

Laws of Adaptation

A course on biological evolution in eight lectures
by Carlo Matessi

Lecture 4

Invading mutants, or the “long-term dynamics” point of view

Part II – constant selection

Wednesday October 11, 15:00-16:00

Monomorphic LTE under constant selection (Matessi and Di Pasquale, 1996)

- ❖ $x^\circ \in D \subseteq \mathbb{R}$: value of a genetically transmissible adaptive trait common to all members of an infinite population
- ❖ $g_0 g_0$: common genotype (*wild type*) of all population members associated to the trait in question, so that $x^\circ = X(g_0 g_0)$
- ❖ To verify that x° is a LTE, the number of genes controlling the trait is irrelevant; what counts is the number of genes that can be affected simultaneously by a mutation. Thus, for example, if only one gene at the time can mutate, one-gene selection dynamics would be sufficient for LTE analysis
- ❖ Simultaneous mutations of multiple genes are exceedingly rare and cannot have any significant effect on the course of an adaptation process, since very few of these, if any, can possibly appear in a relatively short time
- ❖ However a LTE should persist for an infinite (very long) time, during which even a very rare event becomes certain (very likely)
- ❖ This suggests that mutations of multiple genes, and therefore multiple genes selection dynamics, should be taken into account to verify that x° is a LTE
- ❖ After a mutational event produces a multiple mutation (a gamete that is mutated in several loci), many different mutant-carrying gametes are generated by recombination with the wild type gamete g_0

$$p_i(t+1) = \frac{\sum_{jk} w_{jk} p_j(t) p_k(t) R(i|jk)}{\bar{w}(t)}, \quad \bar{w}(t) = \sum_{jk} w_{jk} p_j(t) p_k(t)$$

- ❖ $\mathcal{G} = \{g_1, \dots, g_n\}$: the set of all possible mutant gametes generated by recombination following the original mutation event

frequency of g_i at generation $t = \varepsilon_i(t)$; $\sum_i \varepsilon_i(t) = \varepsilon(t) \cong 0$; frequency of g_0 at generation $t = 1 - \varepsilon(t) \cong 1$

$$\varepsilon_i(t+1) = \frac{2}{w_{00}} \sum_{j=1}^n w_{0j} R(i|0j) \varepsilon_j(t) + o(\varepsilon)$$

- ❖ Hence we have to deal with a short term dynamics simplified to a linear system of recursion equations
- ❖ x° is a LTE if and only if $\varepsilon(t) \rightarrow 0$ as $t \rightarrow \infty$, for all possible \mathcal{G}
- ❖ Partition \mathcal{G} in subsets $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_M$, such that \mathcal{G}_i is the subset of gametes that carry exactly i mutant genes (M being the current number of simultaneously mutated genes) and label the gametes in \mathcal{G} so that

$$\forall g_i \in \mathcal{G}_m \text{ and } g_j \in \mathcal{G}_n : n > m \Rightarrow j > i$$

- ❖ By the properties of recombination

$$R(i|0i) = \frac{1}{2} \text{ if } g_i \in \mathcal{G}_1, \quad 0 \leq R(i|0i) \leq \frac{1}{2} \text{ if } g_i \notin \mathcal{G}_1$$

$$\forall g_i \in \mathcal{G}_m \text{ and } g_j \in \mathcal{G}_n \text{ with } g_j \neq g_i : R(i|0j) > 0 \text{ only if } n > m \text{ otherwise } R(i|0j) = 0$$

- ❖ Hence

$$\varepsilon_i(t+1) = \frac{w_{0i}}{w_{00}} \varepsilon_i(t) + \frac{2}{w_{00}} \sum_{j>i} w_{0j} R(i|0j) \varepsilon_j(t) + o(\varepsilon), \quad \forall g_i \in \mathcal{G}_1$$

$$\varepsilon_i(t+1) = \frac{2w_{0i}R(i|0i)}{w_{00}} \varepsilon_i(t) + \frac{2}{w_{00}} \sum_{j>i} w_{0j} R(i|0j) \varepsilon_j(t) + o(\varepsilon), \quad \forall g_i \notin \mathcal{G}_1$$

- ❖ Thus, the transition matrix of this recursion is upper-triangular and its eigenvalues are given by its diagonal elements

$$\begin{aligned}\varepsilon_i(t+1) &= \frac{w_{0i}}{w_{00}} \varepsilon_i(t) + \frac{2}{w_{00}} \sum_{j>i} w_{0j} R(i|0j) \varepsilon_j(t) + o(\varepsilon) \quad , \quad \forall g_i \in \mathcal{G}_1 \\ \varepsilon_i(t+1) &= \frac{2w_{0i}R(i|0i)}{w_{00}} \varepsilon_i(t) + \frac{2}{w_{00}} \sum_{j>i} w_{0j} R(i|0j) \varepsilon_j(t) + o(\varepsilon) \quad , \quad \forall g_i \notin \mathcal{G}_1\end{aligned}$$

- ❖ Hence the condition for x° being a LTE [$\varepsilon(t) \rightarrow 0$ as $t \rightarrow \infty$, $\forall \mathcal{G}$] reduces to

$$w_{00} > w_{0i} \quad , \quad \forall g_i \in \mathcal{G}_1 \quad \text{and} \quad w_{00} > 2R(i|0i)w_{0i} \quad , \quad \forall g_i \notin \mathcal{G}_1 \quad ; \quad \forall \mathcal{G}$$

- ❖ Fitness being a function, $W(\cdot)$, of the trait value determined by the genotype, $X(\cdot)$, this can be written

$$W(x^\circ) > W(X(g_0g_i)) \quad , \quad \forall g_i \in \mathcal{G}_1 \quad \text{and} \quad W(x^\circ) > 2R(i|0i)W(X(g_0g_i)) \quad , \quad \forall g_i \notin \mathcal{G}_1 \quad ; \quad \forall \mathcal{G}$$

- ❖ Provided that there are no constraints of genetic expression, this is equivalent to

$$W(x^\circ) > W(x) \quad \text{and} \quad W(x^\circ) > 2R(i|0i)W(x) \quad , \quad \forall g_i \notin \mathcal{G}_1 \quad ; \quad \forall x \in D$$

- ❖ But, considering that $2R(i|0i) \leq 1$, $\forall g_i \notin \mathcal{G}_1$, the second inequality is implied by the first; so that we conclude:

- ❖ For a trait subject to constant selection, a monomorphic LTE obtains at x° if and only if fitness is maximized at x°

Properties of invading mutants

- ❖ The result presented below, determined by Eshel and Feldman in 1984 for two genes with an arbitrary number of alleles, was later shown by Lieberman (1988) to hold for any number of genes
- ❖ Consider a generic transient state of long term evolution (a short term equilibrium but not a LTE), of a continuous trait subject to selection, polymorphic with respect to two genes

r : frequency of recombination between the two genes

$\{A_1, \dots, A_n\}$, $\{B_1, \dots, B_m\}$: alleles

p_{ij} : (equilibrium) frequency of gamete $g_{ij} \equiv (A_i B_j)$

$X(g_{ij}g_{kl}) = x_{ijkl}$: trait value of genotype $g_{ij}g_{kl}$; $x_{ijkl} = x_{klij} = x_{ilkj}$

$W(x_{ijkl}) = w_{ijkl}$: fitness of x_{ijkl}

- ❖ Suppose that a mutant allele of gene A, A_0 , appears, determining the formation of:

$\{g_{01}, \dots, g_{0m}\}$: mutant-carrying gametes

$\{\varepsilon_{01}(t), \dots, \varepsilon_{0m}(t)\}$: their frequencies ; $\varepsilon_{01}(t) + \dots + \varepsilon_{0m}(t) = \varepsilon(t) \approx 0$

$\{p_{11} - \delta_{11}(t), \dots, p_{nm} - \delta_{nm}(t)\}$: perturbed frequencies of resident gametes ; $\delta_{11}(t) + \dots + \delta_{nm}(t) = \varepsilon(t)$

$$p_i(t+1) = \frac{\sum_{jk} w_{jk} p_j(t) p_k(t) R(i|jk)}{\bar{w}(t)}, \quad \bar{w}(t) = \sum_{jk} w_{jk} p_j(t) p_k(t)$$

- ❖ The frequencies of mutant gametes satisfy *initially* the following (linearized) system of recurrence equations

$$\varepsilon_{0i}(t+1) = \frac{1-r}{\bar{w}} \sum_{k=1}^n \sum_{l=1}^m w_{0ikl} p_{kl} \varepsilon_{0i}(t) + \frac{r}{\bar{w}} \sum_{k=1}^n \sum_{l=1}^m w_{0lki} p_{ki} \varepsilon_{0l}(t) + o(\varepsilon(t)) \quad ; \quad \bar{w} = \sum_{i=1}^n \sum_{k=1}^m \sum_{j=1}^n \sum_{l=1}^m w_{ikjl} p_{ik} p_{jl}$$

- ❖ The transition matrix, \mathbf{Q} , of this system is positive:

$$(\mathbf{Q})_{ii} = \frac{1-r}{\bar{w}} \sum_{k=1}^n \sum_{l=1}^m w_{0ikl} p_{kl} + \frac{r}{\bar{w}} \sum_{k=1}^n w_{0iki} p_{ki} \quad , \quad (\mathbf{Q})_{ij} = \frac{r}{\bar{w}} \sum_{k=1}^n w_{0jki} p_{ki}$$

- ❖ Its dominant eigenvalue, λ , and associated right eigenvector, \mathbf{q} , are therefore real and positive
- ❖ Thus λ is the asymptotic rate of growth of the mutant gametes sub-population and \mathbf{q} , that can be normalized so that its elements sum to one, is the asymptotic (stationary) composition of such sub-population
- ❖ The mutant allele invades if $\lambda > 1$, because then $\varepsilon(t) \rightarrow \infty$ as $t \rightarrow \infty$; the mutant on the opposite is eliminated if $\lambda < 1$, because then $\varepsilon(t) \rightarrow 0$ as $t \rightarrow \infty$

$$(\mathbf{Q})_{ii} = \frac{1-r}{\bar{w}} \sum_{k=1}^n \sum_{l=1}^m w_{0ikl} p_{kl} + \frac{r}{\bar{w}} \sum_{k=1}^n w_{0iki} p_{ki}, \quad (\mathbf{Q})_{ij} = \frac{r}{\bar{w}} \sum_{k=1}^n w_{0jki} p_{ki}$$

- ❖ By definition, λ satisfies to

$$\lambda \mathbf{q} = \mathbf{Q} \mathbf{q} \Leftrightarrow \lambda q_i = \sum_j (\mathbf{Q})_{ij} q_j = \frac{1-r}{\bar{w}} \sum_{k=1}^n \sum_{l=1}^m w_{0ikl} p_{kl} q_i + \frac{r}{\bar{w}} \sum_{k=1}^n \sum_{l=1}^m w_{0lki} p_{kl} q_l, \quad \forall i = 1, \dots, m$$

- ❖ Summing over $i = 1, \dots, m$ this gives

$$\lambda = \frac{\sum_i \sum_{kl} q_i p_{kl} w_{0ikl}}{\sum_{ij} \sum_{kl} p_{ij} p_{kl} w_{ijkl}} = \frac{w_{mut}}{\bar{w}}$$

- ❖ The quantity w_{mut} at the numerator has a neat interpretation: it is the (hypothetical) mean fitness that the sub-population of mutant individuals initially present in the population (of genotype $g_{0i}g_{kl}$) would eventually attain if the linearized recursion retained its validity indefinitely. The quantity at the denominator, on the other hand, is the actual mean fitness of the resident population, before introduction of mutants
- ❖ It can therefore be concluded that a mutant allele invades if

$$w_{mut} > \bar{w}$$

namely, if it can (potentially) increase the mean fitness of its carriers above the level prevailing in the resident population where it is introduced. Conversely, a mutation cannot invade if it would lower the mean fitness of its carrier below that of the residents

- ❖ Successful invasions *tend* to increase the mean fitness of the population
- ❖ This however does not mean that at any step of long term evolution fitness indeed increases, because: (i) the actual frequency of the newly introduced genotypes, $g_{0i}g_{kl}$, at a new equilibrium following invasion, bears no relation to the hypothetical frequency distribution, \mathbf{q} , implied by the linearized recursion; (ii) other genotypes ($g_{0i}g_{0j}$), that will be common when the newly introduced gametes cease to be rare, do not contribute to w_{mut} and they could very well have a low fitness, without compromising the capacity of the mutation to invade

Polymorphic LTE under constant selection

- ❖ The Eshel-Feldman-Lieberman result permits us to reach an immediate but quite general and important conclusion concerning LTE under constant selection. Namely that, under very mild conditions, *polymorphic LTE cannot exist* and that, therefore, the only LTE are attained at trait values that maximize fitness, irrespective of the number of genes controlling the trait
- ❖ The proof is trivial. Consider any polymorphic state of a population, with gametes $\mathcal{G} = \{g_1, \dots, g_n\}$, phenotypes $\{X(g_i g_j) = x_{ij} \in [a, b] : i, j = 1, \dots, n\}$, and fitnesses $\{W(x_{ij}) : i, j = 1, \dots, n\}$
- ❖ Assume that $W(\cdot)$ is continuous in $[a, b]$, and that its maximum here is attained at a unique point x° . Assume moreover that there are no restrictions to expression of genes controlling the trait in question preventing us from doing the following
- ❖ Introduce a mutant allele of one of the genes of the gametes in \mathcal{G} and let $\mathcal{G}^* = \{g_1^*, \dots, g_n^*\}$ be the set of mutant gametes obtained by replacing in each of the resident gametes its current allele at the target gene with the mutant allele. To the mutant genotypes thus arising assign now a trait value such that

$$\left| X(g_i^* g_j) - x^\circ \right| \leq \delta \quad \forall g_i^* g_j \in \mathcal{G}^* \times \mathcal{G}$$

where δ has been chosen so small, possibly even $\delta = 0$, to ensure that

$$W\left(X(g_i^* g_j)\right) \geq \max_{x_{ij}} W(x_{ij}) \quad \forall g_i^* g_j$$

Notice that there will be at least one x_{ij} such that $W(x_{ij}) < W(X(g_k^* g_l)) \quad \forall g_k^* g_l$, because the resident population is polymorphic. It then follows that, by the Eshel-Feldman-Lieberman result, the mutation represented by \mathcal{G}^* certainly invades and that therefore \mathcal{G} is not a LTE

Convergence to LTE under constant selection

- ❖ The Eshel-Feldman-Lieberman result provides encouraging evidence that a trait, evolving under constant selection, will in the long run converge to an optimal value at a LTE, but it is far from proving it. The topic is difficult because, if taken in absolute generality, it would amount to consider the complete non linear short-term dynamics of multi-gene selection systems, in order to determine the new equilibrium state attained by a population after a successful invasion
- ❖ Further support in favour of convergence to LTE can be gleaned, however, if we consider invasion of monomorphic transient states and, besides ignoring multiple gene mutations, make some other simplifying assumption:
 - (i) assume a continuous fitness function, $W(\cdot)$, and restrict attention to an interval (a,b) around an internal LTE, x° , sufficiently small that $W(\cdot)$ is monotone in each of the two “half” intervals (a,x°) and (x°,b)
 - (ii) mutations have a small effect; namely any one-gene mutation changing a resident gamete g , to a mutant g^* has phenotypic effect limited by

$$\left| X(g^*g^*) - X(gg) \right| \leq \delta \quad \text{where} \quad \delta \ll (x^\circ - a), (b - x^\circ)$$

- (iii) “partial dominance”: *heterozygotes* (g^*g) have a phenotype intermediate between those of the corresponding *homozygotes* (g^*g^* and gg)

$$\min \{ X(g^*g^*), X(gg) \} < X(g^*g) < \max \{ X(g^*g^*), X(gg) \}$$

- ❖ Take now, as starting point of an evolutionary path, a monomorphic population, of genotype (gg) and trait value $X(gg) = x \in (a,b)$, but $x \notin [x^\circ - \delta, x^\circ + \delta]$, and suppose that a mutation occurs at one of the genes controlling the trait in question, producing a mutant gamete g^* , such that $X(g^*g) = x'$ and $X(g^*g^*) = x''$
- ❖ The mutation invades if and only if $W(x') > W(x)$; in which case it follows that $|x' - x^\circ| < |x - x^\circ|$, by the hypothesis of small mutation effects; moreover, by the partial dominance hypothesis, $|x'' - x^\circ| < |x' - x^\circ|$, so that $W(x'') > W(x')$, implying that g would not be able to invade g^*
- ❖ Since only one gene is variable in this population after invasion, short term selection dynamics guarantees maximization of mean fitness; hence at the new short term equilibrium the mutant gamete g^* will have entirely replaced the resident gamete g and the population will be once again monomorphic, but with phenotype $X(g^*g^*) = x''$. As a result the population has moved closer to the LTE at x° .



- ❖ Iteration of such a process, by which the population stays monomorphic and moves toward the LTE, will unavoidably take the population within the small neighborhood of the LTE, $(x^\circ - \delta, x^\circ + \delta)$
- ❖ Once this has occurred we can no longer predict the further path of evolution because the population is now most likely to become polymorphic. The reason is that a mutation might easily arise such that $X(gg)$ and $X(g^*g^*)$ are on opposite sides of x° , and moreover $W(X(g^*g)) > W(X(gg)), W(X(g^*g^*))$, a fitness configuration under which one-gene short term dynamics leads to a polymorphism, with the two genes coexisting. Notice that in this case, g^* invades g but also g could invade back g^* . If the trait is controlled by more than one gene, such initial one-gene polymorphism can rapidly evolve into a multi-gene polymorphism, a situation where analysis is forbidding

