12 i.e.,  $\delta\mathbf{s}_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2) = \delta\mathbf{s}_{\text{MLI}}(\tilde{\mathbf{s}})$ . Then, provided that Lemma D.1 holds, substitution of  
 13  $\delta\mathbf{s}_{\text{MLI}}(\tilde{\mathbf{s}}) = \delta\mathbf{s}_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2)$  and  $\tilde{\mathbf{s}}$  into Eqs. (D.1) immediately gives Lemma D.2. The proof of  
 14  $\delta\mathbf{s}_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2) = \delta\mathbf{s}_{\text{MLI}}(\tilde{\mathbf{s}})$  is as follows.

15 The MLI mutation  $\delta\mathbf{s}_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2)$  is given by the  $\delta\mathbf{s}$  that maximizes Eq. (9b) in Section 4,

$$16 \quad P_{\text{MLI}}(\delta\mathbf{s} + \mathbf{s}_{\text{MLI}}; \mathbf{s}_1, \mathbf{s}_2) = T\mu\hat{n}_{\text{MLI}}M(\delta\mathbf{s})f(\delta\mathbf{s} + \mathbf{s}_{\text{MLI}}; \mathbf{s}_1, \mathbf{s}_2) \quad (\text{D.11})$$

17 for  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_1$  (with  $P_{\text{MLI}} = P_1$  and  $\hat{n}_{\text{MLI}} = \hat{n}_1$ ) or  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_2$  (with  $P_{\text{MLI}} = P_2$  and  $\hat{n}_{\text{MLI}} = \hat{n}_2$ ).

18 Here, the invasion-fitness function under dimorphism,  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$ , is approximately given by

$$19 \quad f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = (y' - y_2) + D(x' - x_1)(x' - x_2) - \frac{y_2 - y_1}{x_2 - x_1}(x' - x_2), \quad (\text{D.12})$$

20 as long as the dimorphic residents are still close to the base point of the expansion,

21  $\mathbf{s}_0 = (0, 0)^T$  (see Appendix E for the derivation). This function can be expressed in the form  
 22 of a monomorphic invasion-fitness function, Eq. (12a),

$$23 \quad f(\delta\mathbf{s} + \mathbf{s}_{\text{MLI}}; \mathbf{s}_1, \mathbf{s}_2) = f(\delta\mathbf{s} + \tilde{\mathbf{s}}; \tilde{\mathbf{s}}), \quad (\text{D.13})$$

24 by choosing  $\tilde{\mathbf{s}}$  as in Eq. (D.10). With this relationship, Eq. (D.11) yields

$$\begin{aligned}
P_{\text{MLI}}(\delta\mathbf{s} + \mathbf{s}_{\text{MLI}}; \mathbf{s}_1, \mathbf{s}_2) &= T \mu \hat{n}_{\text{MLI}} M(\delta\mathbf{s}) f(\delta\mathbf{s} + \tilde{\mathbf{s}}; \tilde{\mathbf{s}}) \\
&= \frac{\hat{n}_{\text{MLI}}}{\tilde{n}} T \mu \tilde{n} M(\delta\mathbf{s}) f(\delta\mathbf{s} + \tilde{\mathbf{s}}; \tilde{\mathbf{s}}) \\
&= \frac{\hat{n}_{\text{MLI}}}{\tilde{n}} P(\delta\mathbf{s} + \tilde{\mathbf{s}}; \tilde{\mathbf{s}}),
\end{aligned} \tag{D.14}$$

where  $\tilde{n}$  is the equilibrium population size of the monomorphic resident  $\tilde{\mathbf{s}}$ . Since both  $\hat{n}_{i_{\text{MLI}}}$  and  $\tilde{n}$  do not depend on  $\delta\mathbf{s}$ , any fixed  $\mathbf{s}_1$  and  $\mathbf{s}_2$  fulfill  $P_{\text{MLI}}(\delta\mathbf{s} + \mathbf{s}_{\text{MLI}}; \mathbf{s}_1, \mathbf{s}_2) \propto P(\delta\mathbf{s} + \tilde{\mathbf{s}}; \tilde{\mathbf{s}})$ . Therefore, the  $\delta\mathbf{s} = \delta\mathbf{s}_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{s}'_{\text{MLI}}(\mathbf{s}_1, \mathbf{s}_2) - \mathbf{s}_{\text{MLI}}$  that maximizes  $P_{\text{MLI}}(\delta\mathbf{s} + \mathbf{s}_{\text{MLI}}; \mathbf{s}_1, \mathbf{s}_2)$  is identical to the  $\delta\mathbf{s} = \delta\mathbf{s}_{\text{MLI}}(\tilde{\mathbf{s}}) = \mathbf{s}'_{\text{MLI}}(\tilde{\mathbf{s}}) - \tilde{\mathbf{s}}$  that maximizes  $P(\delta\mathbf{s} + \tilde{\mathbf{s}}; \tilde{\mathbf{s}})$ . This completes the proof.

## Appendix E: Invasion-fitness functions under dimorphism

Here we approximate dimorphic invasion-fitness functions  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  by a form similar to that derived for monomorphic invasion-fitness functions, Eq. (D.12) in Appendix D. First, we assume that  $|\mathbf{s}_2 - \mathbf{s}_1|$  is sufficiently small for the function  $f$  to be approximated using the first and second derivatives only. However, the direct Taylor expansion of  $f$  with respect to  $\mathbf{s}'$ ,  $\mathbf{s}_1$ , and  $\mathbf{s}_2$  up to second order cannot generally satisfy the consistency condition  $f(\mathbf{s}_1; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}_2) = 0$ .

Under the restrictive assumption that the two residents are near an evolutionarily singular point, this problem can be solved by allowing the invasion-fitness function to have a rational form composed of its first and second derivatives (Durinx et al., 2008). In our study, however, the two residents may be distant from a singular point in the  $y$ -direction. Here we therefore generalize the result by Durinx et al. (2008) by showing that for arbitrary dimorphic residents dimorphic invasion-fitness functions can be approximated by a rational form composed of its first and second derivatives.

### Special case

We first consider a special case that arises when the two residents  $\mathbf{s}_1$  and  $\mathbf{s}_2$  are both located exactly on the  $x$ -axis, i.e.,  $\mathbf{s}_1 = (x_1, 0)^T$  and  $\mathbf{s}_2 = (x_2, 0)^T$ , in which case the dimorphic invasion-fitness function  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  can be approximated by a quadratic function

1 of  $\mathbf{s}'$ ,  $\mathbf{s}_1$ , and  $\mathbf{s}_2$  as follows. To satisfy  $f(\mathbf{s}_1; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}_2) = 0$ , the quadratic function  
 2 must be expressed as

$$3 \quad f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2} \tilde{D}_{xx} (x' - x_1)(x' - x_2) + \left[ \tilde{D}_{xy} x' + \frac{1}{2} \tilde{D}_{yy} y' + \tilde{C}_{xy1} x_1 + \tilde{C}_{xy2} x_2 + \tilde{G}_y \right] y', \quad (\text{E.1})$$

4 with unknown constant parameters  $\tilde{D}_{xx}$ ,  $\tilde{D}_{xy}$ ,  $\tilde{D}_{yy}$ ,  $\tilde{C}_{xy1}$ ,  $\tilde{C}_{xy2}$ , and  $\tilde{G}_y$  to be specified.

5 For satisfying  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}'; \mathbf{s}_2, \mathbf{s}_1)$ ,  $\tilde{C}_{xy1} = \tilde{C}_{xy2} = \tilde{C}_{xy}$  is required. The other parameters  
 6 can be determined by comparing with the monomorphic invasion-fitness functions

$$7 \quad f(\mathbf{s}'; \mathbf{s}_1) = \mathbf{G}(\mathbf{s}' - \mathbf{s}_1) + \mathbf{s}_1^T \mathbf{C}(\mathbf{s}' - \mathbf{s}_1) + \frac{1}{2} (\mathbf{s}' - \mathbf{s}_1)^T \mathbf{D}(\mathbf{s}' - \mathbf{s}_1) \quad (\text{E.2a})$$

8 and

$$9 \quad f(\mathbf{s}'; \mathbf{s}_2) = \mathbf{G}(\mathbf{s}' - \mathbf{s}_2) + \mathbf{s}_2^T \mathbf{C}(\mathbf{s}' - \mathbf{s}_2) + \frac{1}{2} (\mathbf{s}' - \mathbf{s}_2)^T \mathbf{D}(\mathbf{s}' - \mathbf{s}_2), \quad (\text{E.2b})$$

10 where  $\mathbf{s}_0 = (0, 0)^T$  is assumed without loss of generality.

11 Now we consider a continuous shift in the resident phenotypes  $\mathbf{s}_1$  and/or  $\mathbf{s}_2$  in a way that  
 12 maintains their coexistence (i.e.,  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$  and  $f(\mathbf{s}_2; \mathbf{s}_1) > 0$ ), such that the population  
 13 size of  $\mathbf{s}_2$ , denoted by  $\hat{n}_2$ , converges to zero. Then  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  has to converge to  $f(\mathbf{s}'; \mathbf{s}_1)$ .

14 As derived in the last subsection in this appendix,  $\hat{n}_2 \rightarrow +0$  while  $\hat{n}_1 > 0$  implies

15  $f(\mathbf{s}_2; \mathbf{s}_1) = 0$  and  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ . This consideration yields the consistency condition

$$16 \quad f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}'; \mathbf{s}_1) \text{ for } f(\mathbf{s}_2; \mathbf{s}_1) = 0 \text{ and } f(\mathbf{s}_1; \mathbf{s}_2) > 0. \quad (\text{E.3a})$$

17 In the same manner, considering the case of  $\hat{n}_1 \rightarrow +0$  while  $\hat{n}_2 > 0$  yields another

18 consistency condition,

$$19 \quad f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}'; \mathbf{s}_2) \text{ for } f(\mathbf{s}_1; \mathbf{s}_2) = 0 \text{ and } f(\mathbf{s}_2; \mathbf{s}_1) > 0. \quad (\text{E.3b})$$

20 First, we examine the consistency condition Eq. (E.3a). The condition  $f(\mathbf{s}_2; \mathbf{s}_1) = 0$  is

21 transformed into

$$22 \quad x_2 = x_1 - \frac{2C_{xx}x_1 + 2G_x}{D_{xx}}, \quad (\text{E.4})$$

23 which upon substitution into  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}'; \mathbf{s}_1)$  gives

$$\begin{aligned}
f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) &= \left[ \frac{\tilde{D}_{xx}}{D_{xx}} G_x (x' - x_1) + \frac{\tilde{D}_{xx}}{D_{xx}} C_{xx} x_1 (x' - x_1) + \frac{1}{2} \tilde{D}_{xx} (x' - x_1)^2 \right] \\
&\quad + \left[ \tilde{D}_{xy} (x' - x_1) y' + \frac{1}{2} \tilde{D}_{yy} y'^2 \right] \\
&\quad + \left( \tilde{G}_y - \frac{2\tilde{C}_{xy} G_x}{D_{xx}} \right) y' + \left( \tilde{D}_{xy} + 2\tilde{C}_{xy} - \frac{2\tilde{C}_{xy} C_{xx}}{D_{xx}} \right) x_1 y' \\
1 \quad &= \left[ G_x (x' - x_1) + C_{xx} x_1 (x' - x_1) + \frac{1}{2} D_{xx} (x' - x_1)^2 \right] \tag{E.5} \\
&\quad + \left[ D_{xy} (x' - x_1) y' + \frac{1}{2} D_{yy} y'^2 \right] \\
&\quad + G_y y' + C_{xy} x_1 y' \\
&= f(\mathbf{s}'; \mathbf{s}_1).
\end{aligned}$$

2 This equation must be satisfied for arbitrary  $x' - x_1$ ,  $x_1$ , and  $y'$  as long as  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ .  
3 Comparing the coefficients for  $x' - x_1$ ,  $x_1$ , and  $y'$  at each order, we can specify the un-  
4 known parameters as

$$\begin{aligned}
5 \quad &\tilde{D}_{xx} = D_{xx}, \quad \tilde{D}_{xy} = D_{xy}, \quad \tilde{D}_{yy} = D_{yy}, \\
&\tilde{C}_{xy} = \frac{D_{xx} (C_{xy} - D_{xy})}{2(D_{xx} - C_{xx})}, \quad \tilde{G}_y = G_y + \frac{G_x (C_{xy} - D_{xy})}{D_{xx} - C_{xx}}, \tag{E.6}
\end{aligned}$$

6 where  $D_{xx} - C_{xx} > 0$  is ensured by  $f(\mathbf{s}_1; \mathbf{s}_2) = 0$  and  $f(\mathbf{s}_2; \mathbf{s}_1) > 0$ .

7 As the parameters thus specified also satisfy the other consistency condition, Eq. (E.3b),  
8 Eq. (E.1) with [EqEqs. \(E.6\)](#) is an appropriate quadratic approximation of the dimorphic  
9 invasion-fitness function  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$ . It can be shown that Eq. (E.1) with [EqEqs. \(E.6\)](#) can be  
10 further transformed into

$$11 \quad f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = w_1 f(\mathbf{s}'; \mathbf{s}_1) + w_2 f(\mathbf{s}'; \mathbf{s}_1) + h, \tag{E.7}$$

12 where  $w_1 = f_{12} / (f_{12} + f_{21})$ ,  $w_2 = f_{21} / (f_{12} + f_{21})$ , and  $h = f_{12} f_{21} / (f_{12} + f_{21})$ , with  
13  $f_{12} = f(\mathbf{s}_1; \mathbf{s}_2)$  and  $f_{21} = f(\mathbf{s}_2; \mathbf{s}_1)$ .

## 14 General case

15 Next, we consider the general case in which the two residents  $\mathbf{s}_1$  and  $\mathbf{s}_2$  are [located](#) neither  
16 [located](#) on the  $x$ -axis (as in the previous subsection) [nor](#) near an evolutionarily singular

1 point (as in Durinx et al., 2008). As the dimorphic invasion-fitness function obtained for the  
 2 special case above, Eq. (E.7), has a form independent of the coordinate system (i.e.,  
 3 independent of how the  $x$ - and  $y$ -axes are chosen), it is expected that even in this general  
 4 case the function is obtained in a form identical to Eq. (E.7). Below, we confirm this  
 5 conjecture.

6 First, to treat this general case analogously to the special case, we introduce a new coordi-  
 7 nate system  $\check{\mathbf{s}} = (\check{x}, \check{y})^T$  so that  $\mathbf{s}_1$  and  $\mathbf{s}_2$  are both located on the  $\check{x}$ -axis, i.e.,  $\check{\mathbf{s}}_1 = (\check{x}_1, 0)^T$   
 8 and  $\check{\mathbf{s}}_2 = (\check{x}_2, 0)^T$ , by an affine coordinate transformation,

$$9 \quad \mathbf{s} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \begin{pmatrix} \check{x} \\ \check{y} \end{pmatrix} + \begin{pmatrix} 0 \\ y_1 - ax_1 \end{pmatrix} = \mathbf{A}\check{\mathbf{s}} + \mathbf{b}, \quad (\text{E.8})$$

10 with  $a = (y_2 - y_1)/(x_2 - x_1)$ . Then, in the new coordinate system,  $f(\mathbf{s}'; \mathbf{s}_1)$  and  $f(\mathbf{s}'; \mathbf{s}_2)$   
 11 are expressed as

$$12 \quad \begin{aligned} \check{f}(\check{\mathbf{s}}'; \check{\mathbf{s}}_1) &= f(\mathbf{s}'; \mathbf{s}_1) = f(\mathbf{A}\check{\mathbf{s}}' + \mathbf{b}; \mathbf{A}\check{\mathbf{s}}_1 + \mathbf{b}) \\ &= \check{\mathbf{G}}(\check{\mathbf{s}}' - \check{\mathbf{s}}_1) + \check{\mathbf{s}}_1^T \check{\mathbf{C}}(\check{\mathbf{s}}' - \check{\mathbf{s}}_1) + \frac{1}{2}(\check{\mathbf{s}}' - \check{\mathbf{s}}_1)^T \check{\mathbf{D}}(\check{\mathbf{s}}' - \check{\mathbf{s}}_1) \end{aligned} \quad (\text{E.9a})$$

13 and

$$14 \quad \begin{aligned} \check{f}(\check{\mathbf{s}}'; \check{\mathbf{s}}_2) &= f(\mathbf{s}'; \mathbf{s}_2) = f(\mathbf{A}\check{\mathbf{s}}' + \mathbf{b}; \mathbf{A}\check{\mathbf{s}}_2 + \mathbf{b}) \\ &= \check{\mathbf{G}}(\check{\mathbf{s}}' - \check{\mathbf{s}}_2) + \check{\mathbf{s}}_2^T \check{\mathbf{C}}(\check{\mathbf{s}}' - \check{\mathbf{s}}_2) + \frac{1}{2}(\check{\mathbf{s}}' - \check{\mathbf{s}}_2)^T \check{\mathbf{D}}(\check{\mathbf{s}}' - \check{\mathbf{s}}_2), \end{aligned} \quad (\text{E.9b})$$

15 where  $\mathbf{s}_0 = (0, 0)^T$ , i.e.,  $\check{\mathbf{s}}_0 = \mathbf{b}$ , is assumed without loss of generality, and where  $\check{\mathbf{G}} = \mathbf{G}\mathbf{A}$ ,  
 16  $\check{\mathbf{C}} = \mathbf{A}^T \mathbf{C}\mathbf{A}$ , and  $\check{\mathbf{D}} = \mathbf{A}^T \mathbf{D}\mathbf{A}$ . Therefore, in the same manner as in the special case, we obtain  
 17 the quadratic approximation of the dimorphic invasion-fitness function, in a form identical to  
 18 Eq. (E.7),

$$19 \quad \check{f}(\check{\mathbf{s}}'; \check{\mathbf{s}}_1, \check{\mathbf{s}}_2) = \check{w}_1 \check{f}(\check{\mathbf{s}}'; \check{\mathbf{s}}_1) + \check{w}_2 \check{f}(\check{\mathbf{s}}'; \check{\mathbf{s}}_2) + \check{h}, \quad (\text{E.10})$$

20 where  $\check{w}_1 = \check{f}_{12} / (\check{f}_{12} + \check{f}_{21})$ ,  $\check{w}_2 = \check{f}_{21} / (\check{f}_{12} + \check{f}_{21})$ , and  $\check{h} = \check{f}_{12}\check{f}_{21} / (\check{f}_{12} + \check{f}_{21})$ , with  
 21  $\check{f}_{12} = \check{f}(\check{\mathbf{s}}_1; \check{\mathbf{s}}_2)$  and  $\check{f}_{21} = \check{f}(\check{\mathbf{s}}_2; \check{\mathbf{s}}_1)$ .

22 Here, Eq. (E.10) directly gives the dimorphic invasion-fitness function  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  in the  
 23 original coordinate system,



$$1 \quad f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = \tilde{f}(\mathbf{A}^{-1}(\mathbf{s}' - \mathbf{b}); \mathbf{A}^{-1}(\mathbf{s}_2 - \mathbf{b}), \mathbf{A}^{-1}(\mathbf{s}_2 - \mathbf{b})) = \tilde{f}(\tilde{\mathbf{s}}'; \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2). \quad (\text{E.11})$$

2 In addition,  $\tilde{f}(\tilde{\mathbf{s}}'; \tilde{\mathbf{s}}_1) = f(\mathbf{s}'; \mathbf{s}_1)$  and  $\tilde{f}(\tilde{\mathbf{s}}'; \tilde{\mathbf{s}}_2) = f(\mathbf{s}'; \mathbf{s}_2)$  both hold according to Eqs. (E.9a)  
3 and (E.9b), which upon substitution into Eq. (E.10) yields Eq. (E.7), as expected. Therefore,  
4 in general a dimorphic invasion-fitness function can be approximated by a rational form  
5 composed of its first and second derivatives, Eq. (E.7).

6 Even if the trait space is multivariate, with arbitrary dimension  $L$ , denoted by  
7  $\mathbf{s} = (u_1, \dots, u_L)^\top$ , we can find, for arbitrary  $\mathbf{s}'$ ,  $\mathbf{s}_1$ , and  $\mathbf{s}_2$ , a bivariate subspace that contains  
8 all three of these phenotypes. In this subspace, Eq. (E.7) holds, where the monomorphic and  
9 dimorphic invasion-fitness functions defined on  $\mathbf{s} = (u_1, \dots, u_L)^\top$  can be directly used, as long  
10 as we consider  $\mathbf{s}'$  to be restricted to this subspace. As we can always find such a subspace  
11 for any  $\mathbf{s}'$ ,  $\mathbf{s}_1$ , and  $\mathbf{s}_2$ , the dimorphic invasion-fitness function  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  on  
12  $\mathbf{s} = (u_1, \dots, u_L)^\top$  is given by Eq. (E.7), by using the monomorphic function  $f(\mathbf{s}'; \mathbf{s})$  on  
13  $\mathbf{s} = (u_1, \dots, u_L)^\top$ .

14 Finally, substituting the monomorphic invasion-fitness function in Eq. (12a) in the main  
15 text into Eq. (E.7) yields the corresponding dimorphic invasion-fitness function, Eq. (D.12).

## 16 Derivation of consistency condition

17 Here we derive the consistency condition  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}'; \mathbf{s}_1)$  for  $f(\mathbf{s}_2; \mathbf{s}_1) = 0$  and  
18  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ , by proving that  $\hat{n}_2 \rightarrow +0$  holds for  $f(\mathbf{s}_2; \mathbf{s}_1) \rightarrow +0$  while  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ .

19 First, we assume a protected dimorphism of  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , i.e.,  $f(\mathbf{s}_2; \mathbf{s}_1) > 0$  and  
20  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ . This is possible only when the resident phenotypes  $\mathbf{s}_1$  and  $\mathbf{s}_2$  are both in the  
21 neighborhood of an evolutionarily singular point in the one-dimensional trait subspace  
22 defined by the  $\tilde{x}$ -axis in Eq. (E.8). Then, the population dynamics of  $\mathbf{s}_1$  and  $\mathbf{s}_2$  can be  
23 approximated by Lotka-Volterra equations (Durinx et al., 2008), giving a unique interior  
24 stable equilibrium  $(\hat{n}_1, \hat{n}_2)$ ; see also Appendix M. Thus, when  $f(\mathbf{s}_2; \mathbf{s}_1)$  converges to zero  
25 while  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ ,  $\hat{n}_1$  converges to zero, while  $\hat{n}_2 > 0$ . In this case,  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  must  
26 converge to  $f(\mathbf{s}'; \mathbf{s}_1)$ . Therefore,  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) = f(\mathbf{s}'; \mathbf{s}_1)$  for  $f(\mathbf{s}_2; \mathbf{s}_1) = 0$  and  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$   
27 holds.

## 1 Appendix F: Mutational step sizes of MLI mutants

2 Here we prove that MLI mutants  $\mathbf{s}'_{\text{MLI}}$ , which maximize Eqs. (11) in Section 4, satisfy  
 3  $|\mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}}| \leq \sqrt{2}\sigma$ , where  $\mathbf{s}_{\text{MLI}}$  is the parental resident, when invasion-fitness functions are  
 4 approximated by quadratic functions of  $\mathbf{s}'$  (e.g., Eqs. 2) and mutation probability  
 5 distributions are approximated by multivariate Gaussian functions.

### 6 Main proof

7 We consider a **multivariate** trait space  $\mathbf{s} = (u_1, \dots, u_L)^T$  with arbitrary dimension  $L$  that is  
 8 normalized and rescaled so that mutation is isotropic with standard deviation 1. We assume  
 9 that the mutation probability distribution is approximated by a multivariate Gaussian function

$$10 \quad M(\delta\mathbf{s}) = \frac{1}{(2\pi)^{L/2}} \exp\left(-\frac{1}{2}|\delta\mathbf{s}|^2\right), \quad (\text{F.1})$$

11 with standard deviation 1 in all directions.

12 In a manner similar to Appendix A, we can expand the invasion-fitness function in  
 13 ~~around~~  $\mathbf{s}'$  about  $\mathbf{s}_{\text{MLI}}$  as

$$14 \quad f(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N) = \mathbf{G}\delta\mathbf{s} + \frac{1}{2}\delta\mathbf{s}^T \mathbf{D}\delta\mathbf{s} + \text{h.o.t.}, \quad (\text{F.2})$$

15 where  $\delta\mathbf{s} = \mathbf{s}' - \mathbf{s}_{\text{MLI}}$ ,  $\mathbf{G}$  is an  $L$ -dimensional row vector, and  $\mathbf{D}$  is an  $L \times L$  symmetric  
 16 matrix. Notice that  $\mathbf{G}$  and  $\mathbf{D}$  are both functions of  $\mathbf{s}_1, \dots, \mathbf{s}_N$  so that  $f(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N) = 0$   
 17 holds for all  $i = 1, \dots, N$ . For example,  $f(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N)$  for  $N = 2$  (i.e., for a dimorphism) is  
 18 given by Eq. (E.7) in Appendix E. We assume that the higher-order terms in Eq. (F.2) can be  
 19 neglected.

20 According to Eq. (11) in Section 4,  $\delta\mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}}$  maximizes

$$21 \quad \begin{aligned} P_{\text{MLI}}(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N) &= T \mu \hat{n}_{\text{MLI}} M(\delta\mathbf{s}) f(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N) \\ &= \frac{T \mu \hat{n}_{\text{MLI}}}{(2\pi)^{L/2}} \exp\left(-\frac{1}{2}|\delta\mathbf{s}|^2\right) \left[ \mathbf{G}\delta\mathbf{s} + \frac{1}{2}\delta\mathbf{s}^T \mathbf{D}\delta\mathbf{s} \right], \end{aligned} \quad (\text{F.3})$$

22 where  $\mathbf{s}_{\text{MLI}} = \mathbf{s}_{i_{\text{MLI}}}$ ,  $P_{\text{MLI}} = P_{i_{\text{MLI}}}$ ,  $\hat{n}_{\text{MLI}} = \hat{n}_{i_{\text{MLI}}}$ , and the conversion of negative invasion fitnesses  
 23 to zero (subscript “+”) is not needed here, as  $\delta\mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}}$  always provides positive  
 24 invasion fitnesses.

1 First, we consider the special case  $\mathbf{G} = \mathbf{0}$ . In this case,  $\mathbf{D} \neq \mathbf{0}$  is required for neglecting  
 2 the higher-order terms. When  $\mathbf{D}$  is negative definite, the MLI mutation is given by  
 3  $\delta \mathbf{s}_{\text{MLI}} = \mathbf{0}$ , otherwise,  $\delta \mathbf{s}_{\text{MLI}} = \pm \sqrt{2} \mathbf{v}_1$ , where  $\mathbf{v}_1$  is the eigenvector of the maximum  
 4 eigenvalue of  $\mathbf{D}$ , with  $|\mathbf{v}_1| = 1$ . Thus,  $|\delta \mathbf{s}_{\text{MLI}}| \leq \sqrt{2}$  holds for  $\mathbf{G} = \mathbf{0}$ .

5 Second, we consider the case  $\mathbf{G} \neq \mathbf{0}$ . We express  $\delta \mathbf{s}$  as  $\delta \mathbf{s} = z \mathbf{e}_z$ , where  $\mathbf{e}_z$  is the unit  
 6 vector parallel to the mutational step  $\delta \mathbf{s}$ , with  $|\mathbf{e}_z| = 1$ , and  $z$  is a scalar. Substituting  
 7  $\delta \mathbf{s} = z \mathbf{e}_z$  into Eq. (F.3) yields

$$8 \quad P_{\text{MLI}}(\mathbf{s}'; \mathbf{s}_1, \dots, \mathbf{s}_N) = A_0 \exp\left(-\frac{1}{2} z^2\right) \left[ G_z z + \frac{1}{2} D_z z^2 \right] = H(z), \quad (\text{F.4})$$

9 where  $A_0 = (T \mu \hat{n}_{\text{MLI}}) / (2\pi)^{L/2} > 0$ ,  $G_z = \mathbf{G} \mathbf{e}_z$ ,  $D_z = \mathbf{e}_z^T \mathbf{D} \mathbf{e}_z$ , and the second equality defines  
 10  $H(z)$ .

### 11 **Lemma F.1:**

12 The  $z = z_{\text{MLI}}$  that maximizes  $H(z)$  always satisfies  $|z_{\text{MLI}}| \leq \sqrt{2}$  for arbitrary  $G_z$   
 13 and  $D_z$ , except for the special case  $G_z = D_z = 0$ .

14 If there exists an  $\mathbf{e}_z$  with  $G_z = D_z = 0$ , then  $H(z) = 0$  for all  $z$ , so  $z_{\text{MLI}}$  cannot be  
 15 determined. However, as  $\mathbf{G} \neq \mathbf{0}$ , there always exist other  $\mathbf{e}_z$  with  $G_z \neq 0$ , which yield  
 16  $H(z) > 0$  for some  $z$  (e.g., for  $0 < z < 2G_z / |D_z|$  when  $\mathbf{e}_z = \mathbf{G}^T$ ). Thus,  $\delta \mathbf{s}_{\text{MLI}}$  is chosen  
 17 along those  $\mathbf{e}_z$ , which satisfy  $|z_{\text{MLI}}| \leq \sqrt{2}$  according to Lemma F.1. Therefore,  $|\delta \mathbf{s}_{\text{MLI}}| \leq \sqrt{2}$   
 18 holds also for  $\mathbf{G} \neq \mathbf{0}$ . This completes the proof.

### 19 **Proof of Lemma F.1**

20 When  $G_z = 0$ ,  $D_z$  must be non-zero, as Lemma F.1 excludes the special case  $G_z = D_z = 0$ .  
 21 Then,  $z_{\text{MLI}}$  is given by  $z_{\text{MLI}} = \pm \sqrt{2}$  for  $D_z > 0$ , or by  $z_{\text{MLI}} = 0$  for  $D_z < 0$ . Thus,  
 22  $|z_{\text{MLI}}| \leq \sqrt{2}$  holds.

23 When  $G_z \neq 0$ , we multiply  $\mathbf{e}_z$  by  $-1$  as necessary so that  $G_z > 0$  always holds,  
 24 without loss of generality. Then, for any negative  $z$ ,  $H(-z) > H(z)$  holds. Thus,  $z_{\text{MLI}}$ ,  
 25 which maximizes  $H(z)$ , satisfies  $z_{\text{MLI}} \geq 0$ . In addition, as  $H(z) \geq 0$  holds for  $z_{\text{MLI}}$ , the  
 26 expression in the square bracket in Eq. (F.4) satisfies  $G_z + \frac{1}{2} D_z z_{\text{MLI}} \geq 0$ . Thus,  $z_{\text{MLI}}$  satisfies

1 
$$0 \leq z_{\text{MLI}} \text{ for } D_z \geq 0, \quad (\text{F.5a})$$

2 and

3 
$$0 \leq z_{\text{MLI}} \leq \frac{2G_z}{-D_z} \text{ for } D_z < 0. \quad (\text{F.5b})$$

4 Notice that  $H(z)$  is a smooth function of  $z$ . As  $H(0) = 0$ ,  $H(+\infty) = 0$ , and  $H(z) > 0$  for  
 5  $z < 2G_z / |D_z|$ ,  $H(z)$  has a positive maximum for finite  $z = z_{\text{MLI}}$ , fulfilling

6 
$$\frac{dH(z)}{dz} = A_0 \exp\left(-\frac{1}{2}z^2\right) \left[ G_z(1-z^2) + \frac{1}{2}D_z z(2-z^2) \right] = 0. \quad (\text{F.6})$$

7 Thus, when  $D_z \geq 0$ , the  $z_{\text{MLI}}$ , which satisfies both Eqs. (F.5a) and (F.6), must satisfy

8  $1 \leq z_{\text{MLI}} \leq \sqrt{2}$ . On the other hand, when  $D_z < 0$ , Eq. (F.6) can be expressed as

9 
$$\frac{dH(z)}{dz} = A_0 \exp\left(-\frac{1}{2}z^2\right) \left[ \left(G_z + \frac{1}{2}D_z z\right)(1-z^2) + \frac{1}{2}D_z z \right] = 0, \quad (\text{F.7})$$

10 where  $G_z + \frac{1}{2}D_z z \geq 0$  holds for  $z = z_{\text{MLI}}$  according to Eq. (F.5b). Then, the  $z_{\text{MLI}}$ , which  
 11 satisfies both Eqs. (F.5b) and (F.7), must satisfy  $0 \leq z_{\text{MLI}} \leq 1$ . Therefore,  $|z_{\text{MLI}}| \leq \sqrt{2}$  holds  
 12 for both  $G_z = 0$  and  $G_z \neq 0$ . This completes the proof.

## 13 **Appendix G: Proof of Lemma 1 (conditions for dimorphic** 14 **emergence)**

15 Here we prove Lemma 1 in Section 5, specifying conditions on MLI mutants to ensure  
 16 dimorphic emergence. While this lemma refers to MLI mutants, it also holds for non-MLI  
 17 mutants as long as they satisfy the conditions for dimorphic emergence, inequalities (13).

18 Thus, here we do not distinguish between  $\mathbf{s}'$  and  $\mathbf{s}'_{\text{MLI}}$ , and denote mutants simply by  $\mathbf{s}'$ .

19 We consider a monomorphic resident  $\mathbf{s} = (x, y)^T$ . A sufficient condition for protected  
 20 dimorphism of  $\mathbf{s}$  and  $\mathbf{s}'$ , and thus for dimorphic emergence, is given by mutual invasibility,

21 
$$\begin{aligned} f(\mathbf{s}'; \mathbf{s}) &= \delta y + D\delta x^2 + Cx\delta x > 0, \\ f(\mathbf{s}; \mathbf{s}') &= -\delta y + D\delta x^2 - C(x + \delta x)\delta x > 0. \end{aligned} \quad (\text{G.1})$$

22 These inequalities can be combined into

23 
$$0 < \delta y + D\delta x^2 + Cx\delta x < (2D - C)\delta x^2. \quad (\text{G.2})$$

1 We first suppose that the population is not close to the convergence-stable line  $x = 0$ , so that  
 2  $|x|$  is significantly larger than 1. Since the magnitudes of  $\delta x$  and  $\delta y$  are of order 1,  
 3  $|(2D - C)\delta x^2| \ll |Cx\delta x|$  holds. In this case, inequalities (G.2) cannot hold. This means that  
 4 invasion by  $s'$  always replaces  $s$ , which corresponds to directional evolution. To satisfy  
 5 inequalities (G.2), the population has to come close to the convergence-stable line through  
 6 directional evolution, such that  $|x|$  becomes sufficiently small. Such convergence is ensured  
 7 if all invading mutants satisfy

$$8 \quad \begin{cases} x' > x & \text{for } x < 0 \\ x' < x & \text{for } x > 0. \end{cases} \quad (\text{G.3})$$

9 This implies  $\delta x > 0$  for  $x < 0$  and  $\delta x < 0$  for  $x > 0$ . The population monotonically  
 10 converges to the convergence-stable line if all invading mutants satisfy this condition, as long  
 11 as the resident is monomorphic.

12 On this basis, we now suppose that the population has come close to the convergence-  
 13 stable line and that a mutant has arisen such that the resident and the mutant straddle the  
 14 convergence-stable line,  $xx' < 0$ . As  $C$  is negative, both  $Cx\delta x = -C(-xx' + x^2)$  and  
 15  $-C(x + \delta x)\delta x = -C(x'^2 - xx')$  are always positive. Thus, inequalities (G.2) hold if

$$16 \quad x'x < 0 \quad (\text{G.4})$$

17 and

$$18 \quad \frac{|\delta y|}{\delta x^2} < D. \quad (\text{G.5})$$

19 Since directional evolution proceeds toward the convergence-stable line under inequalities  
 20 (G.3), the situation  $xx' < 0$  inevitably occurs, unless protected dimorphism emerges even  
 21 before that. Thus, for an arbitrary initial resident  $s_a$ , if all subsequent invading mutants  
 22 satisfy inequalities (G.3) and (G.5), then the population monotonically converges to the line  
 23  $x = 0$ , until protected dimorphism has emerged, which inevitably occurs once  $s$  and  $s'$   
 24 straddle the line. Inequalities (G.3) and (G.5) are identical to inequalities (13) in Section 5.  
 25 This completes the proof.

# 1 Appendix H: Proof of Lemma 2 (conditions for dimorphic 2 divergence)

## 3 Preparation

4 Here we prove Lemma 2 in Section 5, specifying conditions on MLI mutants to ensure  
5 dimorphic divergence. Similarly to Lemma 1, Lemma 2 also holds for non-MLI mutants as  
6 long as they satisfy the conditions for dimorphic divergence, inequalities (14). Thus, as in  
7 Appendix G, here we do not distinguish between  $s'$  and  $s'_{\text{MLI}}$ , and denote mutants simply by  
8  $s'$ .

9 After the emergence of a protected dimorphism, composed of two residents denoted by  $s_1$   
10 and  $s_2$ , the next invasion event occurs based on the dimorphic invasion fitness  $f(s'; s_1, s_2)$ .  
11 As long as the two residents remain close to the base point of the expansion,  $s_0 = (0, 0)^T$ , this  
12 dimorphic invasion fitness is approximately given by Eq. (D.12),

$$13 \quad f(s'; s_1, s_2) = (y' - y_2) + D(x' - x_1)(x' - x_2) - \frac{y_2 - y_1}{x_2 - x_1}(x' - x_2), \quad (\text{H.1})$$

14 as shown in Appendix E. This dimorphic invasion-fitness function  $f(s'; s_1, s_2)$  and the  
15 monomorphic invasion-fitness function  $f(s'; s)$  given by Eq. (12a) together determine  
16 whether a sequence of invading mutants can bring about dimorphic divergence.

## 17 Conditions for a single step of dimorphic divergence

18 We define dimorphic divergence as the directional evolution of two resident morphs in  
19 opposite directions along the  $x$ -axis. Such a compound evolutionary process is formed by  
20 repetition of a unit process defined as follows.

21 **Definition of a single step of dimorphic divergence:** An invading mutant replaces  
22 only either of [the two](#) residents and coexists with the other resident, and the phenotypic  
23 distance along the  $x$ -axis between the new residents is larger than that between the  
24 [previous](#) residents.

25

1 For a more formal description, we now consider arbitrary phenotypes  $s'$ ,  $s_1$ , and  $s_2$ . The  
 2 resident that is replaced by  $s'$  is denoted by  $s_p$  (i.e.,  $s_p = s_1$  or  $s_p = s_2$ ). The other  
 3 resident that coexists with  $s'$  is denoted by  $s_q$  (i.e.,  $s_q = s_2$  for  $s_p = s_1$ , or  $s_q = s_1$  for  
 4  $s_p = s_2$ ). Although  $s'$  usually replaces its parental resident (i.e.,  $s_p$  is its parental resident),  
 5  $s'$  may replace the other non-parental resident (i.e.,  $s_p$  is the non-parental resident) when  
 6 the two residents are close to each other. Then, the definition above of a single step of  
 7 dimorphic divergence is fulfilled under the following conditions.

8 **Sufficient conditions for a single step of dimorphic divergence:**

9 (a)  $f(s'; s_p, s_q) > 0$  ( $s'$  can invade)

10 (b)  $f(s_q; s') > 0$  and  $f(s'; s_q) > 0$  ( $s'$  coexists with  $s_q$ , i.e.,  $s_q$  owing to mutual in-  
 11 vasibility)

12 (c)  $f(s_p; s_q, s') < 0$  ( $s'$  excludes  $s_p$ )

13 (d)  $|x' - x_q| > |x_2 - x_1|$  (phenotypic divergence along the  $x$ -axis becomes larger)

14 On this basis, we prove the following lemma in Appendix I.

15 **Lemma H.1:**

16 Sufficient conditions for satisfying conditions (a) to (d) above are given by

17 
$$|y_2 - y_1| < D(x_2 - x_1)^2, \quad (\text{H.2a})$$

18 
$$(x_p - x_q)\delta x > 0, \quad (\text{H.2b})$$

19 and

20 
$$|\delta y| < D\delta x^2, \quad (\text{H.2c})$$

21 where  $(\delta x, \delta y)^T = (x' - x_p, y' - y_p)^T = s' - s_p$ .

22 Therefore, for dimorphic residents  $s_1$  and  $s_2$  satisfying inequality (H.2a), any mutant  
 23 satisfying inequalities (H.2b) and (H.2c) ensures a single step of dimorphic divergence.

24 The set of mutants  $s'$  that satisfy inequalities (H.2b) and (H.2c) is illustrated as the white  
 25 regions in Fig. 3d.

## 1 **Conditions for the whole process of dimorphic divergence**

2 We suppose that an invading mutant  $s'$  and residents  $s_1$  and  $s_2$  satisfy inequalities (H.2).  
3 Without loss of generality, we assume  $s_p = s_2$  and  $s_q = s_1$  (differently from Lemma 2,  
4 which assumes  $x_1 < x_2$  instead). Then,  $s'$  excludes only  $s_2$  and coexists with  $s_1$ . In  
5 addition, as proved in Appendix I, the new dimorphism composed of  $s'$  and  $s_1$  satisfies

$$6 \quad |y' - y_1| < D(x' - x_1)^2. \quad (\text{H.3})$$

7 When  $s'$  is renamed as ~~its replacing~~ the replaced resident  $s_2$ , inequality (H.3) gives  
8 inequality (H.2a). Thus, for the next step of dimorphic divergence, only inequalities (H.2b)  
9 and (H.2c) have to hold, and the same applies for subsequent steps of dimorphic divergence.  
10 Therefore, for any initial dimorphic residents satisfying inequality (H.2a), the whole process  
11 of subsequent dimorphic divergence is ensured, if all of the subsequent invading mutants  
12 satisfy inequalities (H.2b) and (H.2c). In addition, any initial protected dimorphism emerged  
13 under the conditions for dimorphic emergence, inequalities (13) in Section 5, clearly satisfies  
14 inequality (H.2a). Thus, provided that the initial dimorphism has emerged under the  
15 conditions for dimorphic emergence, sufficient conditions on the subsequent invading  
16 mutants for dimorphic divergence are given by inequalities (H.2b) and (H.2c). Inequalities  
17 (H.2b) and (H.2c) are equivalent to inequalities (14) in Section 5 when  $x_1 < x_2$  is assumed.  
18 This completes the proof ~~of Lemma 2~~.

## 19 **Appendix I: Proof of Lemma H.1 (conditions for a single** 20 **step of dimorphic divergence)**

21 Here we prove Lemma H.1 in Appendix H. We also show that conditions for a single step of  
22 dimorphic divergence, inequalities (H.2), ensure inequality (H.3).

### 23 **Main proof**

24 Without loss of generality, we assume that  $s_p = s_2$  and  $s_q = s_1$  (differently from Lemma 2,  
25 which assumes  $x_1 < x_2$ ). Then, the conditions (a) to (d) in Appendix H, for a single step of  
26 dimorphic divergence, become  $f(s'; s_1, s_2) > 0$ ,  $f(s_1; s') > 0$ ,  $f(s'; s_1) > 0$ ,  $f(s_2; s_1, s') < 0$ ,



1 and  $|x' - x_1| > |x_2 - x_1|$ . We denote the phenotypic differences of the original and the new  
2 residents by  $\Delta \mathbf{s} = \mathbf{s}_p - \mathbf{s}_q = \mathbf{s}_2 - \mathbf{s}_1$ , i.e.,  $(\Delta x, \Delta y)^T = (x_2 - x_1, y_2 - y_1)^T$ , and  $\Delta \mathbf{s}' = \mathbf{s}' - \mathbf{s}_q$ , i.e.,  
3  $(\Delta x', \Delta y')^T = (x' - x_1, y' - y_1)^T$ , respectively. We also define  $\tan \varphi = \Delta y / \Delta x$  and  
4  $\tan \varphi' = \Delta y' / \Delta x'$ .

5 On this basis, we prove the following two lemmas in the subsequent subsections.

6 **Lemma I.1:**

7 All inequalities  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) > 0$ ,  $f(\mathbf{s}_1; \mathbf{s}') > 0$ ,  $f(\mathbf{s}'; \mathbf{s}_1) > 0$ ,  $f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}') < 0$ , and  
8  $|x' - x_1| > |x_2 - x_1|$  hold if

$$9 \quad \Delta x' \Delta x > 0 \text{ and } |\Delta x| < |\Delta x'| \quad (\text{I.1a})$$

10 and

$$11 \quad |\tan \varphi' - \tan \varphi| < D |\Delta x' - \Delta x|. \quad (\text{I.1b})$$

12

13 **Lemma I.2:**

14 Inequalities (I.1) hold if inequalities (H.2) hold.

15

16 By these [Lemmas I.1 and I.2](#) lemmas, the proof of [Lemma H.1](#) is completed.

17 **Proof of Lemma I.1**

18 We assume that inequalities (I.1) hold. Inequalities (I.1a) immediately give  $|x' - x_1| > |x_2 - x_1|$ .

19 As for the signs of the invasion fitnesses, dividing  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)$  by  $\Delta x' - \Delta x = x' - x_2$  yields

$$20 \quad \frac{f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2)}{\Delta x' - \Delta x} = \left( \frac{y' - y_2}{x' - x_2} - \frac{y_2 - y_1}{x_2 - x_1} \right) + D[(x' - x_1) - (x_2 - x_1)] \quad (\text{I.2a})$$

$$= (\tan \varphi' - \tan \varphi) + D \Delta x'.$$

21 Similarly,  $f(\mathbf{s}'; \mathbf{s}_1)$ ,  $f(\mathbf{s}_1; \mathbf{s}')$ , and  $f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}')$  satisfy the following equations,

$$22 \quad \frac{f(\mathbf{s}'; \mathbf{s}_1)}{\Delta x'} = \tan \varphi' - \tan \varphi + D(\Delta x' - \Delta x) + \frac{f(\mathbf{s}_2; \mathbf{s}_1)}{\Delta x}, \quad (\text{I.2b})$$

$$1 \quad \frac{f(\mathbf{s}_1; \mathbf{s}')}{\Delta x'} = -\tan \varphi' + \tan \varphi + D(\Delta x' - \Delta x) - C(\Delta x' - \Delta x) + \frac{f(\mathbf{s}_1; \mathbf{s}_2)}{\Delta x}, \quad (\text{I.2c})$$

2 and

$$3 \quad \frac{f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}')}{\Delta x} = -(\tan \varphi' - \tan \varphi) - D(\Delta x' - \Delta x). \quad (\text{I.2d})$$

4 When  $\Delta x > 0$ , Eq. (I.1a) gives  $\Delta x' > 0$  and  $\Delta x' - \Delta x > 0$ . As  $D > 0$  and  $C < 0$ , and

5  $f(\mathbf{s}_2; \mathbf{s}_1) > 0$  and  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$ , all inequalities  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) > 0$ ,  $f(\mathbf{s}'; \mathbf{s}_1) > 0$ ,  $f(\mathbf{s}_1; \mathbf{s}') > 0$ ,

6 and  $f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}') < 0$  hold under inequality (I.1b), according to Eqs. (I.2). When  $\Delta x < 0$ , Eq.

7 (I.1a) gives  $\Delta x' < 0$  and  $\Delta x' - \Delta x < 0$ . Thus, in the same manner, all inequalities

8  $f(\mathbf{s}'; \mathbf{s}_1, \mathbf{s}_2) > 0$ ,  $f(\mathbf{s}'; \mathbf{s}_1) > 0$ ,  $f(\mathbf{s}_1; \mathbf{s}') > 0$ , and  $f(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}') < 0$  hold under inequality (I.1b).

9 This completes the proof.

## 10 Proof of Lemma I.2

11 We assume that inequalities (H.2) hold. Inequality (H.2b) can be expressed as

12  $(x_2 - x_1)\delta x = \Delta x(\Delta x' - \Delta x) = \Delta x\Delta x' - \Delta x^2 > 0$ , which gives inequalities (I.1a). In addition,

13 inequalities (I.1a) give  $|\delta x| + |\Delta x| = |\Delta x'|$ . Thus, inequalities (H.2a) and (H.2c) yield inequality

14 (I.1b),

$$\begin{aligned}
 |\tan \varphi' - \tan \varphi| &= \left| \frac{\delta y + \Delta y}{\Delta x'} - \frac{\Delta y}{\Delta x} \right| = \left| \frac{\Delta x \delta y - \delta x \Delta y}{\Delta x' \Delta x} \right| \\
 &< \frac{|\Delta x \delta y| + |\delta x \Delta y|}{\Delta x' \Delta x} \\
 &< \frac{D \delta x^2 |\Delta x| + D |\delta x| \Delta x^2}{\Delta x' \Delta x} \\
 &= \frac{D |\delta x| |\Delta x| (|\delta x| + |\Delta x|)}{\Delta x' \Delta x} \\
 &= D |\delta x| = D |\Delta x' - \Delta x|.
 \end{aligned} \quad (\text{I.3})$$

16 This completes the proof.

## 17 Proof of inequalities (H.2) ensuring inequality (H.3)

18 We assume that inequalities (H.2) hold. In this case, inequality (H.2b) gives  $\Delta x \delta x > 0$ . Then,

19 inequalities (H.2a) and (H.2b) yield inequality (H.3),

$$\begin{aligned}
|\Delta y'| &\leq |\delta y| + |\Delta y| \\
&< D\delta x^2 + D\Delta x^2 \\
&= D(\delta x + \Delta x)^2 - 2D\delta x\Delta x \\
&\leq D(\delta x + \Delta x)^2 = D\Delta x'^2.
\end{aligned} \tag{I.4}$$

2 This completes the proof.

### 3 **Appendix J: Directional evolution sufficient for** 4 **evolutionary branching**

5 Here we derive Eq. (19) in Section 5, specifying the sufficient amount of directional evolution  
6 along  $y$  for the emergence of a dimorphism of  $\mathbf{s}_1$  and  $\mathbf{s}_2$  satisfying  $|\Delta x| = |x_2 - x_1| > \Delta x^*$   
7 for an arbitrary  $\Delta x^* > 0$ . Provided that inequality (18),  $D > 1/\sqrt{2}$ , holds, Lemmas D.1 and  
8 D.2 in Appendix D ensure that

$$\begin{cases} \delta x_{\text{MLI}} > 0 & \text{for } x < 0 \\ \delta x_{\text{MLI}} < 0 & \text{for } x > 0 \end{cases} \tag{J.1a}$$

10 for a monomorphic resident population, and that

$$\begin{cases} \delta x_{\text{MLI}} < 0 & \text{for } \mathbf{s}_{\text{MLI}} = \mathbf{s}_1 \\ \delta x_{\text{MLI}} > 0 & \text{for } \mathbf{s}_{\text{MLI}} = \mathbf{s}_2 \end{cases} \tag{J.1b}$$

12 for a dimorphic resident population, where  $(\delta x_{\text{MLI}}, \delta y_{\text{MLI}})^T = \delta \mathbf{s}_{\text{MLI}} = \mathbf{s}'_{\text{MLI}} - \mathbf{s}_{\text{MLI}}$ ,  $\mathbf{s}_{\text{MLI}}$  is the  
13 parental resident of  $\mathbf{s}'_{\text{MLI}}$ , and  $x_1 < x_2$  is assumed without loss of generality. Also, for both  
14 monomorphic and dimorphic populations, Lemmas D.1 and D.2 ensure that

$$1 < |\delta x_{\text{MLI}}| < \sqrt{2}, \tag{J.2a}$$

$$0 < \delta y_{\text{MLI}} < \frac{1}{\sqrt{2}}. \tag{J.2b}$$

17 Then, under inequalities (J.1) and (J.2), the process of evolutionary branching along an MLIP  
18 amounts to monotonic monomorphic convergence toward  $x = 0$ , followed by monotonic  
19 dimorphic divergence after the emergence of a dimorphism. Thus, if  $x_a$  is negative, this  
20 evolutionary dynamics in the trait space traces the shape of a lower-case letter “y”. By  
21 contrast, if  $x_a$  is positive, the shape is that of a mirrored “y”.

1 The sum of the lengths of all branches and trunks of such a tree shape is given by

$$2 \quad l(\mathbf{I}; \mathbf{s}_a) = \sum_{k=1}^K |\delta \mathbf{s}(k)| = \sum_{k=1}^K |\mathbf{s}'(k) - \mathbf{s}_p(k)|, \quad l(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a) = \sum_{k=1}^K |\delta \mathbf{s}_{\text{MLI}}(k)| = \sum_{k=1}^K |\mathbf{s}'_{\text{MLI}}(k) - \mathbf{s}_{\text{MLI}}(k)|, \quad (\text{J.3})$$

3 where  $\mathbf{s}'(k) = (x'(k), y'(k))^T$ ,  $\mathbf{s}'_{\text{MLI}}(k) = (x'_{\text{MLI}}(k), y'_{\text{MLI}}(k))^T$  is the MLI mutant invading at the  
4  $k$  th invasion event,  $\mathbf{s}_p(k) = (x_p(k), y_p(k))^T$ ,  $\mathbf{s}_{\text{MLI}}(k) = (x_{\text{MLI}}(k), y_{\text{MLI}}(k))^T$  is its parental  
5 resident,  $K$  is the total number of invasion events,

6  $\mathbf{I} = (\mathbf{s}'(1), \dots, \mathbf{s}'(k), \dots, \mathbf{s}'(K))$ ,  $\mathbf{I}_{\text{MLI}} = (\mathbf{s}'_{\text{MLI}}(1), \dots, \mathbf{s}'_{\text{MLI}}(k), \dots, \mathbf{s}'_{\text{MLI}}(K))$ , and  $\mathbf{s}_a$  is the initial resident  
7 phenotype. Eq. (J.3) describes the length of the evolutionary path formed by the mutant-  
8 invasion sequence  $\mathbf{I}_{\text{MLI}}$  from  $\mathbf{s}_a$ . By projecting this evolutionary path onto the  $x$ - and  $y$ -  
9 axes, we obtain its lengths along  $x$  and  $y$  as

$$10 \quad \begin{aligned} l_x(\mathbf{I}; \mathbf{s}_a) &= \sum_{k=1}^K |\delta x(k)| = \sum_{k=1}^K |x'(k) - x_p(k)|, & l_x(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a) &= \sum_{k=1}^K |\delta x_{\text{MLI}}(k)| = \sum_{k=1}^K |x'_{\text{MLI}}(k) - x_{\text{MLI}}(k)|, \\ l_y(\mathbf{I}; \mathbf{s}_a) &= \sum_{k=1}^K |\delta y(k)| = \sum_{k=1}^K |y'(k) - y_p(k)|, & l_y(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a) &= \sum_{k=1}^K |\delta y_{\text{MLI}}(k)| = \sum_{k=1}^K |y'_{\text{MLI}}(k) - y_{\text{MLI}}(k)|, \end{aligned} \quad (\text{J.4})$$

11 respectively. We furthermore decompose  $l_x$  into the portions before and after dimorphic  
12 emergence,

$$13 \quad \begin{aligned} l_x(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a) &= \sum_{k=1}^{k_D-1} |\delta x_{\text{MLI}}(k)| + \sum_{k=k_D}^K |\delta x_{\text{MLI}}(k)| \\ &= l_{\text{xC}} + l_{\text{xD}}, \end{aligned} \quad (\text{J.5})$$

14 where  $\delta x_{\text{MLI}}(k) = x'_{\text{MLI}}(k) - x_{\text{MLI}}(k)$ , and the  $k_D$  th invasion event is assumed to bring about  
15 the emergence of dimorphism.

16 Under inequalities (J.1) and (J.2), the population monotonically converges toward the evo-  
17 lutionary-branching line until it becomes dimorphic, which inevitably occurs before the resi-  
18 dent and mutant straddle the line  $x = 0$ . In other words, the population starts diversification  
19 before the length of its evolutionary path along  $x$  exceeds  $|x_a|$ . Thus,

$$20 \quad l_{\text{xC}} \leq |x_a|. \quad (\text{J.6})$$

21 In addition, monotonic diversification along  $x$  continues after the emergence of  
22 dimorphism, with

$$l_{\text{xD}} = |x_2(K) - x_1(K)|. \quad (\text{J.9})$$

Moreover, as  $|\delta y_{\text{MLI}}|/|\delta x_{\text{MLI}}| < 1/\sqrt{2}$  always holds under ~~inequalities~~ (J.3) and (J.42),

$$l_y(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a) = \sum_{k=1}^K |\delta y_{\text{MLI}}(k)|$$

$$< \frac{1}{\sqrt{2}} \sum_{k=1}^K |\delta x_{\text{MLI}}(k)| = \frac{1}{\sqrt{2}} (l_{\text{xC}} + l_{\text{xD}}) \leq \frac{1}{\sqrt{2}} (|x_a| + |x_2(K) - x_1(K)|) \quad (\text{J.7})$$

holds. In addition, obviously,

$$\bar{y}(K) \leq \max(y_1(K), y_2(K)) \leq l_y(\mathbf{I}_{\text{MLI}}; \mathbf{s}_a) + y_a, \quad (\text{J.8})$$

where  $\bar{y}(K)$  is the mean value of  $y$  after  $K$  invasion events, given by  $\bar{y}(K) = y(K)$  for monomorphism and by  $\bar{y}(K) = (\hat{n}_1(K)y_1(K) + \hat{n}_2(K)y_2(K)) / (\hat{n}_1(K) + \hat{n}_2(K))$  for dimorphism. Substituting inequality (J.8) into inequality (J.7) yields

$$\bar{y}(K) - y_a < \frac{|x_a| + |x_2(K) - x_1(K)|}{\sqrt{2}}. \quad (\text{J.9})$$

Thus, for the emergence of dimorphism with  $|x_2(K) - x_1(K)| > \Delta x^*$ , Eq. (19) in Section 5,

$$\bar{y} - y_a < \frac{|x_a| + \Delta x^*}{\sqrt{2}} \quad (\text{J.10})$$

is the sufficient amount of directional evolution, where  $\bar{y} = \bar{y}(K)$ . Analogously, the sufficient number of invasion events can be derived as  $\lceil \frac{|x_a| + \Delta x^*}{\sqrt{2}} \rceil$ .

## Appendix K: Procedures for the numerical calculation of evolutionary dynamics

Here we explain the procedures for the numerical calculation of the evolutionary dynamics shown in Section 6. These calculations are conducted in a normalized and rescaled trait space such that mutation is isotropic with standard deviation 1. For calculating MLIPs, the MLI mutant at each invasion event is determined so that it maximizes the invasion-event probability density defined in Eq. (9b) in Section 4. For calculating OSIPs, each invading mutant is stochastically chosen according to the invasion-event probability density (see also

1 Ito and Dieckmann, 2007). See Appendix M for the details of how to calculate invasion-event  
2 probability densities. When an invasion has occurred, the coexisting phenotypes at the next  
3 population dynamical equilibrium are determined by checking invasion fitnesses among  
4 residents and the mutant.

5 For each calculation of an MLIP or OSIP, the trait  $x_a$  of the initial resident is drawn  
6 randomly from a uniform distribution with  $-10 \leq x_a \leq 10$ , while the trait  $y_a$  of the initial  
7 resident is set to 0 without loss of generality. For evaluating the occurrence of evolutionary  
8 branching in OSIPs, it is numerically observed that 99.997 percent of failures (i.e., collapse of  
9 protected dimorphisms) occur for  $|\Delta x| < 10$ , where  $\Delta x = x_2 - x_1$  describes the phenotypic  
10 difference in  $x$  between the two residents. Thus, we conclude that evolutionary branching  
11 has occurred when a dimorphism with  $|\Delta x| > 10$  has emerged. Then, the sufficient directional  
12 evolution in  $y$  along MLIPs is given by  $\hat{L}_{y_0} = L_{y_0}(|x_a|, 10) = (|x_a| + 10) / \sqrt{2}$ , according to  
13 inequality (19) in Section 5.

14 We calculate PSIPs using the polymorphic stochastic model (Dieckmann and Law, 1996),  
15 which describes individual births and deaths as stochastic events. For illustration, we use the  
16 birth and death rates defined for the resource-competition model studied by Ito and  
17 Dieckmann (2007), which is a linear combination of the MacArthur-Levins resource-  
18 competition model (MacArthur, 1972) in  $x$  and a constant selection gradient in  $y$ . This  
19 model (detailed in Appendix N) is a simple but ecologically plausible realization of the nor-  
20 mal form for invasion-fitness functions with significant sensitivity difference considered in  
21 this study, as given by Eq. (3). The initial monomorphic phenotype is assigned as described  
22 for OSIPs above. To examine the process of evolutionary branching, phenotypes whose phe-  
23 notypic distance is less than 2 are clustered together. When an initial single cluster splits into  
24 two clusters,  $|\Delta x|$  is calculated as the phenotypic distance between the averages of  $x$  with-  
25 in the two clusters. We conclude that evolutionary branching has occurred when  $|\Delta x|$  ex-  
26 ceeds 10, analogous to the criterion used for OSIPs, as described above.

## 1 Appendix L: Mutation distributions

2 Here we specify three additional mutation distributions used in our numerical calculations,  
3 which are the bivariate fixed-step distribution, the univariate Gaussian distribution, and the  
4 univariate fixed-step distribution.

5 The bivariate fixed-step distribution describes mutations that are limited to an ellipse  
6  $\delta X^2 / \tilde{\sigma}_X^2 + \delta Y^2 / \tilde{\sigma}_Y^2 = 1$ , as illustrated in Fig. 4b5b. Along this ellipse, mutations are distribut-  
7 ed uniformly. The standard deviations of the mutational step sizes, calculated according to Eq.  
8 (1) in Section 3, are  $\sigma_X = \sqrt{V_{XX}} = \tilde{\sigma}_X / \sqrt{2}$  and  $\sigma_Y = \sqrt{V_{YY}} = \tilde{\sigma}_Y / \sqrt{2}$ , with  $V_{XY} = 0$ .

9 The univariate Gaussian distribution describes separate mutations in  $X$  and  $Y$ , with the  
10 corresponding relative mutation rates given by  $\mu_X$  and  $\mu_Y$  ( $\mu_X + \mu_Y = 1$ ), respectively, as  
11 illustrated in Fig. 4e5c. The mutational step sizes follow Gaussian distributions with standard  
12 deviations  $\tilde{\sigma}_X$  and  $\tilde{\sigma}_Y$ , respectively. The standard deviations of mutational step sizes,  
13 calculated according to Eq. (1) in Section 3, are  $\sigma_X = \sqrt{\mu_X} \tilde{\sigma}_X$  and  $\sigma_Y = \sqrt{\mu_Y} \tilde{\sigma}_Y$ , with  
14  $V_{XY} = 0$ .

15 The univariate fixed-step distribution describes separate mutations in  $X$  and  $Y$ , with the  
16 corresponding relative mutation rates given by  $\mu_X$  and  $\mu_Y$  ( $\mu_X + \mu_Y = 1$ ), respectively.  
17 The only possible mutations are  $\delta \mathbf{S} = (\tilde{\sigma}_X, 0)^T$ ,  $(-\tilde{\sigma}_X, 0)^T$ ,  $(0, \tilde{\sigma}_Y)^T$ , and  $(0, -\tilde{\sigma}_Y)^T$ , which  
18 occur with probabilities  $\mu_X / 2$ ,  $\mu_X / 2$ ,  $\mu_Y / 2$ , and  $\mu_Y / 2$ , as illustrated in Fig. 5d. The  
19 standard deviations of mutational step sizes, calculated according to Eq. (1) in Section 3, are  
20  $\sigma_X = \sqrt{\mu_X / (\mu_X + \mu_Y)} \tilde{\sigma}_X$  and  $\sigma_Y = \sqrt{\mu_Y / (\mu_X + \mu_Y)} \tilde{\sigma}_Y$ , with  $V_{XY} = 0$ .

21 After normalizing and rescaling the trait space  $\mathbf{S} = (X, Y)^T$  into  $\mathbf{s} = (x, y)^T$  such that the  
22 standard deviations of mutational step sizes become equal to 1 in all directions, the MLIP  
23 condition is applied. MLIP conditions for the additional three mutation distributions described  
24 above are applied by using the  $\sigma = \sigma_X = \sigma_Y$  for these distributions in Eqs. (12c) and (18).  
25 For clarity, here we refer to these conditions as approximate MLIP conditions (as they were  
26 derived for bivariate Gaussian distributions, but are now applied to other distributions).

27 Alternatively, MLIP conditions for the three other distributions can be derived directly by  
28 using these mutation distributions in the invasion-event probability function, analogously to  
29 the derivation for bivariate Gaussian distribution, as defined by Eq. (15) in Section 5. If the

1  $X$  - and  $Y$  -axes correspond to the sensitive and insensitive directions, respectively, the  
2 resultant exact conditions are obtained as  $(\tilde{\sigma}_X^2 D_{XX}) / (2\tilde{\sigma}_Y |G_Y|) > \sqrt{3} / 2$ ,  
3  $(\mu_X \tilde{\sigma}_X^2 D_{XX}) / (2\mu_Y \tilde{\sigma}_Y |G_Y|) > \sqrt{e} / 2$ , and  $(\mu_X \tilde{\sigma}_X^2 D_{XX}) / (2\mu_Y \tilde{\sigma}_Y |G_Y|) > 1$  for the bivariate  
4 fixed-step, univariate Gaussian, and univariate fixed-step distributions, respectively. These  
5 results can be compared with  $D = (\sigma_X^2 D_{XX}) / (2\sigma_Y |G_Y|) > 1 / \sqrt{2}$  for the bivariate Gaussian  
6 distribution.

7 As it turns out, the exact MLIP conditions tend to overestimate the likelihood of  
8 evolutionary branching in OSIPs and PSIPs when univariate fixed-step or univariate  
9 Gaussian mutations with large  $\mu_X / \mu_Y$  are considered (results not shown). Thus, using the  
10 approximate MLIP conditions seems to be more robust than using the exact ones.

## 11 **Appendix M: Calculation of invasion-event probability** 12 **densities in MLIPs and OSIPs**

### 13 **Necessary elements for calculation**

14 Here we explain how invasion-event probability densities, as defined by Eq. (9b) in Section 4,  
15 are determined in the calculation of MLIPs and OSIPs. For any given composition of resident  
16 phenotypes, their equilibrium frequencies and the invasion fitness of possible mutants are  
17 required for the calculation (absolute population sizes are needed only for determining the  
18 waiting time for each invasion event, which is not needed for the numerical results we present  
19 in this study).

20 As for invasion-fitness functions, we have those for monomorphism, Eq. (12a) in Section 5,  
21 and for dimorphism, Eq. (D.12) in Appendix D. Invasion-fitness functions for higher degrees  
22 of polymorphism are not needed, because trimorphism is impossible under the dimorphic  
23 invasion-fitness function, Eq. (D.12). Thus, the monomorphic and dimorphic invasion-fitness  
24 functions are sufficient for the calculation of invasion-event probability densities.

25 As for equilibrium frequencies of resident phenotypes, the frequency of a monomorphic  
26 resident is of course always 1. For a dimorphism of  $s_1$  and  $s_2$ , the corresponding



1 frequencies are approximately given by  $q_1 = f(\mathbf{s}_1; \mathbf{s}_2) / [f(\mathbf{s}_1; \mathbf{s}_2) + f(\mathbf{s}_2; \mathbf{s}_1)]$  and  $q_2 = 1 - q_1$ ,  
 2 as explained in the next subsection.

### 3 **Equilibrium frequencies of dimorphic phenotypes**

4 Here we approximate the equilibrium frequencies of the dimorphic residents  $\mathbf{s}_1$  and  $\mathbf{s}_2$ .  
 5 Without loss of generality, we consider a normalized but ~~not~~non-scaled trait space, so that  
 6 mutation is isotropic with standard deviation  $\sigma \ll 1$ .

7 As explained in Appendix E, the population dynamics of  $\mathbf{s}_1$  and  $\mathbf{s}_2$  can be approximated  
 8 by Lotka-Volterra equations,

$$\begin{aligned}
 \frac{1}{n_1} \frac{dn_1}{dt} &= r_1 \left( 1 - \frac{n_1 + \alpha_{12} n_2}{K_1} \right), \\
 \frac{1}{n_2} \frac{dn_2}{dt} &= r_2 \left( 1 - \frac{\alpha_{21} n_1 + n_2}{K_2} \right),
 \end{aligned}
 \tag{M.1}$$

10 where  $r_2 - r_1$ ,  $\alpha_{21} - \alpha_{12}$ , and  $K_2 - K_1$  are  $O(|\mathbf{s}_2 - \mathbf{s}_1|)$ . The corresponding equilibrium  
 11 population sizes are given by

$$\begin{aligned}
 \hat{n}_1 &= \frac{K_1 - \alpha_{12} K_2}{1 - \alpha_{12} \alpha_{21}}, \\
 \hat{n}_2 &= \frac{K_2 - \alpha_{21} K_1}{1 - \alpha_{12} \alpha_{21}},
 \end{aligned}
 \tag{M.2}$$

13 while  $f(\mathbf{s}_1; \mathbf{s}_2)$  and  $f(\mathbf{s}_2; \mathbf{s}_1)$  are given by

$$\begin{aligned}
 f(\mathbf{s}_1; \mathbf{s}_2) &= \frac{r_1}{K_1} [K_1 - \alpha_{12} K_2], \\
 f(\mathbf{s}_2; \mathbf{s}_1) &= \frac{r_2}{K_2} [K_2 - \alpha_{21} K_1].
 \end{aligned}
 \tag{M.3}$$

15 Then, the following relationship holds,

$$\begin{aligned}
\frac{\hat{n}_1}{\hat{n}_1 + \hat{n}_2} &= \frac{f(\mathbf{s}_1; \mathbf{s}_2)}{f(\mathbf{s}_1; \mathbf{s}_2) + \frac{r_1 K_2}{r_2 K_1} f(\mathbf{s}_2; \mathbf{s}_1)} \\
&= \frac{f(\mathbf{s}_1; \mathbf{s}_2)}{f(\mathbf{s}_1; \mathbf{s}_2) + (1 + \varepsilon) f(\mathbf{s}_2; \mathbf{s}_1)} \\
&= \frac{f(\mathbf{s}_1; \mathbf{s}_2)}{f(\mathbf{s}_1; \mathbf{s}_2) + f(\mathbf{s}_2; \mathbf{s}_1)} \frac{1}{1 + \varepsilon \frac{f(\mathbf{s}_2; \mathbf{s}_1)}{f(\mathbf{s}_1; \mathbf{s}_2) + f(\mathbf{s}_2; \mathbf{s}_1)}} \\
&= w_1 \left[ 1 - \varepsilon w_2 + O(\varepsilon^2) \right] \\
&= w_1 - \varepsilon w_1 w_2 + O(\varepsilon^2),
\end{aligned} \tag{M.4}$$

2 where  $w_1 = f(\mathbf{s}_1; \mathbf{s}_2) / [f(\mathbf{s}_1; \mathbf{s}_2) + f(\mathbf{s}_2; \mathbf{s}_1)]$ ,  $w_2 = 1 - w_1$ , and  $\varepsilon = (r_1 K_2) / (r_2 K_1) - 1$   
3  $= O(|\mathbf{s}_2 - \mathbf{s}_1|)$ . Notice that  $f(\mathbf{s}_1; \mathbf{s}_2) > 0$  and  $f(\mathbf{s}_2; \mathbf{s}_1) > 0$  both hold for a protected  
4 dimorphism of  $\mathbf{s}_1$  and  $\mathbf{s}_2$ . Thus, as long as  $|\mathbf{s}_2 - \mathbf{s}_1| = O(\sigma)$ ,

$$\frac{\hat{n}_1}{\hat{n}_1 + \hat{n}_2} = w_1 + O(\sigma) \tag{M.5}$$

6 is a good approximation. Therefore, after rescaling this trait space such that the standard  
7 deviation of mutational step sizes is equal to 1, Eq. (M.5) is a good approximation as long as  
8  $|\mathbf{s}_2 - \mathbf{s}_1|$  is of order 1.

## 9 **Appendix N: Specific model for calculation of PSIPs**

10 Here we explain the model used for the calculation of PSIPs in Section 6. We consider a  
11 normalized bivariate trait space  $\mathbf{s} = (x, y)^T$ , in which mutation is isotropic with standard  
12 deviation  $\sigma \ll 1$ . We define individual birth and death rates following Ito and Dieckmann  
13 (2007), as explained below. The trait  $x$  affects the death rate through resource competition  
14 (as, e.g., when beak size in birds determines the size of seeds they compete for). The death  
15 rate  $d(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N)$  of phenotype  $\mathbf{s}_i = (x_i, y_i)^T$  depends on the trait values  $x_i$  and  $x_j$ , as  
16 well as on the abundances  $n_j$  of extant phenotypes  $j = 1, \dots, N$ ,

$$d(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N) = \sum_j \frac{\alpha(x_j - x_i) n_j}{K(x_i)}. \tag{N.1}$$

18 Here

$$K(x_i) = K_0 \exp\left(-\frac{1}{2} x_i^2 / \sigma_K^2\right) \tag{N.2}$$

1 is the carrying capacity of phenotype  $s_i$ , given by a Gaussian function with variance  $\sigma_K^2$   
 2 and mean  $x_i = 0$ . The function

$$3 \quad \alpha(x_j - x_i) = \exp\left(-\frac{1}{2}(x_j - x_i)^2 / \sigma_a^2\right) \quad (\text{N.3})$$

4 describes the strength of competition between phenotype  $x_i$  and phenotype  $x_j$ ; it is also  
 5 given by a Gaussian function, with variance  $\sigma_a^2$  and mean  $x_j - x_i = 0$ . Accordingly, the  
 6 strength of competition is maximal between identical phenotypes and monotonically declines  
 7 with phenotypic distance. If the birth rate  $b(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N)$  is assumed to be constant and equal  
 8 to 1, the birth and death rates imply the MacArthur-Levins resource-competition model  
 9 (MacArthur, 1972),  $\frac{d}{dt}n_i = [b(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N) - d(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N)]n_i$  for  $i = 1, \dots, N$ , in the limit of  
 10 infinite population size.

11 Directional selection on  $y$  can be due to any ecological interaction (e.g., competition,  
 12 exploitation, or mutualism) and may act on any morphological, physiological, or life-history  
 13 trait  $y$ . A simple way of introducing a fitness gradient in  $y$  is

$$14 \quad b(\mathbf{s}_i; \mathbf{s}_1, \dots, \mathbf{s}_N) = 1 + b_1(y_i - \bar{y}), \quad (\text{N.4})$$

15 where  $\bar{y}$  denotes the population average of trait value  $y$ ,  $\bar{y} = \sum_j y_j n_j / \sum_j n_j$ , and  $b_1$  is  
 16 a constant describing the constant directional selection pressure on  $y$ .

17 We now consider a monomorphic resident  $\mathbf{s}$  with sufficiently large population size. The  
 18 invasion fitness of  $\mathbf{s}'$  with respect to  $\mathbf{s}$  is given by

$$19 \quad \begin{aligned} f(\mathbf{s}'; \mathbf{s}) &= b(\mathbf{s}'; \mathbf{s}) - d(\mathbf{s}'; \mathbf{s}) \\ &= 1 - \frac{\alpha(x' - x)\hat{n}}{K(x')} + b_1(y' - y), \end{aligned} \quad (\text{N.5})$$

20 where  $\hat{n} = K(x)$  is the equilibrium population size of  $\mathbf{s}$ . Since trait  $y$  contributes to  
 21 invasion fitness only through the linear term  $b_1(y' - y)$ , the condition for significant  
 22 sensitivity difference, Eq. (4) in Section 3, is immediately satisfied whenever  $b_1$  is  
 23 sufficiently small such that  $G_y = b_1 = O(\sigma)$ . In this case, the invasion fitness can be expanded  
 24 around  $\mathbf{s}_0 = (0, y_0)^T$  with arbitrary  $y_0$ , in the form of Eq. (3) in Section 3,

$$25 \quad f(\mathbf{s}'; \mathbf{s}) = G_x \delta x + C_{xx} x \delta x + \frac{1}{2} D_{xx} \delta x^2 + G_y \delta y, \quad (\text{N.6})$$

1 where

$$\begin{aligned}
 G_x &= \left. \frac{\partial f(\mathbf{s}'; \mathbf{s})}{\partial x'} \right|_{\mathbf{s}'=\mathbf{s}=\mathbf{s}_0} = 0, \\
 G_y &= \left. \frac{\partial f(\mathbf{s}'; \mathbf{s})}{\partial y'} \right|_{\mathbf{s}'=\mathbf{s}=\mathbf{s}_0} = b_1,
 \end{aligned}
 \tag{N.7}$$

3 and

$$\begin{aligned}
 C_{xx} &= \left. \frac{\partial^2 f(\mathbf{s}'; \mathbf{s})}{\partial x'^2} \right|_{\mathbf{s}'=\mathbf{s}=\mathbf{s}_0} + \left. \frac{\partial^2 f(\mathbf{s}'; \mathbf{s})}{\partial x \partial x'} \right|_{\mathbf{s}'=\mathbf{s}=\mathbf{s}_0} = -\frac{1}{\sigma_K^2}, \\
 D_{xx} &= \left. \frac{\partial^2 f(\mathbf{s}'; \mathbf{s})}{\partial x'^2} \right|_{\mathbf{s}'=\mathbf{s}=\mathbf{s}_0} = \frac{1}{\sigma_a^2} - \frac{1}{\sigma_K^2}.
 \end{aligned}
 \tag{N.8}$$

5 Since  $G_x = 0$  and  $C_{xx} < 0$  hold,  $x = 0$  is a convergence-stable line. Then,  $C$  and  $D$ , in  
 6 the rescaled trait space, are given by

$$\begin{aligned}
 C &= \frac{\sigma C_{xx}}{|G_y|} = -\frac{\sigma}{|b_1| \sigma_K^2}, \\
 D &= \frac{\sigma D_{xx}}{2|G_y|} = \frac{\sigma}{2|b_1|} \left( \frac{1}{\sigma_a^2} - \frac{1}{\sigma_K^2} \right).
 \end{aligned}
 \tag{N.9}$$

8 As  $C$  is always negative, the convergence-stable line is an evolutionary-branching line if the  
 9 MLIP condition,  $D > 1/\sqrt{2}$ , holds.

## 10 **Appendix O: Extension of conditions for evolutionary-** 11 **branching lines to multivariate trait spaces**

12 Here we extend the conditions for evolutionary-branching lines – Eqs. (4), (6a), (6b), and (18)  
 13 – to multivariate trait spaces. We consider an arbitrary  $L$ -variate trait space  $\mathbf{S} = (U_1, \dots, U_L)^T$   
 14 with a mutational variance-covariance matrix  $\Lambda$ , which has real and non-negative  
 15 eigenvalues,  $\sigma_1^2, \dots, \sigma_L^2$ . The maximum eigenvalue of  $\Lambda$ , denoted by  $\sigma_1^2$ , gives the  
 16 maximum standard deviation  $\sigma_1$  of mutational step sizes among all directions. The trait  
 17 space is normalized by an affine coordinate transformation of  $\mathbf{S}$  into  $\mathbf{s} = (u_1, \dots, u_L)^T$  with  
 18 isotropic mutation  $\sigma = \sigma_1$  (see Appendix P).

1 The invasion fitness in the normalized space can be written analogously to the bivariate  
 2 case (Appendix A) as

$$3 \quad f(\mathbf{s}'; \mathbf{s}) = \mathbf{G}\delta\mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C}\delta\mathbf{s} + \frac{1}{2}\delta\mathbf{s}^T \mathbf{D}\delta\mathbf{s} + O(\sigma^3), \quad (\text{O.1})$$

4 where  $\mathbf{G}$  is a row vector of length  $L$ , and  $\mathbf{C}$  and  $\mathbf{D}$  are  $L \times L$  matrices, with  $\mathbf{D}$  being  
 5 symmetric. In a manner similar to the bivariate case, we can define

6 **Significant sensitivity difference in multivariate trait spaces:** After normalization  
 7 to make mutation isotropic, the normalized invasion-fitness function, Eq. (O.1), can be  
 8 made to satisfy

$$9 \quad \frac{|G_j| + |C_{ij}| + |C_{ji}| + |C_{jj}| + |D_{ij}| + |D_{jj}|}{|G_i| + |C_{ii}| + |D_{ii}|} = O(\sigma) \quad (\text{O.2})$$

10 for all  $i = 1, \dots, \hat{L}$  and all  $j = \hat{L} + 1, \dots, L$ , by rotating the axes, where  $\hat{L} < L$ ,  $G_j$  is the  
 11  $j$ th component of  $\mathbf{G}$ , and  $C_{ij}$  and  $D_{ij}$  are the  $(i, j)$  components of  $\mathbf{C}$  and  $\mathbf{D}$ , re-  
 12 spectively.

13 If the invasion-fitness function, Eq. (O.1), has significant sensitivity difference, the trait space  
 14 can be decomposed into an  $\hat{L}$ -variate sensitive subspace  $\mathbf{x} = (x_1, \dots, x_{\hat{L}})^T = (u_1, \dots, u_{\hat{L}})^T$  and an  
 15  $\tilde{L} (= L - \hat{L})$ -variate insensitive subspace  $\mathbf{y} = (y_1, \dots, y_{\tilde{L}})^T = (u_{\hat{L}+1}, \dots, u_L)^T$ . Notice that Eq. (O.2)  
 16 allows decomposition of Eq. (O.1) into  $f(\mathbf{s}'; \mathbf{s}) = g(\mathbf{x}'; \mathbf{x}) + \varepsilon h(\mathbf{s}'; \mathbf{s})$  with a small  $\varepsilon = O(\sigma)$ ,  
 17 while  $h(\mathbf{s}'; \mathbf{s})$  is kept smooth and finite, as in the bivariate case (Appendix B). In this case,  
 18 Eq. (O.1) is transformed into

$$19 \quad \begin{aligned} f(\mathbf{s}'; \mathbf{s}) &= \sum_{i=1}^{\hat{L}} G_i \delta x_i + \sum_{i=1}^{\hat{L}} \sum_{i'=1}^{\hat{L}} C_{ii'} (x_{i'} - x_{0i'}) \delta x_i \\ &\quad + \frac{1}{2} \sum_{i=1}^{\hat{L}} \sum_{i'=1}^{\hat{L}} D_{ii'} \delta x_i \delta x_{i'} + \sum_{j=\hat{L}+1}^L G_j \delta y_{j-\hat{L}} + O(\sigma^3) \\ &= \mathbf{G}_x \delta \mathbf{x} + (\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_{xx} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{x}^T \mathbf{D}_{xx} \delta \mathbf{x} + \mathbf{G}_y \delta \mathbf{y} + O(\sigma^3) \\ &= \mathbf{G}_x \delta \mathbf{x} + (\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_{xx} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{x}^T \mathbf{D}_{xx} \delta \mathbf{x} + |\mathbf{G}_y| \delta y + O(\sigma^3), \end{aligned} \quad (\text{O.3})$$

20 where  $\delta y = (\mathbf{G}_y / |\mathbf{G}_y|) \delta \mathbf{y}$  is the element of  $\delta \mathbf{y}$  parallel to the fitness gradient  $\mathbf{G}_y$  in the  
 21 insensitive subspace. Notice that the insensitive subspace contributes to the invasion fitness

1 above only through this element  $\delta y$ . Thus, the dimensionality of the local evolutionary  
 2 dynamics can be contracted to  $\hat{L}+1$ . If the sensitive subspace is univariate, then  $\mathbf{G}_x$ ,  $\mathbf{C}_{xx}$ ,  
 3 and  $\mathbf{D}_{xx}$  become scalars, in which case Eq. (O.3) yields

$$4 \quad f(\mathbf{s}'; \mathbf{s}) = G_x \delta x + C_{xx} (x - x_0) \delta x + \frac{1}{2} D_{xx} \delta x^2 + |\mathbf{G}_y| \delta y + O(\sigma^3). \quad (\text{O.4})$$

5 Thus, the local evolutionary dynamics ~~around~~  $\mathbf{s}_0$  can be contracted into that in a  
 6 bivariate trait space  $(x, y)^T$ . Then, by denoting  $|\mathbf{G}_y|$  by  $|G_y|$ , the conditions for  
 7 evolutionary-branching lines in bivariate trait species can be applied ~~as they are,~~  
 8 ~~providing directly,~~ yielding Theorem 3 in Section 7.

9 On the other hand, when the sensitive subspace is more than univariate, the MLIP condition  
 10 cannot be applied. Yet, the following considerations apply. For  $|\mathbf{G}_y| = 0$ , it is expected that an  
 11 evolutionarily singular point  $\mathbf{x}_b$  in the sensitive subspace (i.e.,  $|\mathbf{G}_x| = 0$  for  $\mathbf{x}_0 = \mathbf{x}_b$ ) will  
 12 attract a monomorphic population and induce its evolutionary branching, if the following  
 13 conditions hold. First,  $\mathbf{x}_b$  is strongly convergence stable, i.e.,

$$14 \quad \lambda_{C_i} < 0, \quad (\text{O.5})$$

15 for all  $i = 1, \dots, \hat{L}$ , where  $\lambda_{C_1}, \dots, \lambda_{C_{\hat{L}}}$  are the eigenvalues of  $\mathbf{C}_{xx}$  (Leimar, 2008). Second,  
 16  $\mathbf{x}_b$  is also evolutionarily unstable, i.e.,

$$17 \quad \lambda_{D_{\max}} > 0. \quad (\text{O.6})$$

18 Here  $\lambda_{D_{\max}}$  is the maximum eigenvalue of  $\mathbf{D}_{xx}$ , with eigenvector  $\mathbf{v}_{D_{\max}}$ . This  $\lambda_{D_{\max}}$  is  
 19 always real, as  $\mathbf{D}_{xx}$  is a symmetric matrix. The inequality above means that the fitness  
 20 landscape has a positive curvature (i.e., ~~second derivative~~) in the direction of  $\mathbf{v}_{D_{\max}}$  with  
 21 selection favoring evolutionary diversification mainly in this direction. As mutual invasibility  
 22 in this direction is also ensured in this case, i.e.,  $\mathbf{v}_{D_{\max}}^T (\mathbf{D}_{xx} - \mathbf{C}_{xx}) \mathbf{v}_{D_{\max}} > 0$  always holds  
 23 under inequalities (O.5) and (O.6), dimorphic divergence may proceed without collapse,  
 24 resulting in evolutionary branching (Dieckmann and Metz, in preparation). In this case, it is  
 25 expected that the diversifying residents stay in the neighborhood of the line  $\mathbf{x}_b + \alpha \mathbf{v}_{D_{\max}}$   
 26 (with real parameter  $\alpha$ ). Thus, if the  $x$ -axis is chosen as  $x = \mathbf{v}_{D_{\max}}^T (\mathbf{x} - \mathbf{x}_b)$ , then the  
 27 condition

$$D = \frac{\sigma \lambda_{D_{\max}}}{2|\mathbf{G}_y|} > \frac{1}{\sqrt{2}} \quad (\text{O.7})$$

and inequalities (O.5) together may be a useful indicator of the likelihood of evolutionary branching.

In summary, Eq. (O.2) allows a multivariate trait space to be decomposed into a sensitive subspace and an insensitive one; if the sensitive subspace is univariate, the situation is reduced to a bivariate one, in which case the MLIP condition is applicable. On the other hand, if the sensitive subspace is ~~more than univariate~~ multivariate, there is no assurance of the validity of the MLIP condition. Yet, inequalities (O.5) and (O.7) may still be useful.

## Appendix P: Checking conditions for evolutionary-branching lines without prior normalization

Here we explain how conditions for evolutionary-branching lines can be checked without the prior normalization of trait spaces. We consider a non-normalized  $L$ -variate trait space  $\mathbf{S} = (U_1, \dots, U_L)^T$  that has a mutation variance-covariance matrix  $\Lambda$ , which is diagonalized as

$$\Lambda = \mathbf{P}_\Lambda \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \sigma_L^2 \end{pmatrix} \mathbf{P}_\Lambda^{-1}, \quad (\text{P.1})$$

where  $\mathbf{P}_\Lambda$  is an orthogonal matrix,  $\mathbf{P}_\Lambda^{-1} = \mathbf{P}_\Lambda^T$ , and the square roots of all eigenvalues  $\sigma_1, \dots, \sigma_L$  are positive. ~~The~~ Without loss of generality,  $\sigma_1$  is assumed to be the largest ~~one~~ without loss of generality eigenvalue.

In this trait space, the invasion-fitness function can be written analogously to the bivariate case (Appendix A) as

$$F(\mathbf{S}'; \mathbf{S}) = \check{\mathbf{G}} \delta \mathbf{S} + (\mathbf{S} - \mathbf{S}_0)^T \check{\mathbf{C}} \delta \mathbf{S} + \frac{1}{2} \delta \mathbf{S}^T \check{\mathbf{D}} \delta \mathbf{S} + O(\sigma_1^3), \quad (\text{P.2})$$

where  $\check{\mathbf{G}} = \mathbf{F}_m$ ,  $\check{\mathbf{C}} = \mathbf{F}_{mm} + \mathbf{F}_{rm}$ , and  $\check{\mathbf{D}} = \mathbf{F}_{mm}$ , with the subscripts  $m$  and  $r$  corresponding to the mutant  $\mathbf{S}'$  and the resident  $\mathbf{S}$ , respectively, and where  $\mathbf{F}_m = \partial F(\mathbf{S}'; \mathbf{S}) / \partial \mathbf{S}' \big|_{\mathbf{S}' = \mathbf{S} = \mathbf{S}_0}$  is the gradient row vector at  $\mathbf{S}_0$ , while  $\mathbf{F}_{mm} = \partial^2 F(\mathbf{S}'; \mathbf{S}) / \partial \mathbf{S}'^2 \big|_{\mathbf{S}' = \mathbf{S} = \mathbf{S}_0}$  and  $\mathbf{F}_{rm} = \partial^2 F(\mathbf{S}'; \mathbf{S}) / \partial \mathbf{S}' \partial \mathbf{S} \big|_{\mathbf{S}' = \mathbf{S} = \mathbf{S}_0}$  are ~~the~~ Hessian matrices.

1 This trait space can first be normalized to yield isotropic mutation with standard deviation  
 2  $\sigma \ll 1$ , where  $\sigma$  otherwise is arbitrary, by an affine transformation to coordinates  $\mathbf{s}$ ,

$$3 \quad \mathbf{S} = \mathbf{Q}^T \mathbf{s}, \quad (\text{P.3a})$$

4 with

$$5 \quad \mathbf{Q} = \frac{1}{\sigma} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \sigma_L \end{pmatrix} \mathbf{P}_\Lambda^T, \quad (\text{P.3b})$$

6 because  $\Lambda = (\sigma \mathbf{Q})^T (\sigma \mathbf{Q})$  gives  $\delta \mathbf{S}^T \Lambda^{-1} \delta \mathbf{S} = [\mathbf{Q}^T \delta \mathbf{s}]^T [\sigma^2 \mathbf{Q}^T \mathbf{Q}]^{-1} [\mathbf{Q}^T \delta \mathbf{s}] = |\delta \mathbf{s}|^2 / \sigma^2$  (see  
 7 also Appendix Q). (Notice that the matrix  $\mathbf{Q}$  can be interpreted as the Cholesky  
 8 decomposition of  $\Lambda / \sigma_1^2$ .) If  $\sigma_1 \ll 1$ , then it is natural to choose  $\sigma = \sigma_1$ .

9 Next, to adjust the axes of  $\mathbf{s} = (u_1, \dots, u_L)^T$  so as to maximize the sensitivity difference-  
 10 ~~between these two traits~~, we add a coordinate rotation to Eq. (P.3),

$$11 \quad \mathbf{S} = \mathbf{Q}^T \mathbf{B} \mathbf{s}, \quad (\text{P.4})$$

12 with an orthogonal matrix  $\mathbf{B}$  for describing the rotation. Substituting this equation into Eq.  
 13 (P.2) yields

$$14 \quad f(\mathbf{s}'; \mathbf{s}) = F(\mathbf{Q} \mathbf{B} \mathbf{s}'; \mathbf{Q} \mathbf{B} \mathbf{s}) = \mathbf{G} \delta \mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C} \delta \mathbf{s} + \frac{1}{2} \delta \mathbf{s}^T \mathbf{D} \delta \mathbf{s} + O(\sigma^3), \quad (\text{P.5a})$$

15 where

$$16 \quad \begin{aligned} \mathbf{G} &= \check{\mathbf{G}} \mathbf{Q}^T \mathbf{B}, \\ \mathbf{C} &= \mathbf{B}^T \mathbf{Q} \check{\mathbf{C}} \mathbf{Q}^T \mathbf{B}, \\ \mathbf{D} &= \mathbf{B}^T \mathbf{Q} \check{\mathbf{D}} \mathbf{Q}^T \mathbf{B}, \end{aligned} \quad (\text{P.5b})$$

17 which are Eqs. (23) in Section 7.

18 Finally, we explain how to rotate the axes of  $\mathbf{s} = (u_1, \dots, u_L)^T$  such that the condition for  
 19 significant sensitivity difference, Eq. (21) with  $\hat{L} = 1$ ,

$$20 \quad \frac{|G_j| + |C_{1j}| + |C_{j1}| + |C_{jj}| + |D_{1j}| + |D_{jj}|}{|G_1| + |C_{11}| + |D_{11}|} = O(\sigma) \quad (\text{P.6})$$



1 for all  $j = 2, \dots, L$ , may hold ( $L = 2$  gives the condition for significant sensitivity difference  
2 for the bivariate case, Eq. (4)). The exact approach is to find the matrix  $\mathbf{B}^*$  that minimizes  
3 the left-hand side of Eq. (P.6). However, this  $\mathbf{B}^*$  may not be easy to determine. Fortunately,  
4 as explained below,  $\mathbf{B}^*$  can be obtained approximately without such minimization when the  
5 MLIP condition, inequality (18) in Section 5, holds. In particular, when the condition for sig-  
6 nificant sensitivity difference and the MLIP condition both hold, all components of  $\mathbf{D}$  are  
7  $O(\sigma)$ , except for  $D_{11}$ , which is of order  $\sigma^0$ . In this case, when the trait space is bivariate  
8 ( $L = 2$ ), each of the two eigenvectors of  $\mathbf{D}$  is almost parallel to one axis of the trait space,  
9 ~~wherewith~~ with the corresponding eigenvalues satisfying  $\tilde{\lambda}_{D1} = D_{11} + O(\sigma)$  and  $\tilde{\lambda}_{D2} = O(\sigma)$ . In  
10 other words,  $\mathbf{B}^*$  approximately diagonalizes the symmetric matrix  $\mathbf{Q}\check{\mathbf{D}}\mathbf{Q}^T$ . In this case,  
11  $\mathbf{B}^*$  is approximately obtained by requiring that  $\mathbf{B}^{*T}\mathbf{Q}\check{\mathbf{D}}\mathbf{Q}^T\mathbf{B}^*$  becomes diagonal, i.e.,

$$12 \quad \mathbf{B}^* \approx (\tilde{\mathbf{v}}_{D1}, \tilde{\mathbf{v}}_{D2}), \quad (\text{P.7})$$

13 where  $\tilde{\mathbf{v}}_{D1}$  and  $\tilde{\mathbf{v}}_{D2}$  are the eigenvectors of  $\mathbf{Q}\check{\mathbf{D}}\mathbf{Q}^T$ .

14 Analogously, when the trait space is more than bivariate ( $L > 2$ ),  $\mathbf{B}^*$  is approximately ob-  
15 tained as in Eq. (24) in Section 7,

$$16 \quad \mathbf{B}^* \approx (\tilde{\mathbf{v}}_{D1}, \dots, \tilde{\mathbf{v}}_{DL}), \quad (\text{P.8})$$

17 where  $\tilde{\mathbf{v}}_{D1}, \dots, \tilde{\mathbf{v}}_{DL}$  are the eigenvectors of  $\mathbf{Q}\check{\mathbf{D}}\mathbf{Q}^T$ , corresponding to the eigenvalues  
18  $\tilde{\lambda}_{D1}, \dots, \tilde{\lambda}_{DL}$ , where the eigenvalues are assumed to be ordered such that  $\tilde{\lambda}_{D1} > \tilde{\lambda}_{Dj}$  for all  
19  $j = 2, \dots, L$ , without loss of generality. (When  $\tilde{\mathbf{v}}_{D2}, \dots, \tilde{\mathbf{v}}_{DL}$  are difficult to obtain, possibly  
20 because of too small  $\tilde{\lambda}_{D2}, \dots, \tilde{\lambda}_{DL}$ , those vectors can be chosen ~~arbitrary~~ arbitrarily, as long as  
21  $\tilde{\mathbf{v}}_{D1}, \dots, \tilde{\mathbf{v}}_{DL}$  form an orthogonal coordinate system.) If conditions for evolutionary-branching  
22 lines hold for this  $\mathbf{B}^*$ , an evolutionary-branching line passes through  $\mathbf{S}_0$ .

## 23 Appendix Q: Directional evolution in MLIPs

24 Here we derive MLIPs of monomorphic populations when directional selection is the  
25 dominant selection pressure. We consider an arbitrary  $L$ -variate trait space  $\mathbf{S} = (U_1, \dots, U_L)^T$   
26 and a monomorphic population with phenotype  $\mathbf{S}$ . The ~~invasion~~ invasion-fitness function is  
27 written as Eq. (P.2) in Appendix P,

$$1 \quad F(\mathbf{S}'; \mathbf{S}) = \tilde{\mathbf{G}}\delta\mathbf{S} + (\mathbf{S} - \mathbf{S}_0)^T \tilde{\mathbf{C}}\delta\mathbf{S} + \frac{1}{2}\delta\mathbf{S}^T \tilde{\mathbf{D}}\delta\mathbf{S} + \text{h.o.t.} \quad (\text{Q.1})$$

2 Here we assume a mutation distribution given by a multivariate Gaussian function,

$$3 \quad M(\delta\mathbf{S}) = \frac{1}{(2\pi)^{L/2} |\mathbf{\Lambda}|^{1/2}} \exp\left(-\frac{1}{2}\delta\mathbf{S}^T \mathbf{\Lambda}^{-1}\delta\mathbf{S}\right), \quad (\text{Q.2})$$

4 where  $\mathbf{\Lambda}$  is an  $L \times L$  variance-covariance matrix. By normalizing this trait space using  
5  $\mathbf{S} = \mathbf{Q}^T \mathbf{s}$  in Eq. (P.3a) in Appendix P, the argument of the exponential function in Eq. (Q.2) is  
6 transformed into

$$\begin{aligned} 7 \quad -\frac{1}{2}\delta\mathbf{S}^T \mathbf{\Lambda}^{-1}\delta\mathbf{S} &= -\frac{1}{2}\delta\mathbf{s}^T \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^T \delta\mathbf{s} \\ &= -\frac{1}{2\sigma^2} \delta\mathbf{s}^T \mathbf{Q}\mathbf{Q}^{-1}(\mathbf{Q}^T)^{-1}\mathbf{Q}^T \delta\mathbf{s} \\ &= -\frac{|\delta\mathbf{s}|^2}{2\sigma^2}. \end{aligned} \quad (\text{Q.3})$$

8 Then, the mutation distribution in the normalized trait space is given by

$$9 \quad M(\delta\mathbf{s}) = A_0 M(\mathbf{Q}\delta\mathbf{s}) = \frac{A_0}{(2\pi)^{L/2} |\mathbf{\Lambda}|^{1/2}} \exp\left(-\frac{1}{2}|\delta\mathbf{s}|^2 / \sigma^2\right), \quad (\text{Q.4})$$

10 where  $A_0$  is determined by  $\int M(\delta\mathbf{s}) \mathbf{d}\delta\mathbf{s} = 1$ . In addition, substituting  $\mathbf{S} = \mathbf{Q}^T \mathbf{s}$  into Eq.  
11 (Q.1) gives the normalized invasion-fitness function

$$12 \quad f(\mathbf{s}'; \mathbf{s}) = \tilde{\mathbf{G}}\mathbf{Q}^T \delta\mathbf{s} + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{Q}\tilde{\mathbf{C}}\mathbf{Q}^T \delta\mathbf{s} + \frac{1}{2}\delta\mathbf{s}^T \mathbf{Q}\tilde{\mathbf{D}}\mathbf{Q}^T \delta\mathbf{s} + O(\sigma^3). \quad (\text{Q.5})$$

13 Without loss of generality, we chose the base point ~~of~~  $\mathbf{s}_0$  of the expansion, ~~at~~ as the  
14 resident phenotype  $\mathbf{s}_0 = \mathbf{s}$  after each invasion event. If the gradient  $\tilde{\mathbf{G}}\mathbf{Q}^T$  is the dominant  
15 component of the invasion fitness, such that we can neglect the higher-order terms, the MLI  
16 mutant  $\mathbf{s}'_{\text{MLI}}$  is obtained as

$$17 \quad \mathbf{s}'_{\text{MLI}} = \mathbf{s} + \sigma \frac{\mathbf{Q}\tilde{\mathbf{G}}^T}{|\mathbf{Q}\tilde{\mathbf{G}}^T|}, \quad (\text{Q.6})$$

18 by substituting Eqs. (Q.4) and (Q.5) into Eq. (8a) in Section 4. The expected waiting time for  
19 the next invasion event is given by

1 
$$T = \frac{\sqrt{2\pi}}{\left| \mathbf{Q}\tilde{\mathbf{G}}^T \right| \mu \hat{\eta} \sigma}. \quad (\text{Q.7})$$

2 Then, the directional change of the resident  $\mathbf{s}$  per unit time is given by

3 
$$\frac{\mathbf{s}'_{\text{MLI}} - \mathbf{s}}{T} = \frac{2}{\sqrt{2\pi}} \frac{\mu \hat{\eta} \sigma^2}{2} \mathbf{Q}\tilde{\mathbf{G}}^T. \quad (\text{Q.8})$$

4 Substituting  $\mathbf{s} = (\mathbf{Q}^T)^{-1} \mathbf{S}$  yields

5 
$$\begin{aligned} \frac{d\mathbf{S}}{dt} &\approx \mathbf{Q}^T \frac{\mathbf{s}'_{\text{MLI}} - \mathbf{s}}{T} = \frac{2}{\sqrt{2\pi}} \frac{\mu \hat{\eta}}{2} \sigma^2 \mathbf{Q}^T \mathbf{Q}\tilde{\mathbf{G}}^T \\ &= \sqrt{\frac{2}{\pi}} \frac{\mu \hat{\eta}}{2} \Lambda \tilde{\mathbf{G}}^T. \end{aligned} \quad (\text{Q.9})$$

6 As the canonical equation of adaptive dynamics theory (Dieckmann and Law, 1996) is given  
7 by

8 
$$\frac{d\mathbf{S}}{dt} = \frac{\mu \hat{\eta}}{2} \Lambda \tilde{\mathbf{G}}^T, \quad (\text{Q.10})$$

9 we see that Eq. (Q.9) differs from the canonical equation only in speed, by a factor of  
10  $\sqrt{2/\pi} \approx 0.798$ .