

Optimality principles in evolutionary genetics

Narayan Behera

A population evolves due to changes in its gene frequencies arising due to mutations, natural selection, random genetic drifts and migrations. Svirezhev introduced an integral variational principle, in analogy with the least-action principle of classical mechanics, by defining a Lagrangian which remained stationary on the trajectory followed by the population undergoing selection. This principle can also be extended to multiple loci in some simple cases. However, in a two-locus model or more general models, there is no straightforward extension of this principle if linkage and epistasis are present. The local optimality principle can be geometrically formulated in a Riemannian metric space of gene frequencies so that, under evolutionary pressures, the population trajectory moves in that direction along which the increment of the mean fitness is maximum.

The causes of evolution

Gene is a physical entity, transmitted from parent to offspring, that influences a hereditary trait. From a biochemical point of view, genes are fundamental units of genetic information that correspond to the sequence of nucleotides in a segment of DNA. The position of a particular gene in a chromosome is called a locus. Genes can exist in different forms or states. For example, a gene for hemoglobin may exist in a normal form or in any one of a number of alternative forms that result in hemoglobin molecules that are more or less abnormal. These alternative forms of a gene are called alleles. The **gene pool** of a population is defined as the collection of genes belonging to all the members of a population taken together. **Genotype** is the hereditary or genetic constitution of an individual. Genotype thus refers to the particular alleles present in an organism at all loci that affect the trait in question. The sum total of observable structural and functional properties of an organism, which is the product of the interaction between the genotype and the environment, is the **phenotype** of the individual. **Epistasis** is the mode of interaction among nonallelic genes which results in the phenotype of an organism (of a given genotype) that is not mere 'sum' of the phenotypes caused by genotypes at each locus taken separately. **Linkage** is the association in the inheritance of the genes that are located in a single chromosome (i.e. linkage group), and are termed the linked ones. Breakage of the joint transmission of the linked gene occurs as a result of recombination. The formulation of a problem in popu-

lation genetics involves first assuring that some trait is the expression of a specific genetic mechanism. The net numbers of surviving progenies that an organism (genotype) leaves behind, to the next generation (relative to that of other genotypes), is defined as its **fitness**. Fitness is usually measured in terms of malthusian parameter, which is the difference between the birth rate and death rate.

In a sexually reproducing population, the genes get scrambled in each generation owing to segregation and recombination, the genotypes constantly changing as individuals die and new ones are born. The unique expression of a particular trait of a genotype depends on the manner in which the alleles of a gene interact during the development of an organism. For the expression of any of the alleles of one gene, dominance refers to the concealment of the presence of one allele by the strong phenotypic effect of another. The basic mechanisms by which genes are passed on from one generation to the next were discovered by Mendel in 1865 (and were rediscovered by genetics experiments in the beginning of this century). His law of segregation states that every somatic cell of an individual carries a pair of genes for each character; the members of each pair separate during meiosis so that each gamete carries one gene of each pair, and gametes pair at random to form the genotype of the offspring. The law of independent assortment says that each member of any pair of alleles is equally likely to be combined with either member of another pair of alleles, since they associate independently.

Genotypes undergo shuffling during gamete formation owing to segregation and recombination, and are reassembled in each generation by the process of fertilization. Due to random mating, the probability of choosing a particular genotype for a mate is equal to the relative frequency of that genotype in the population. In the absence of factors that change the gene frequencies (described below) coupled with random mating, the

Narayan Behera is in Evolutionary Population Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; and also at Theoretical Sciences Unit, Jawaharlal Nehru Center for Advanced Scientific Research, Jakkur, Bangalore 560 012, India.

population will have constant gene and genotype frequencies from generation to generation. This is called the Hardy-Weinberg principle, which is true only for an infinitely large population. Many populations in nature have births and deaths occurring more or less continuously, with both reproduction and mortality at various ages.

Thus, the changes in the gene frequencies in a population is the process of **evolution**. Evolution is caused by the following four basic forces acting on the population. (i) *Mutation*: Mutation occurs by 'chance' due to error in copying of chromosomes. It is a statistical fluctuation undergoing in nature. One allelic form changes into another under mutation. Mutation is the ultimate source of new or novel genes and it also prevents old alleles from ever being entirely eliminated. Recurrent mutation is fundamental to the process of evolution because it helps to maintain a supply of genetic variation for selection to act on. (ii) *Natural selection*: The fitter genotypes survive to produce more progeny for the next generation compared to the less fit genotypes which cannot leave their genes in the genepool. As long as the alleles which are present in the population do not all have the same fitness values, natural selection will change the state of the genepool. Selection occurs when one genotype leaves a different number of progeny than another. This may happen because of differences in survival, in mating or in fertility. This is the **Darwinian paradigm**. The idea of natural selection is an extension of the idea of artificial selection (as practiced by animal and plant breeders) except for the fact that in the former no conscious agent is involved to bring about selection. The spirit of the Darwinian theory of natural selection is that natural selection uses the genetic variation existing in a population to produce individuals that are better adapted to their environment. Through natural selection, alleles that enhance survival and reproduction increase gradually in frequency from generation to generation, and the population becomes progressively better able to survive and reproduce in the environment. A general result of natural selection is the progressive genetic improvement in the population which constitutes the process of evolutionary adaptation. (iii) *Genetic drift*: Genetic drift is the change in the general structure due to errors or 'noise' that creeps into the transmission of the genepool from generation to generation. Random sampling of alleles in a finite population results in chance changes in allele frequencies. The random sampling from the gamete pool means that some alleles may be over represented in current generation relative to their frequencies in previous generation and some alleles may be underrepresented. An allele can get fixed in the population after the passage of few generations by pure chance. (iv) *Migration*: Many species occupy a rather broad area and they may also be subdivided into several more or less separate populations. Most subpopulations are connected to one another through

recurrent interchange of migrants. The population subdivision prevents random mating from occurring. However, the migration among the subpopulations can result in change in the gene frequencies.

Selection changes the variance in a given population. Whether the variance increases or decreases depends on the gene frequencies, dominance, epistasis, linkage and the mating system. Natural selection is based on total fitness, and the effect on any measurable trait depends on the relation between that trait and the expectation of surviving and reproducing. Natural selection, like classical mechanics, has both static and dynamic aspects. The former involves the relatively stable situation that results from the balance of various opposing forces – mutations, selection, migrations, and random fluctuations.

Fundamental theorem of natural selection

The variance of any trait in a population is in general determined partly by genetic factors and partly by environmental factors. We are primarily interested in that part of the variance which is determined solely by genetic differences among individuals in the population. Genic values are measures of the contribution of each gene, i.e. averaged over all genotypes into which this gene enters, and are measured as a deviation from the mean. Fisher enunciated his Fundamental Theorem of Natural Selection as: the rate of increase of mean fitness of a population at any time is equal to its genic variance in fitness at that time¹. This holds true when the fitness depends on the genes at one locus only in a randomly mating system. The theorem accounts for the effect of gene frequency changes alone, isolated from the other things that are happening. **Fitness** measures the evolutionary improvement of a population brought about by the changes in the gene frequencies. Natural selection tends to preserve those genes which increase the fitness of their carriers. Intuition tells us that the rate at which selection changes the fitness will be related to the variability of the population. Under natural selection, the population follows a trajectory in the space of gene frequency such that the fitness of the population increases at a rate equal to the genic variance. This is the first instance of an optimality principle in population genetics. The Fundamental Theorem is compared with the second law of thermodynamics where entropy always increases in a physical process. The status of Fisher's theorem has long been controversial in evolutionary theory. It is well known that it is not exactly true in any but the simplest models of population genetics. When fitness depends on the presence of genes at more than one locus, the total change in mean fitness can be negative, even under random mating system. This essentially happens due to recombination and nonadditive gene effects. If it is reinterpreted as an approximate statement, i.e. if it is taken to state that the rate of change of fitness is

approximately equal to the genic variance, then its range of applicability is significantly enhanced. Thus, the new interpretation of Fisher's theorem, put forward by Ewens, focuses on the partial change in the mean fitness, rather than the total change in the fitness². Fisher attributed this component of the change in mean fitness to the changes in gene frequency caused by natural selection, as opposed to the change due to the environment. This partial change is the genic variance which is irrespective of the mating system. Furthermore, this result holds for multiple loci with multiple alleles at each locus, and for arbitrary recombination structure between the various loci. Thus, the partial increase in mean fitness is always nonnegative, and is due to the average effects of the alleles which form the basic components in the new interpretation. Fisher's theorem has an interesting analog for the rate of change of a character, which is correlated to its fitness, wherein one replaces 'genic variance' by 'genic covariance' in the statement of Fisher's theorem to obtain an interesting true generalization^{3,4}. Thus, using the approximate form of Fisher's Theorem, one can calculate or measure the genic variance and, subsequently, predict that the rate of change of a population's fitness will be approximately equal to it. Fisher dealt with a continuous population and was not clear whether the theorem could be applied to discrete generations as well. Kimura was the first to add explicit terms into the theorem to account for environmental changes (or other causes of changes in the genotypic fitnesses)⁵. His treatment has been widely regarded as clarifying and extending Fisher's Fundamental Theorem. While the Fundamental Theorem states that the average fitness of a population increases as long as the population is not in equilibrium, Kimura's maximum principle states that the change in gene frequencies occurs in such a way that the increase in average fitness is maximal.

Price's analysis of Fundamental Theorem recognizes the fact that the theorem refers to the partial change⁶. Finally, Ewens' clearer formulation of the theorem led to its wider acceptance². Ewens believed that Fisher was interested in fitness changes only through changes in the gene frequency and not through changes in the expected value of the fitness. The theorem, although exact, measures only one component of the fitness change. So the modern interpretation of the Fundamental Theorem (following Ewens and Price) is: 'the rate of increase in the mean fitness of any population at any time ascribable to natural selection acting through changes in gene frequencies is exactly equal to genic variance in the fitness at that time'. However, while Ewens' new formulation does not require any specific mating systems, and holds good for both discrete and continuous time, Fisher's result is true for random mating only. Thus, it is clear that the calculation of the partial change is exact and involves no approximations. This supports the argument that Fisher himself viewed the theorem as exact (for some background discussion, see ref. 7).

Price was perhaps correct in saying that Fisher viewed the partial changes in mean fitness as those due to single-locus gene-frequency changes. These changes essentially form the basis of evolution according to Fisher (in contrast, for example, to the evolution of co-adapted gene complexes favoured by Wright⁸). This may be the central tenet of the Fundamental Theorem. Nagylaki has shown that, under most circumstances, the terms additional to the genic variance are of a smaller order than the genic variance; the major exception occurs when the population is near an equilibrium⁹. Nagylaki believed that the biological significance of the partial change in the mean fitness remains to be demonstrated¹⁰. Ewens also recorded a negative assessment of the theorem as a biological statement². Even if the Fundamental Theorem is inexact and incomplete, it captures the essence of the way selection works, and encapsulates a great deal of evolutionary insight into a simple expression¹¹. The presence or absence of dominance, in the effect of genes, can play an important role for the optimality principle to hold true¹².

Global optimization

Hamilton's principle in classical mechanics can be stated as follows: Of all the possible paths along which a dynamical system may move from one point to another within a specified time interval (consistent with any constraints), the actual path followed is that which minimizes the time integral of the difference between the kinetic and potential energies. The variational statement of the principle states that the integral of the difference of kinetic and potential energies is an extremum, and not necessarily a minimum. But in almost all important applications in dynamics, the minimum condition occurs.

The intuition that leads to the expectation that the optimality principles must exist is, that if selection acts to increase a well-defined quantity (called 'fitness') over a period of time during which selection acts, then some quantity correlated with fitness ought to be maximized. In a standard, continuous time, one-locus selection model with random mating and constant (frequency- and time-independent) fitness, Svirezhev showed that for the trajectory followed by the population in its configuration in space, the integral of the 'Lagrangian' is:

$$\frac{1}{8} \left(\sum_i (\dot{p}_i^2 / p_i) + \sum_i p_i (m_i - \bar{m})^2 \right),$$

(where p_i is the frequency of the i th allele, m_i its (marginal) fitness, \bar{m} is the mean fitness of the population and $\dot{p}_i = dp_i/dt$ is stationary¹³). This principle is a very close analog to Hamilton's principle in classical mechanics where the Lagrangian is equal to T (the kinetic energy) $- V$ (the potential energy) of a system: this integral is known as the action. This is reminiscent of the

principle of least action and Hamilton's principle in classical mechanics whereby a minimization process leads to Newton's laws of motion (see Lanczos¹⁴ for a detailed discussion of the history and the limitation of this principle).

The range of validity of Svirezhev's principle has recently been explored for multilocus genetic system in a systematic way^{15,16}. To be precise, that principle will be taken to state that a population undergoing selection follows a trajectory that keeps the integral of a Lagrangian function stationary, where the Lagrangian has the form indicated above. This particular form of the Lagrangian, that Svirezhev had introduced, is interesting because, leaving out the factor of 1/8, it is exactly equal to the variance. The second term is obviously half the variance (half because of diploidy) and the first term is also half the variance because,

$$\dot{p}_i = p_i(m_i - \bar{m}).$$

However, these statements are true only on the natural-selection trajectory. The potential utility of this principle is clearly the same as Fisher's: one writes down the formula for the genic variance, and then expects the population to follow the trajectory where (to some approximation) the integral of one-eighth of the variance remains stationary. The general strategy followed throughout this paper is therefore: (i) the equations governing the change of allelic or gametic frequencies over time are written down (which are derived using the standard techniques of population genetics), (ii) using Svirezhev's principle, a Lagrangian function is written down, (iii) the Euler-Lagrange equations corresponding to the Lagrangian are derived, using the method of Lagrange multipliers to incorporate the known constraints. This is the stage at which the calculus of variations plays a role. That the integral of the Lagrangian is stationary (over a trajectory) is equivalent to the condition that the Euler-Lagrange equations hold for the motion, (iv) these Euler-Lagrange equations are solved to obtain the equations of motion for the system, and (v) this solution is compared to the equations that were obtained using the standard techniques. When linkage disequilibrium is zero, the rate of change of allelic frequency can be correlated with Malthusian parameters (hence with experimentally observable quantities).

Genetic flexibility, for adaptation to a fluctuating environment, demands recombination to generate adaptive combinations different from those previously selected. The roles of linkage and epistasis are very important in the context of genomic evolution. In the above analysis, the rate of change of allelic frequency is used, for mathematical convenience, to prove the variational principle. With linkage and epistasis, the rates of change of allele frequencies no longer suffice to determine the genotypic dynamics of the population, instead gametic

(or haplotype) frequencies must be used. Svirezhev's principle does not hold good even in simple models when linkage and epistasis are incorporated¹⁶.

Population trajectories in a metric space

All trajectories of the population equations have been examined only in the Euclidean space. In the Euclidean space, the trajectories of the population equations do not provide the steepest ascent for the mean fitness. Hence, nothing can be said about the dynamics of the movement of the population towards the stable equilibrium condition. However, in a suitable Riemannian metric space, the trajectories can be constructed as a gradient of mean fitness, so that, under evolutionary pressure, the population selects that direction out of all the possible directions along which the increment of the mean fitness is maximum. It would be better if these trajectories possessed some external features, for instance if they were trajectories of a vector field of a gradient of mean fitness. The local principle of optimality can be formulated as follows. If we examine the motion of a population in a certain Riemannian metric space of gene frequencies with the basic form of the matrix tensor g_{ij} defined as:

$$g_{ij} = \frac{1}{2} \frac{p_i \delta_{ij} + p_j \delta_{ji}}{p_i p_j},$$

where δ_{ij} is the delta of Kroneker, then all trajectories of its motion in this space lie on the unit sphere. At any moment in time, from all possible directions on this sphere of the population, such a path is chosen that the increase of its average fitness is a maximum, and the population moves at a rate proportional to a certain degree of the genetic variation in the population. This rate V equals¹³

$$|V| = \sqrt{\sum p_i (m_i - \bar{m})^2}.$$

In the n -locus system (when all possible linkage disequilibrium are zero), the corresponding surface on which the phase point moves is the product space of n higher-dimensional unit spheres. When linkage and epistasis are present, the metric tensor will be dependent on fitness parameters in addition to gametic frequencies; and in a two-locus two-allele genetic system, the population trajectories will lie on a three-dimensional surface with a complicated geometry¹⁶.

The particular form of metric tensor, used by Svirezhev for mathematical convenience, was given suitable biological justification¹⁷. Thus, overall action of natural selection minimizes the genetic distance between parental and daughter gene frequency values such that the partial

increase in mean fitness is the natural selection value. This genetic distance has the same quadratic functional form as the metric. This new interpretation is based on the concept of the average effect of an allele. The gradient pattern of genetic process of variations in allele frequencies under selection pressure as a motion in a Riemannian space as well as of the properties of this motion in the case of several loci, was considered¹⁸. Therefore while examining the trajectory of a population as the trajectory of a kind of mechanical motion in a special chosen space, we cannot link its geometry to the characteristics of selection. Instead, this geometry is determined by a system of interbreeding and by-laws of heredity. It is interesting therefore to look at the trajectory of a population in such a space, the geometry of which would immediately be linked to characteristics of selection, for example in the metric space with the basic form:

$$ds^2 = a_{ij} dx^i dx^j,$$

$i, j = 1, 2, \dots, n$, where a_{ij} are the coefficients of the relative viability and dx^i is the infinitesimal change in the contravariant vector x^i , the components of which are equal to the allelic frequencies p_i . It is assumed that $\det // a_{ij} // \neq 0$. Thus, in such a system, polymorphism is present, i.e. in the population 'the persistence of two or more alleles at the fixed locus in the population'. However, this method appears fruitful only for the analysis of the stationary condition existing in a polymorphic system¹³.

Conclusion

Minimal principles in physics have a long history. The search for such principles is primarily motivated on the ground that nature always minimizes certain important quantities when a physical process takes place. In 1657, Fermat formulated such a principle by postulating that a light ray always travels from one point to another in a medium by a path that requires the least time. The principle of least action was mathematically formulated by Lagrange (1760) who asserted that dynamical motion takes place with minimum action. Hamilton (1835) defined the dynamical principle on which it is possible to base all of mechanics. Hamilton's principle allows us to calculate the equations of motion of a body completely without recourse to Newtonian theory. The equations of motion are obtained without the necessity of explicitly taking into account the external forces acting on the body. According to Hamilton's principle, in Nature, the motion of a body occurs to minimize the time integral of the difference between the kinetic and potential energies. Furthermore, using this approach, the connection between symmetry properties and the invariance of physical

quantities can be suitably formulated. Hamilton's method is in essence a *posteriori*, because we know beforehand that a result equivalent to Newtonian equations must be obtained which are correlated with experimental facts. Hamilton's principle, based on a single basic postulate, gives satisfying unification of many individual theories. It is a goal of physical theories to describe observed phenomena mathematically with an economy of fundamental postulates.

The failure to extend Svirezhev's principle to models with linkage does not mean that no function exists such that the Euler-Lagrange Equations obtained using it (as the Lagrangian) would give rise to the appropriate dynamics. However, unless there exists some method by which the Lagrangian function can be obtained *without already knowing the dynamics*, the Lagrangian reformulation of these models is of little value. In physics, this is straightforward because the Lagrangian is equal to $T - V$, where T is the kinetic energy and V is the potential energy of the system. Similarly, in these population genetics models, without linkage or epistasis, the Lagrangian is simply one-eighth of the variance written in Svirezhev's form. The failure of the Svirezhev principle however means that this recipe for finding the Lagrangian no longer works. Moreover, there is no straightforward way to change the recipe, for instance, by the introduction of an additional additive term in the Lagrangian¹⁶. In the case of Fisher's Fundamental Theorem, the presence of linkage and epistasis can be taken into account (in two-locus continuous time models) by adding terms to the genic variance (see Crow and Kimura¹⁹). However, the results reported in ref. 16 show that such a strategy is not available for the Lagrangian formulation. This formulation cannot play a particularly useful heuristic role in any but the simplest situations where linkage disequilibrium is zero. The success of Svirezhev's principle in these cases suggests the following conjecture: this principle is applicable in exactly those circumstances where Fisher's Fundamental Theorem, in its exact form, i.e. the rate of change of the mean fitness is exactly equal to the genic variance, is true.

The formulation of the population trajectory as a trajectory of steepest ascent in a Riemannian metric space provides an insightful way of looking at evolution. The analysis of Svirezhev's variational principle in the population genetics models involving: (i) non-random mating, and (ii) multiple loci with non-zero linkage disequilibria are mathematically too cumbersome. In this work some analogies have been formulated between classical mechanics and certain problems in mathematical theory of population genetics.

1. Fisher, R. A., *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford, 1930.
2. Ewens, W. J., *Theor. Pop. Biol.*, 1989, **36**, 167-180.

GENERAL ARTICLES

3. Crow, J. F. and Nagylaki, T., *Am. Nat.*, 1976, **110**, 207–213, 400.
4. Nagylaki, T., *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 1910–1913.
5. Kimura, M., *Heredity*, 1958, **12**, 145–167.
6. Price, G. R., *Annu. Hum. Genet.*, 1972, **36**, 129–140.
7. Edwards, A. W. F., *Biol. Rev. Cambridge Philos. Soc.*, 1994, **69**, 443–474.
8. Wright, S., *Am. Nat.*, 1988, **131**, 115–123.
9. Nagylaki, T., *Genetics*, 1976, **83**, 583–600.
10. Nagylaki, T., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 2402–2406.
11. Crow, J. F., *Theor. Pop. Biol.*, 1990, **38**, 263–275.
12. Narain, P., *J. Genet.*, 1993, **72**, 59–71.
13. Svirzhev, Y. M., *Studies on Theoretical Genetics* (ed. Ratner, V. A.), USSR Academy of Sciences, Novosibirsk, 1972, pp. 86–102.
14. Lanczos, C., *The Variational Principles of Mechanics*, Toronto Press, Toronto, 1959.
15. Behera, N., *J. Genet.*, 1995, **74**, 19–24.
16. Behera, N., *Bull. Math. Biol.*, 1996, **58**, 175–202.
17. Ewens, W. J., *Theor. Pop. Biol.*, 1992, **42**, 333–340.
18. Shahshahani, S., *Mem. Am. Math. Soc.*, 1979, **17**.
19. Crow, J. F. and Kimura, M., *An Introduction to Population Genetics Theory*, Harper and Row, New York, 1970.

ACKNOWLEDGEMENT. I thank Sahotra Sarkar for discussions.

Received 18 March 1998; revised accepted 16 September 1998

MEETINGS/SYMPOSIUMS/SEMINARS

TRendys-1999

Date: 23–24 November 1999

Place: New Delhi

TRendys-1999, a National Forum to discuss with a difference frontier areas and emerging concept in Biochemistry and Molecular Biology is organising its next meeting during 23–24 November 1999 at the Department of Biochemistry, University of Delhi, South Campus New Delhi 110 021. The following are expected to speak at this meeting:

Ramesh Bamezai, New Delhi; S. K. Brahmachari, New Delhi; D. P. Burma, Calcutta; Dipankar Chatterjee, New Delhi; V. S. Chauhan, New Delhi; N. K. Ganguly, New Delhi; Amit Ghosh, Chandigarh; Seyed E Hasnain, Hyderabad; Shahid Jamil, New Delhi; Shekar Mande, Chandigarh; T. Ramasarma, Bangalore; M. R. S. Rao, Bangalore; K. Subba Rao, Hyderabad and Satyajit Rath, New Delhi.

Those desirous of participation may please contact:

Prof. A. K. Tyagi
Department of Biochemistry
University of Delhi South Campus
Benito Juarez Road
New Delhi 110 021
Fax: 011-6885270
E-mail: ATT1000@hotmail.com

National Workshop on Transfer of Forest Techn

Date: 16–17 December 1999

Place: Coimbatore

The technical session would be organised around areas: 1. Tested technologies that can be transferred to SFDs and industries; 2. Technologies that need to be developed through Joint Forest Technology Development Centres; 3. Technologies with future implications which require support by Forest Departments, industries and other agencies.

Contact: Shri. K. Subramanian
Director
Institute of Forest Genetics and Tree Breeding
PB No. 1061, R.S. Puram Post
Coimbatore 641 002
Tel: 0422-431540
Fax: 0422-430549
E-mail: ifgtb@satyam.net.in