Genet. Res., Camb. (2005), **86**, *pp.* 53–63. *With 3 figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672305007688 Printed in the United Kingdom

Neutral mutation as the source of genetic variation in life history traits

KRUNOSLAV BRČIĆ-KOSTIĆ*
Department of Molecular Genetics, Ruđer Bošković Institute, Bijenička 54, HR-10000 Zagreb, Croatia

(*Received 21 December 2004 and in revised form 16 May and 26 June 2005*)

Summary

The mechanism underlying the maintenance of adaptive genetic variation is a long-standing question in evolutionary genetics. There are two concepts (mutation–selection balance and balancing selection) which are based on the phenotypic differences between alleles. Mutation – selection balance and balancing selection cannot properly explain the process of gene substitution, i.e. the molecular evolution of quantitative trait loci affecting fitness. I assume that such loci have non-essential functions (small effects on fitness), and that they have the potential to evolve into new functions and acquire new adaptations. Here I show that a high amount of neutral polymorphism at these loci can exist in real populations. Consistent with this, I propose a hypothesis for the maintenance of genetic variation in life history traits which can be efficient for the fixation of alleles with very small selective advantage. The hypothesis is based on neutral polymorphism at quantitative trait loci and both neutral and adaptive gene substitutions. The model of neutral – adaptive conversion (NAC) assumes that neutral alleles are not neutral indefinitely, and that in specific and very rare situations phenotypic (relative fitness) differences between them can appear. In this paper I focus on NAC due to phenotypic plasticity of neutral alleles. The important evolutionary consequence of NAC could be the increased adaptive potential of a population. Loci responsible for adaptation should be fast evolving genes with minimally discernible phenotypic effects, and the recent discovery of genes with such characteristics implicates them as suitable candidates for loci involved in adaptation.

1. Introduction

Life history traits (lifetime reproductive success, viability, fecundity, longevity, etc.) are closely related to fitness and are assumed to be under strong directional selection (Falconer & Mackay, 1996; Roff, 1997). Genetic variation in these traits is essential to the ability of organisms to adapt to new environments. A fundamental problem in evolutionary genetics is the question of how genetic variation in life history traits originates and is maintained in a population even though natural selection operating on these traits tends to quickly erode genetic variation (Fisher, 1930; Falconer & Mackay, 1996; Roff, 1997).

There are two levels at which genetic variation can be observed. One is molecular genetic variation, commonly measured as average heterozygosity per locus, and the second is quantitative genetic variation, measured as a fraction of the phenotypic variance of the trait (Falconer & Mackay, 1996). In terms of evolutionary potential, the most important fraction of phenotypic variance (V_P) is additive genetic variance (V_A) which is often expressed as heritability (h^2), that is the ratio of additive genetic variance to total phenotypic variance (V_A/V_P). For adaptive evolution to proceed according to the rules of quantitative genetics there must be additive genetic variation in a population such that $h^2>0$ (Falconer & Mackay, 1996; Roff, 1997). Life history traits usually have significantly lower heritabilities than non-fitness-related traits which are under weak stabilizing selection. This fact can be explained by three non-exclusive scenarios. First, a strong directional selection on life history traits quickly eliminates deleterious alleles, and consequently erodes additive genetic variance (Fisher, 1930; Gustaffson, 1986; Roff & Mousseau,

* Tel: +385 1 4561111. Fax: +385 1 4561177. e-mail: brcic@irb.hr

1987; Mousseau & Roff, 1987). Second, life histories have a higher level of residual variance (non-additive genetic and environmental) (Price & Schluter, 1991). And third, fitness traits can be canalized (buffered) against genetic and environmental perturbations (Waddington, 1942). Additive genetic variance can change through natural selection or through the exposure to a novel environment (phenotypic plasticity) (Roff, 1997; Lynch & Walsh, 1998).

Existing concepts for the maintenance of genetic variation in fitness traits are based on the phenotypic differences between alleles (mutation–selection balance and balancing selection) (Charlesworth & Hughes, 1999; Merila & Sheldon, 1999; Hughes & Burleson, 2000; Frankham *et al*., 2002). The problem with mutation–selection balance is in the fact that it is based on deleterious alleles at very low frequencies. Mutation–selection balance can be efficient in explaining the process of gene substitution only if one of these deleterious alleles acquires a relatively high selective advantage. Under overdominant balancing selection the alleles are at intermediate frequencies but such a polymorphism is stable (in reality long-lasting), and consequently overdominant balancing selection has a serious problem in explaining the process of gene substitution (Nei & Graur, 1984; Nei, 1987). In order to explain both polymorphism and gene substitution (molecular evolution) in allozymes Gillespie (1978, 1979) proposed the SAS-CFF model of selection in a random environment which belongs to the balancing selection concept. According to the SAS-CFF model, polymorphism and the rate of gene substitution are similar to that of the neutral model. In this paper, I propose a model which is based on neutral polymorphism at quantitative trait loci (QTLs), and on both adaptive and neutral gene substitutions. This model assumes that neutral alleles are not neutral indefinitely, and that in specific and very rare situations phenotypic (relative fitness) differences between them can appear. In other words, there is a very small (but non-zero) probability that neutral polymorphism can be converted into adaptive polymorphism. This conversion can be achieved by phenotypic plasticity (genotype by environment interaction, G×E) and by epistasis. These two mechanisms are different, and here I focus on neutral–adaptive conversion (NAC) due to phenotypic plasticity (G×E). The NAC model is able to explain gene substitution by the fixation of alleles with small selective advantage. Although it is widely accepted that neutral alleles are 'by definition' not relevant to adaptation (Frankham *et al*., 2002), there were some suggestions that they could be the raw material for adaptive evolution (Dykhuizen & Hartl, 1980; Kimura, 1986; Wagner, 2005). A prerequisite for the model described here is the presence of a substantial amount of neutral DNA polymorphism at QTLs affecting adaptive traits. Since there are no empirical data on polymorphism at QTLs, I use the equilibrium allele frequency distribution for the neutral model to show that in real populations a high level of polymorphism can exist, and that according to the NAC model a population may be able to increase its adaptation due to a higher probability for fixation of alleles with very small selective advantage.

2. Neutral polymorphism at QTLs

It is often remarked that most mutations in genes affecting fitness are deleterious (Frankham *et al*., 2002). This is true for genes with high selective constraints, whose products play important physiological and developmental roles. However, life history traits are also affected by many, possibly thousands of polygenes or QTLs, and most of these have very small effects on fitness (Houle *et al*., 1996; Merila & Sheldon, 1999; Adams *et al*., 2000; Venter *et al*., 2001). Under the term QTL I assume a single gene with several characteristics: (1) a QTL has a small effect on fitness; (2) it has a non-essential function, and has the potential to evolve into a new function or to acquire a new adaptation; and (3) each newly formed mutation at a QTL is neutral or deleterious while advantageous alleles are produced by the conversion of neutral into adaptive polymorphism. Since life histories are the most complex traits, it is reasonable to expect that they are affected by many more QTLs than non-fitness-related quantitative traits (Houle *et al*., 1996; Merila & Sheldon, 1999). QTLs as defined above probably operate under a weak selective constraint. The individual selection coefficient against each QTL is very small even though QTLs affect life history characters which are under strong selection. If a null mutation (or deletion) in such a QTL has a very small effect on fitness (selective disadvantage of several per cent), one would expect that a substantial number of mutations which cause non-synonymous nucleotide changes within such a QTL would be neutral (would not affect fitness). In other words, a relatively high fraction of newly arisen mutations per generation could be neutral. I suggest that neutral mutation in QTLs is crucial to the maintenance of genetic variation in a population.

The fraction of neutral mutations in genes evolving under the neutral model can be roughly estimated as the ratio of the rate of non-synonymous substitutions to synonymous substitutions within the gene, and this reflects the level of selective constraints. From the data of Li (1997) this fraction in mammals can be calculated, and for genes with the highest selective constraints these values are lowest, e.g. *histone 3* (0), *actin α* (0·003) and *ribosomal protein S17* (0·02). As selective constraint decreases, the fraction of neutral mutations increases, e.g. *α-globin* (0·13), *insulin*

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

C-peptide (0·22), *β-globin* (0·30), *relaxin* (0·41) and *apolipoprotein A-1* (0·41). In *Drosophila*, the average ratio of non-synonymous to synonymous substitution rates for the orphan genes (among them a high fraction are fast-evolving) is around 0·17, whereas for the non-orphan genes it is around 0·07 (Domazet-Lošo & Tautz, 2003). Similar values were obtained by Kimura (1983) based on a comparison of rare and common polymorphisms. He estimated that the fraction of neutral mutations for the substrate-specific enzymes of *Drosophila willistoni* is 0·07, and for the non-substrate–specific enzymes is 0·20, whereas in humans the values were 0·11 for substrate-specific and 0·43 for non-substrate-specific enzymes (Kimura, 1983).

It is reasonable to assume that QTLs are under similar (or weaker) selective constraints as the genes under weakest selective constraints mentioned above. Consequently, it is expected that neutral DNA polymorphism would be established within such QTLs. For a diploid random mating population with effective size N_e, let us consider a particular QTL under the infinite alleles model. This model assumes that each new mutation represents an allele that does not already exist in a population. Kimura & Crow (1964) have shown that the distribution of allele frequencies $\Phi(x)$ within a population in equilibrium where the production of new alleles by mutation is balanced with their extinction by random drift is

$$\Phi(x) = 4N_e v(1-x)^{4N_e v-1}x^{-1} \tag{1}$$

where v is the neutral mutation rate per QTL per generation, x is an allele frequency and N_e is the effective population size. The quantity $\Phi(x)dx$ represents the expected number of alleles within the frequencies x and $x+dx$. The homozygosity F per QTL is

$$F = \int_0^1 x^2\Phi(x)dx = \frac{1}{4N_e v+1} \tag{2}$$

and, because heterozygosity is $1-F$, the formula for heterozygosity H per QTL in equilibrium is (Kimura & Crow, 1964)

$$H = \frac{4N_e v}{4N_e v+1}. \tag{3}$$

In addition, useful parameters which measure the degree of polymorphism at a QTL are the effective number of alleles n_e and the actual (average) number of alleles n_a in a population. Since n_e is reciprocal to homozygosity F, it follows that (Kimura & Crow, 1964)

$$n_e = 4N_e v+1 = \frac{1}{1-H}. \tag{4}$$

The n_a can be calculated by numerical integration of formula (1) from $x=1/2N$ to $x=1$, where N is the actual population size (Kimura & Crow, 1964):

$$n_a = 4N_e v \int_{1/2N}^1 (1-x)^{4N_e v-1}x^{-1}dx. \tag{5}$$

It is essential to know the degree of polymorphism at a QTL that can be expected according to the neutral model in real populations. To establish this, I will use the estimates of N_e for wild populations based on protein polymorphism using equation (3) (Nei & Graur, 1984) and my estimates of v. First, I calculate a total mutation rate per QTL per generation u based on the fidelity of DNA replication *in vivo* and on the average number of germ cell divisions (replications) per generation. It is estimated that the average fidelity of DNA replication in *E. coli* is 6×10^{-10} substitutions per nucleotide per replication (Drake, 1991), and it is widely accepted that among eukaryotes it is 10^{-9} substitutions per nucleotide per replication (Maynard-Smith, 1989). In *Drosophila*, on the seventh day after eclosion, there are about 25 germ cell divisions (in both sexes) (Bauer & Aquadro 1997). We do not know the average size of a QTL, but for simplicity I will assume that it is 1000 nucleotides, which is less than the size of the average gene in *Drosophila* (1780 nucleotides) but closer to the size of the average gene without introns (Adams *et al*., 2000). This means that $u=2{\cdot}5\times10^{-5}$ substitutions per QTL per generation, a value in accordance with the known mutation rates among various taxa, which are between 10^{-6} and 10^{-4} mutations per locus per generation (Falconer & Mackay, 1996; Frankham *et al*., 2002). For rodents, where the average number of germ cell divisions in spermatogenesis and oogenesis is around 42 (Li, 1997), it follows that the total mutation rate is $u=10^{-9}\times42\times1000=4{\cdot}2\times10^{-5}$ substitutions per QTL per generation. In humans the average number of germ cell divisions of a 20-year-old man and woman is around 87 (Crow, 2000). In this case, the total mutation rate is $u=10^{-9}\times87\times1000=8{\cdot}7\times10^{-5}$ substitutions per QTL per generation. One can further check the reliability of mutation rate estimates for rodents and humans by comparing the total mutation rate per generation with the rate of nucleotide substitutions per year for neutral DNA (the level of synonymous substitutions within the gene) estimated from phylogenetic data between humans and rodents. The average mutation rate per locus per generation for humans and rodents converted to the rate of nucleotide substitutions per year is $6{\cdot}29\times10^{-9}$, which is in agreement with the phylogenetic data for the highest rates of synonymous substitution per nucleotide per year, e.g. $5{\cdot}16\times10^{-9}$ (*albumin*), $5{\cdot}50\times10^{-9}$ (*interferon* γ), $5{\cdot}56\times10^{-9}$ (*imunoglobulin* k) and $6{\cdot}39\times10^{-9}$ (*relaxin*) (Li, 1997).

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

Table 1. *Expected neutral polymorphism at QTLs in wild* Drosophila *populations*

Fraction of neutral mutation per QTL	Size of QTL	H	n_e	n_a (0·01–0·99)	n_a ($1/2N_e-1$)
0·1	500	0·833	5·99	12·81	62·13
	750	0·882	8·49	16·09	89·89
	1000	0·909	10·99	18·64	116·80
0·2	500	0·909	10·99	18·64	116·80
	750	0·937	16·00	22·34	168·86
	1000	0·952	21·01	24·78	219·22
0·3	500	0·937	16·00	22·34	168·86
	750	0·957	23·47	25·68	243·91
	1000	0·968	30·96	27·43	316·41
0·4	500	0·952	21·01	24·78	219·22
	750	0·968	30·96	27·43	316·41
	1000	0·976	40·98	28·30	410·20
0·5	500	0·961	25·97	26·40	268·31
	750	0·974	38·46	28·20	387·02
	1000	0·980	51·02	28·30	501·47

Data are calculated using equations (3–5). Estimation of mutation rates is based on the fidelity of DNA replication *in vivo* of 10^{-9} substitutions per nucleotide per replication and 25 germ cell divisions per gamete. The effective population size used for estimation of the degree of polymorphism is 10^6. Calculation of actual numbers of alleles was done using Mathematica 4 software (Wolfram Research).
Abbreviations: H, heterozygosity per QTL; n_e, effective number of alleles per QTL; n_a (0·01–0·99), actual number of alleles per QTL (common polymorphism); n_a ($1/2N_e-1$), actual number of alleles per QTL (total polymorphism).

Table 2. *Expected neutral polymorphism at QTLs in wild rodent populations*

Fraction of neutral mutation per QTL	Size of QTL	H	n_e	n_a (0·01–0·99)	n_a ($1/2N_e-1$)
0·1	500	0·456	1·84	4·10	10·50
	750	0·557	2·26	5·35	14·92
	1000	0·627	2·68	6·45	19·21
0·2	500	0·627	2·68	6·45	19·21
	750	0·716	3·52	8·39	27·51
	1000	0·771	4·36	10·07	35·53
0·3	500	0·716	3·52	8·39	27·51
	750	0·791	4·78	10·83	39·45
	1000	0·834	6·04	12·87	50·97
0·4	500	0·771	4·36	10·07	35·53
	750	0·834	6·04	12·87	50·97
	1000	0·870	7·72	15·16	65·86
0·5	500	0·808	5·20	11·54	43·33
	750	0·863	7·30	14·62	62·18
	1000	0·894	9·40	17·08	80·32

Data are calculated using equations (3–5). Estimation of mutation rates is based on the fidelity of DNA replication *in vivo* of 10^{-9} substitutions per nucleotide per replication and 42 germ cell divisions per gamete. The effective population size used for estimation of the degree of polymorphism is 10^5. Calculation of actual numbers of alleles was done using Mathematica 4 software (Wolfram Research).
Abbreviations: H, heterozygosity per QTL; n_e, effective number of alleles per QTL; n_a (0·01–0·99), actual number of alleles per QTL (common polymorphism); n_a ($1/2N_e-1$), actual number of alleles per QTL (total polymorphism).

The neutral mutation rate is $v=f\,u$, where f is the neutral mutation fraction. Since I do not know the exact f for particular QTLs, I estimated the degree of neutral polymorphism using several values for f (0·1, 0·2, 0·3, 0·4 and 0·5), assuming QTLs in the size range of 500, 750 and 1000 nucleotides. I calculated H, n_e, n_a (common polymorphism at the range of allele frequencies from 0·01 to 0·99) and n_a (total polymorphism at the range of allele frequencies from $1/2N_e$ to 1) for *Drosophila* (Table 1), rodents (Table 2) and humans (Table 3). The N_e for *Drosophila* is 10^6 (Nei & Graur, 1984) and using equations (3–5) it follows that a high degree of DNA polymorphism is expected even if the neutral mutation fraction f is 0·1

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

Table 3. *Expected neutral polymorphism at QTLs in human population*

Fraction of neutral mutation	Size of QTL	H	n_e	n_a (0·01–0·99)	n_a ($1/2N_e-1$)
0·1	500	0·15	1·17	1·31	2·68
	750	0·21	1·26	1·80	3·49
	1000	0·26	1·35	2·24	4·29
0·2	500	0·26	1·35	2·24	4·29
	750	0·34	1·52	2·98	5·84
	1000	0·41	1·70	3·62	7·35
0·3	500	0·34	1·52	2·98	5·84
	750	0·44	1·78	3·91	8·09
	1000	0·51	2·04	4·73	10·27
0·4	500	0·41	1·70	3·62	7·35
	750	0·51	2·04	4·73	10·27
	1000	0·58	2·39	5·71	13·08
0·5	500	0·47	1·87	4·19	8·82
	750	0·57	2·30	5·47	12·38
	1000	0·63	2·74	6·60	15·81

Data are calculated using equations (3–5). Estimation of mutation rates is based on the fidelity of DNA replication *in vivo* of 10^{-9} substitutions per nucleotide per replication and 87 germ cell divisions per gamete. The effective population size used for estimation of the degree of polymorphism is 10^4. Calculation of actual numbers of alleles was done using Mathematica 4 software (Wolfram Research).
Abbreviations: H, heterozygosity per QTL; n_e, effective number of alleles per QTL; n_a (0·01–0·99), actual number of alleles per QTL (common polymorphism); n_a ($1/2N_e-1$), actual number of alleles per QTL (total polymorphism).

(Table 1). If f is 0·2, a value reasonable for *Drosophila* according to the data mentioned earlier, the expected H is between 0·91 and 0·95, the expected n_e is 11–21 alleles, n_a (common) is 19–25 alleles and n_a (total) is 117–219 alleles. The N_e for most rodent populations is 10^5 (Nei & Graur, 1984) and the level of expected polymorphism is also high. For values of f between 0·3 and 0·5 the n_a (common) is 8–17 alleles and n_a (total) is 28–80 alleles (Table 2). For humans N_e is 10^4 (Nei & Graur, 1984), and consequently the level of polymorphism is much lower even though humans have a higher mutation rate per locus per generation than rodents and flies. For the values of f between 0·3 and 0·5 the n_a (common) is 3–7 alleles and n_a (total) is 6–16 alleles (Table 3). According to the data in Tables 1–3, it is obvious that under the infinite alleles model with realistic neutral mutation rates and effective population sizes, a high level (*Drosophila* and rodents) and moderate level (humans) of DNA polymorphism at QTLs can be established.

3. Neutral–adaptive conversion and its consequence on gene substitution

(i) *The neutral–adaptive conversion (NAC) hypothesis*

The question is: What is the adaptive significance of neutral polymorphism? In the existing environment it has no adaptive relevance and the population can be close to the selective optimum. However, neutral polymorphism could be relevant for adaptive evolution in a particular novel environment or in a novel genetic background if it is converted into additive genetic variance. In this work I focus on the phenotypic plasticity of neutral alleles ($G \times E$) as the mechanism for such conversion. Additive genetic variance due to unconditionally deleterious alleles is not relevant for adaptation, and is ignored. Only the additive genetic variance which arises as the result of the appearance of advantageous alleles is relevant for adaptation. The classical model assumes that the allele becomes advantageous at the time of its formation or when it is at a very low frequency under mutation–selection balance. According to the model presented here, the additive genetic variance responsible for adaptation is the result of neutral–adaptive conversion (NAC) due to phenotypic plasticity. Let us assume that at any time there is a very small but non-zero probability for the conversion of neutral into adaptive polymorphism at a single QTL. When this conversion occurs under the simplest scenario, one allele of a particular QTL has the highest contribution to fitness in a new environment, and the rest of the alleles within the polymorphism have the same deleterious effect on fitness. Let us assume that alleles act additively within and among loci. We will designate the advantageous allele as A_2 and deleterious alleles as A_1. All deleterious alleles after NAC can be treated as the same allelic type since they show the same phenotype. Under this assumption the fitnesses of the genotypes are A_1A_1 (1), A_1A_2 $(1+s)$, and A_2A_2 $(1+2s)$, where s is the selection coefficient of the advantageous allele. I consider as a life history trait the lifetime reproductive success, which is the best approximation of fitness. The contribution of

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

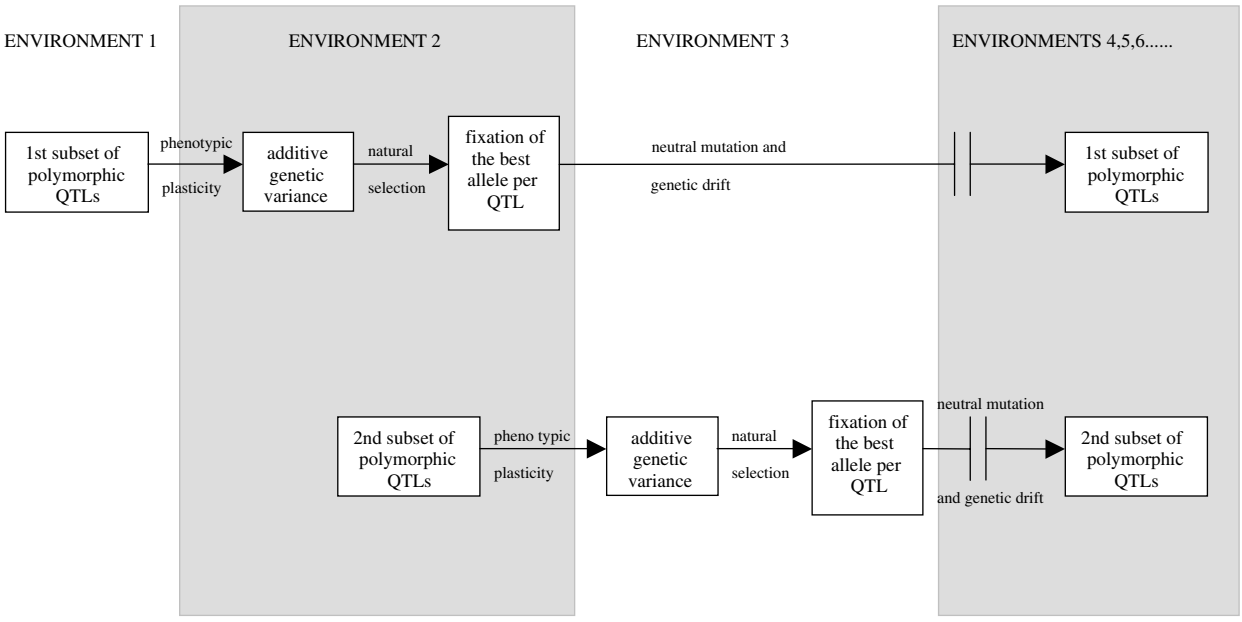

Fig. 1. NAC model for the maintenance of adaptive genetic variation. It is assumed that fast-evolving genes (under low selective constraint) could be QTLs responsible for adaptation. In this case, it is reasonable to estimate that the genome of higher eukaryotes consists of several thousands QTLs. The important assumption of the model is that phenotypic plasticity (genotype by environment interaction) occurs very rarely per a single QTL, and that it will be restricted to a small subset of polymorphic QTLs. This will lead to the sequential release of additive genetic variance, and consequently to adaptive substitutions.

a single QTL to additive genetic variance for fitness after NAC can be expressed in similar form as for the two-allele model (Falconer & Mackay, 1996):

$$V_{Aw} = 2x(1-x)s^2 \tag{6}$$

where x is the frequency of advantageous allele at the time of neutral–adaptive conversion, and $1-x$ is the total frequency of all deleterious alleles (the rest of alleles within polymorphism). If several, i.e. n, QTLs contribute to additive genetic variance responsible for adaptation then

$$V_{Aw} = 2\sum_{i=1}^{n} x_i(1-x_i)s_i^2. \tag{7}$$

Additive genetic variance will change according to the increase in the advantageous allele frequency under selection. If x is below the optimal value of 0·5, then additive variance will initially increase, and when x reaches the optimal value it will finally decrease. Alternatively, the advantageous allele (especially if it is of low frequency) can be lost by random drift. What will happen depends on the probability of fixation of the advantageous allele which we will consider in connection with the process of gene substitution. Since both scenarios require relatively short time periods and the probability for NAC is assumed to be very low, most of the time a life history trait will have a low heritability. This low heritability of a typical life history trait in the equilibrium natural population can be explained by mutation–selection balance and some forms of balancing selection since neutral alleles at QTLs do not contribute to V_A and heritability.

A model for the maintenance of genetic variation in life history traits based on conversion of neutral into adaptive polymorphism assumes: (1) a constant and relatively high neutral mutation rate per QTL per generation across environments; (2) a very low probability for phenotypic plasticity of neutral alleles per QTL; (3) existence of a large number of QTLs affecting a life history trait; and (4) a constant and large effective population size. According to the scheme shown in Fig. 1, a single polymorphic QTL (or several polymorphic QTLs) is plastic to environment 2 and contributes to additive genetic variance and heritability. Subsequently, natural selection erodes the additive genetic variance and heritability. The consequence is a change in the mean phenotype (fitness) of the population. The vast majority of QTLs do not contribute to additive genetic variance in environment 2 and are polymorphic. Another QTL (or another subset of them) will interact with environment 3 and the process of neutral–adaptive conversion and the fixation of the best allele(s) by natural selection will be repeated. In the meantime, the first QTL (or the first subset of QTLs) will be regenerating DNA polymorphism in these ever-changing environments. While these QTLs are monomorphic, additional polymorphic QTLs (or subsets of polymorphic QTLs) will cope with environmental changes if nesessary. It

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

is reasonable to expect that at any time the genome consists of polymorphic QTLs which are possibly the majority of loci, and monomorphic QTLs (loci of recent selective sweeps) (Maynard-Smith & Haigh, 1974) which are recovering neutral polymorphism. It is important to note that selective advantage or neutrality of alleles is conditional, and is dependent on the particular environment.

(ii) *Consequence on gene substitution*

The rate of gene substitution α is generally dependent on the effective population size N_e, on the selection coefficient (advantage) of an allele s, and on the probability of the fixation of an allele P. Under the neutral model, which is the special case, the rate of gene substitution $\alpha_n = v$, where v is the neutral mutation rate per locus per generation. On the other hand, when the allele has a selective advantage the rate of gene substitution $\alpha_a = 4N_e s v_a$, where v_a is the advantageous mutation rate per locus per generation. Both expressions for the rate of gene substitutions were obtained using the general formula for the probability of the fixation of an allele for genic selection (Kimura, 1962):

$$P(x) = \frac{1 - e^{-4N_e sx}}{1 - e^{-4N_e s}} \tag{8}$$

assuming that the initial allele frequency $x = 1/(2N_e)$, and that the actual number of individuals $N = N_e$. It was mentioned earlier that the NAC model assumes that there is a constant and very low (but non-zero) probability r for the conversion of neutral into adaptive polymorphism. According to the NAC model, the rate of gene substitution at a single QTL is

$$\alpha = v(1 - r) + 2N_e vrP^* \tag{9}$$

where P^* is the average probability for the fixation of an advantageous allele from the previously neutral polymorphism. It is intuitively clear that P^* will be larger than the probability for the fixation of an allele which is advantageous at the time of its formation, as is considered in the classical model where $x = 1/(2N_e)$, and the corresponding $P(x)$ is approximately $2s$. Within the previously neutral polymorphism there is a high range of allelic frequencies, and the probability for the fixation of a particular allele depends on its frequency. Which allele will show selective advantage in a new environment is unpredictable, since this is a completely random event, independent of allele frequency. Consequently, the average probability for the fixation of an allele which will contribute to the rate of gene substitution at a QTL in the NAC model is

$$P^* = \frac{1}{n_a} \sum_{i=1}^{n_a} \frac{1 - e^{-4N_e sx_i}}{1 - e^{-4N_e s}} \tag{10}$$

where n_a is the actual number of alleles within the polymorphism. This probability can be written in integral form as

$$P^* = \frac{\int_{1/2N}^{1} \Phi(x)P(x)dx}{\int_{1/2N}^{1} \Phi(x)dx} \tag{11}$$

where $\Phi(x)$ is the distribution of allele frequencies presented in expression (1) and $P(x)$ is the probability of the fixation of a single allele with initial frequency x, as is shown in formula (8).

Using formula (11) I calculate P^* for some data from Table 2, when $N = N_e = 10^5$, and $v = 1{\cdot}68 \times 10^{-5}$ (this mutation rate corresponds to a QTL of 1000 nucleotides when the fraction of neutral mutation per locus is 0·4). If selective advantage is $s = 10^{-3}$, it follows that the average probability for the fixation of an allele from the previously neutral polymorphism is $P^* = 0{\cdot}426$. If the initial allele frequency is $1/(2N_e) = 5 \times 10^{-6}$, then the probability for the fixation of allele P_0 is just $2s = 0{\cdot}002$. This means that in this particular case the probability of fixation of the advantageous allele in the NAC model is 200 times higher than in the classical model. When $s = 10^{-2}$, the probability of fixation of the allele in the NAC model is $P^* = 0{\cdot}658$, which is more than 30 times greater than the probability of fixation of the allele in the classical model ($P_0 = 0{\cdot}02$). Finally, even when the allele has a relatively high selective advantage of $s = 10^{-1}$, the probability of its fixation in the NAC model ($P^* = 0{\cdot}875$) is more than 4 times greater than that in the classical model ($P_0 = 0{\cdot}2$). The dependence of the P^*/P_0 ratio on various s values is shown in Fig. 2. The P^*/P_0 ratio decreases as s increases, and finally approaches unity. The important evolutionary consequence of this feature of the NAC model is that a considerable number of mutations with very small selective advantage could be fixed during the evolution, whereas these mutations would be lost due to genetic drift according to the classical model. Consequently, the NAC mechanism would increase the adaptive potential of the population. If we assume that QTLs are genes with small effects, it is reasonable to expect that most of their advantageous alleles would have a very small selective advantage. The NAC model also suggests that such QTLs should be fast-evolving genes, and that they are not easily discernible at the phenotypic level.

In the NAC model both types of substitutions (neutral and adaptive) are possible if r is very small. On the other hand, according to mutation–selection balance (which is based on deleterious mutations, and a negligible fraction of neutral mutations), gene substitution is expected to occur predominantly with newly

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

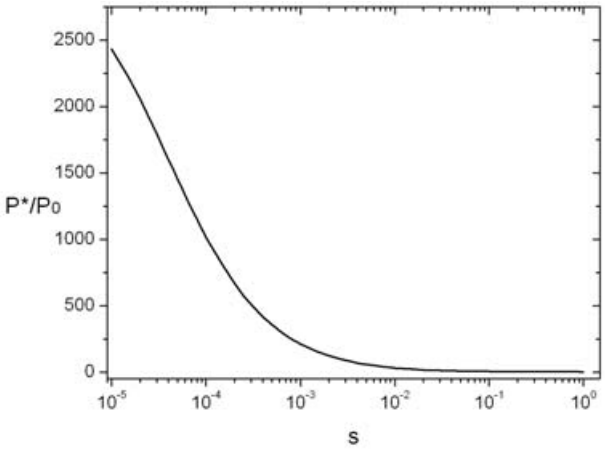

Fig.2. The relationship between the P^*/P_0 ratio and the selection coefficient s. The P^* is the probability for the fixation of an allele in the NAC model and P_0 is the probability for the fixation of an allele in the classical model. The P^*/P_0 ratio decreases as s increases, and finally approaches unity. The curve corresponds to the following data: $N_e = 10^5$ and $v = 1{\cdot}68 \times 10^{-5}$ (this mutation rate is applicable to a QTL of 1000 nucleotides when the fraction of neutral mutations per locus is 0·4). The P^*/P_0 ratio is calculated using Mathematica 4 software (Wolfram Research).

born (or recently formed previously deleterious) advantageous mutations. Mutation–selection balance considers together all of the existing deleterious alleles of a particular locus with similar phenotype (Haldane, 1927; Orr & Betancourt, 2001), or even all the alleles from many loci affecting complex traits such as life histories. In agreement with this, there is a considerable amount of genetic variation in life history traits found in natural populations (Charlesworth & Hughes, 1999; Hughes & Burleson, 2000). However, I assume that a single previously deleterious allele in mutation–selection balance becomes advantageous in a new environment. This seems more realistic than that all deleterious alleles become beneficial. If we consider a particular deleterious allele (a single molecular variant) it would be expected that its frequency is very low, possibly close to $1/(2N_e)$ at the time when it becomes advantageous, and consequently it has a very small chance to become fixed. For example, we can assume that a QTL consists of 1000 nucleotides and that the deleterious mutation rate per QTL per generation is 10^{-5}. The equilibrium deleterious allele frequency x under mutation–selection balance can be approximately expressed as $x = v_d/hs_d$ (Haldane, 1927), where v_d is the deleterious mutation rate, h is the dominance coefficient and s_d is the selection coefficient against the mutant homozygote. According to the above assumption, the applicable mutation rate would be much closer to 10^{-8} (per nucleotide) than to 10^{-5} (per QTL). If we allow that mutations with the potential to become advantageous in a new environment are restricted to 1% (10 nucleotides) of the whole mutation target (1000 nucleotides), then the applicable mutation rate would be 10^{-7}. For the mutation rate of $v_d = 10^{-7}$ and $s_d = 0{\cdot}05$ when $h = \frac{1}{2}$ (no dominance), the equilibrium allele frequency is $x = 4 \times 10^{-6}$ and for $h = \frac{1}{4}$ (partial dominance), $x = 8 \times 10^{-6}$. Both these values are close to the minimal allele frequency in a population of $N_e = 10^5$, which is $x = 5 \times 10^{-6}$. If s_d is very small, much smaller than 0·05 (which is considered to be the typical homozygous mutant fitness reduction in *Drosophila*) (Charlesworth & Hughes, 1999), then the behaviour of such alleles is similar to the behaviour of neutral alleles. Also, there is no reason to believe that deleterious mutations have a better chance to become advantageous in future environmental changes than a non-synonymous neutral mutations.

On the contrary, balancing (overdominant) selection can generate a high DNA polymorphism, even higher than is predicted by the neutral theory. However, overdominant selection polymorphism is stable (not transient), and consequently it decreases the rate of nucleotide substitution in evolution (Nei, 1980; Maruyama & Nei, 1981). Higher rates of molecular evolution of QTLs could be achieved by combining overdominant balancing selection and positive selection, if balancing selection was short-lived due to environmental changes. However, under these circumstances, as well as in the case of combining positive selection and mutation–selection balance, the majority of substitutions in evolution would be adaptive. In the case of the SAS-CFF model, the rate of gene substitution is similar to that predicted by the neutral model (Gillespie, 1979). Recent genomic data show that on average around 35% of all amino acid substitutions per gene in primates are adaptive (Fay *et al.*, 2001), whereas in *Drosophila* this figure is around 45% (Smith & Eyre-Walker, 2002; Fay *et al.*, 2002). This implies that a majority of substitutions are random or neutral, but the rate of gene substitution is higher than expected from the neutral and SAS-CFF models. This result can be satisfactorily explained by the NAC hypothesis.

From the above empirical data, we can take that 40% of all substitutions are adaptive (average value for genes of primates and *Drosophila*). Under the assumption that maintenance of genetic variation and gene substitutions during evolution occurred by a NAC mechanism, we can calculate the probability for the neutral–adaptive conversion r (NAC probability). The proportion of adaptive substitutions a can be obtained as the ratio of the rate of adaptive gene substitution and the total rate of gene substitution in the NAC model,

$$a = \frac{2N_e vrP^*}{v(1-r) + 2N_e vrP^*}. \qquad (12)$$

https://doi.org/10.1017/S0016672305007688 Published online by Cambridge University Press

After algebraic rearrangements, r can be expressed from the above equation as

$$r = \frac{a}{2N_e P^*(1-a)+a}. \quad (13)$$

From the values of $a = 0{\cdot}4$; $N_e = 10^5$; $P^* = 0{\cdot}426$ (when $s = 10^{-3}$ for a QTL of 1000 nucleotides with neutral mutation fraction of 0·4 as shown in Table 2), it implies that the probability for the conversion of neutral into adaptive polymorphism is $r = 8 \times 10^{-6}$. This is really a low probability, and it means that NAC is very difficult to detect in a real experiment. Also, the possibility for detection of NAC is restricted for cases of reasonable phenotypic differences within a relatively short time period before natural selection erodes additive genetic variance. Consequently, this is in agreement with the fact that populations founded in the laboratory from recently captured samples from nature still show low (but non-zero) heritabilities for life history traits.

4. Some empirical and experimental arguments

Functional genomics has shown the existence of rapidly evolving genes with minimally discernible phenotypic effects. It is estimated through database searches and low stringency cross-hybridization studies that one-third of all genes in *Drosophila* are fast-evolving genes (Schmid & Tautz, 1997). These authors also show that the fastest-evolving gene isolated in such a screen had a very high degree of population polymorphism. Additionally, of particular interest are the genes called orphans which have no homologues in distantly related species. It is known that on average *Drosophila* orphan genes evolve more than 3 times faster than other genes, and mutations in most of these have no detectable phenotypes (Domazet-Lošo & Tautz, 2003). It is important to note that among orphan genes there are some that are slow-evolving, and these may be lineage-specific genes which recently acquired a new function and are subject to increased selective constraint (Domazet-Lošo & Tautz, 2003). In yeast, at least one-third of all genes have neither established nor inferred functions and could be candidate QTLs with a marginal effect on fitness. Indeed, it was found that such non-essential genes when completely inactivated have very small but significant fitness contributions, and that some have conditional phenotypes (Smith *et al*., 1996; Thatcher *et al*., 1998). QTL-mapping studies for several adaptive traits on experimental populations of *Drosophila* (Leips & Mackay, 2000) and maize (Jiang *et al*.,1999) have shown that QTLs identified under one environment no longer act as QTLs in a different environment, demonstrating QTL–environment interactions and conditional neutrality of some alleles. Also, it was estimated that a large fraction of

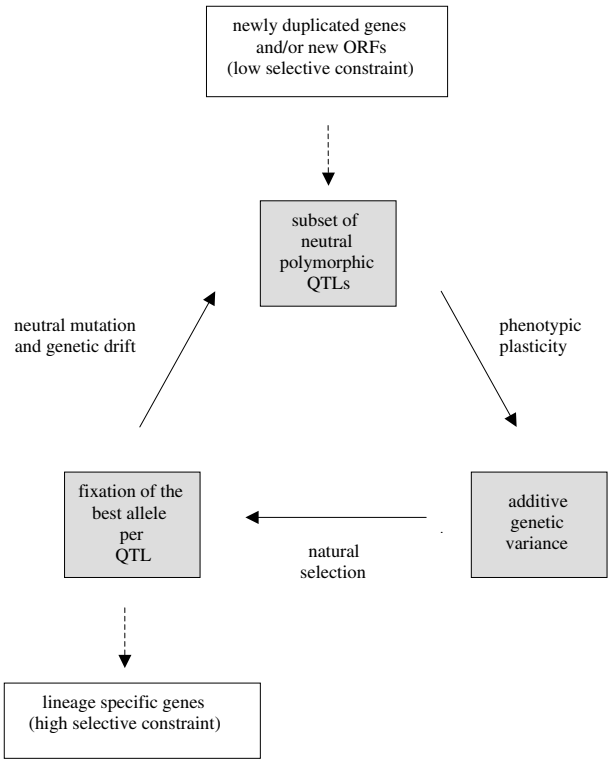

Fig. 3. Hypothetical adaptive evolutionary cycle. The final outcome of fast-evolving adaptive QTLs could be lineage-specific genes, and the raw material for fast evolution could be newly created genes. It is important to note that neutral polymorphism is transient, and that gene substitution at a QTL can be achieved by random fixation of a neutral allele if phenotypic plasticity does not occur during the polymorphic phase.

mutations in *Drosophila* are conditionally neutral (Kondrashov & Houle, 1994). There is also evidence for conditional neutrality due to epistasis. Lauter & Doebley (2002) have shown that some alleles show phenotypic effects in a hybrid genetic background (between teosinte and maize) on traits that are invariant in teosinte. Another example is from *Drosophila* eye development, where extensive genetic variation was revealed by introgression of mutant alleles of the epidermal growth factor receptor (*Egfr*) (Polacysk *et al*., 1998).

Finally, it was shown that *Drosophila* and *Arabidopsis* populations contain a surprising amount of cryptic genetic variation which can be uncovered under stressful environmental conditions by the inactivation of heat-shock protein Hsp90 buffering function (Rutherford & Lindquist, 1998; Queitsch *et al*., 2002; Gibson & Dworkin, 2004). This suggests that one possible mechanism of phenotypic plasticity could be connected with a change in chaperone function required for the maturation of many regulatory proteins involved in the regulation of gene expression and signal transduction pathways. It is important to note that this is not a general example of cryptic

genetic variation. The release of neutral variation due to environmental stress or epistasis is not necessarily connected with canalization (Hermisson & Wagner, 2004).

5. Evolutionary implications

The NAC hypothesis has several important evolutionary consequences. First, it can explain the low heritability of life histories in natural populations (Fisher, 1930; Gustaffson, 1986; Mousseau & Roff, 1987; Roff & Mousseau, 1987), and the ability of such populations to adapt to future environmental changes. This is based on the fact that increased heritability due to NAC lasts for a relatively short time period, until the fixation of advantageous alleles. Also, it predicts a high level of DNA polymorphism at QTLs (if N_e is 10^5 or higher) which can be essential for the final fate of an advantageous allele. Consequently, the NAC model could increase the adaptation of a population by the efficient fixation of the alleles with very small selective advantage. In addition, the important feature of neutral polymorphism is that the identity of alleles is always changing. This is crucial since the environmental change itself is not predictible. On the other hand, balancing selection polymorphism is not dynamic, and most such alleles are very old, even older than the species (Maruyama & Nei, 1981). Finally, the most important phenomenon on a long-term time scale is evolutionary innovation. It is much more likely that such new qualities in evolution can be produced by fast-evolving genes which accumulate sufficient substitutions during evolutionary time such that the final mutation is critical for creating the new quality or function. Such genes with newly acquired functions can become subject to increased selective constraint, at which point they are no longer QTLs with small effects on fitness. On the other hand, new QTLs can arise from newly duplicated genes and/or new open reading frames (ORFs) which are under weak selective constraint (Fig. 3).

According to the neutral theory, and assuming a large number of QTLs, it appears that the total amount of DNA polymorphism in a population could be impressive. Neutral mutations were originally introduced into evolutionary genetics to explain Haldane's dilemma (Haldane, 1957) regarding the genetic load that can be tolerated by a species (Kimura, 1968). Such neutral mutations were placed on the non-coding DNA and some coding DNA with no adaptive significance (Frankham *et al*., 2002). The NAC hypothesis suggests that neutral mutations and random genetic drift could have significant adaptive relevance. Their role could be to produce and accumulate neutral DNA polymorphism at QTLs which can be converted into additive genetic variance by phenotypic plasticity, and then used by natural selection to produce adaptive evolutionary change.

I thank Dijana Žilić, Igor Weber, Bojan Antolović, Hrvoje Šarčević and Miljenko Huzak for helpful discussions, and Mary Sopta and Tomislav Domazet-Lošo for critical reading of the manuscript. This work was supported by the Croatian Ministry of Science (grant 0098 070).

References

Adams, M. D., *et al*. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Bauer, V. L. & Aquadro, C. F. (1997). Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. *Molecular Biology and Evolution* **14**, 1252–1257.

Charlesworth, B. & Hughes, K. A. (1999). The maintenance of genetic variation in life-history traits. In *Evolutionary Genetics: From Molecules to Morphology* (ed. R. S. Singh & C. B. Krimbas), pp. 369–392. Cambridge: Cambridge University Press.

Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics* **1**, 40–47.

Domazet-Lošo, T. & Tautz, D. (2003). An evolutionary analysis of orphan genes in *Drosophila*. *Genome Research* **13**, 2213–2219.

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the USA* **88**, 7160–7164.

Dykhuizen, D. & Hartl, D. L. (1980). Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. *Genetics* **96**, 801–817.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Harlow, UK: Longman.

Fay, J., Wyckoff, G. J. & Wu, C.-I. (2001). Positive and negative selection on the human genome. *Genetics* **158**, 1227–1234.

Fay, J. C., Wyckoff, G. J. & Wu, C.-I. (2002). Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Frankham, R., Ballou, J. D. & Briscoe, D. A. (2002). *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

Gibson, G. & Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nature Reviews Genetics* **3**, 681–690.

Gillespie, J. H. (1978). A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theoretical Population Biology* **14**, 1–45.

Gillespie, J. H. (1979). Molecular evolution and polymorphism in a random environment. *Genetics* **93**, 737–754.

Gustaffson, L. (1986). Lifetime reproductive succes and heritability: empirical support for Fisher's fundamental theorem. *American Naturalist* **128**, 761–764.

Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection. 5. Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 838–844.

Haldane, J. B. S. (1957). The cost of natural selection. *Journal of Genetics* **55**, 511–524.

Hermisson, J. & Wagner, G. P. (2004). The population genetic theory of hidden variation and genetic robustness. *Genetics* **168**, 2271–2284.

Houle, D., Morikawa, B. & Lynch, M. (1996). Comparing mutational variabilities. *Genetics* **143**, 1467–1483.

Hughes, K. A. & Burleson, M. H. (2000). Evolutionary causes of variation in fertility and other fitness components. In *Genetic Influences on Human Sexuality and Fertility* (ed. J. L. Rodgers, D. C. Rowe & W. Miller), pp. 7–34. Dordrecht, The Netherlands: Kluwer Academic.

Jiang, C., Edmeades, G. O., Armstead, I., Lafitte, H. R., Hayward, M. D. & Hoisington, D. (1999). Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers. *Theoretical and Applied Genetics* **99**, 1106–1119.

Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624–626.

Kimura, M. (1983). Rare variant alleles in the light of the neutral theory. *Molecular Biology and Evolution* **1**, 84–93.

Kimura, M. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London, Series B* **312**, 343–354.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kondrashov, A. S. & Houle, D. (1994). Genotype–environment interactions and the estimation of genomic mutation rate in *Drosophila melanogaster*. *Proceedings of the Royal Society of London, Series B* **258**, 221–227.

Lauter, N. & Doebley, J. (2002). Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. *Genetics* **160**, 333–342.

Leips, J. & Mackay, T. F. (2000). Quantitative trait loci for life span in *Drosophila melanogaster*: interactions with genetic background and larval density. *Genetics* **155**, 1773–1788.

Li, W. H. (1997). *Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Maruyama, T. & Nei, M. (1981). Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* **98**, 441–459.

Maynard-Smith, J. (1989). *Evolutionary Genetics*. Oxford: Oxford University Press.

Maynard-Smith J. & Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetical Research* **23**, 23–35.

Merila, J. & Sheldon, B. C. (1999). Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity* **83**, 103–109.

Mousseau, T. A. & Roff, D. A. (1987). Natural selection and the heritability of fitness components. *Heredity* **59**, 181–197.

Nei, M. (1980). Stochastic theory of population genetics and evolution. In *Vito Volterra Symposium on Mathematical Models in Biology* (ed. C. Barigozzi), pp. 17–47. Berlin: Springer.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M. & Graur, D. (1984). Extent of protein polymorphism and the neutral mutation theory. *Evolutionary Biology* **17**, 73–118.

Orr, H. A. & Betancourt, A. J. (2001). Haldane's sieve and adaptation from the standing genetic variation. *Genetics* **157**, 875–884.

Polacysk, P. J., Gasperini, R. & Gibson, G. (1998). Naturally occuring genetic variation affects *Drosophila* photoreceptor determination. *Development, Genes and Evolution* **207**, 462–470.

Price, T. & Schluter, D. (1991). On the low heritability of life-history traits. *Evolution* **45**, 853–861.

Queitsch, C., Sangster, T. A. & Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics*. New York: Chapman and Hall.

Roff, D. A. & Mousseau, T. A. (1987). Quantitative genetics and fitness: lessons from *Drosophila*. *Heredity* **58**, 103–118.

Rutherford, S. L. & Lindquist, S. (1998). Hsp90 as capacitor for morphological evolution. *Nature* **396**, 336–342.

Schmid, K. J. & Tautz, D. (1997). A screen for fast evolving genes from *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **94**, 9746–9750.

Smith, N. G. C. & Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024.

Smith, V., Chou, K. N., Lashkari, D., Botstein, D. & Brown, P. O. (1996). Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274**, 2069–2074.

Thatcher, J. W., Shaw, J. M. & Dickinson, W. J. (1998). Marginal fitness contributions of nonessential genes in yeast. *Proceedings of the National Academy of Sciences of the USA* **95**, 253–257.

Venter, J. C., *et al*. (2001). The sequence of the human genome. *Science* **291**, 1304–1351.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565.

Wagner, A. (2005). Robustness, evolvability, and neutrality. *FEBS Letters* **579**, 1772–1778.