Genet. Res., Camb. (2009), **91**, *pp.* 267–280. © Cambridge University Press 2009
doi:10.1017/S0016672309000196 Printed in the United Kingdom

Enhanced fixation and preservation of a newly arisen duplicate gene by masking deleterious loss-of-function mutations

KENTARO M. TANAKA[1], K. RYO TAKAHASI[2] AND TOSHIYUKI TAKANO-SHIMIZU[1,2,3,4]*

[1] *Department of Genetics, Graduate University for Advanced Studies (SOKENDAI), Mishima, Shizuoka 411-8540, Japan*
[2] *Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan*
[3] *Department of Biosystems Science, Graduate University for Advanced Studies (SOKENDAI), Hayama, Kanagawa 240-0193, Japan*
[4] *Department of Biological Science, Graduate School of Science, The University of Tokyo, Tokyo 113-8654, Japan*

(*Received 12 December 2008 and in revised form 2 March 2009*)

Summary

Segmental duplications are enriched within many eukaryote genomes, and their potential consequence is gene duplication. While previous theoretical studies of gene duplication have mainly focused on the gene silencing process after fixation, the process leading to fixation is even more important for segmental duplications, because the majority of duplications would be lost before reaching a significant frequency in a population. Here, by a series of computer simulations, we show that purifying selection against loss-of-function mutations increases the fixation probability of a new duplicate gene, especially when the gene is haplo-insufficient. Theoretically, the probability of simultaneous preservation of both duplicate genes becomes twice the loss-of-function mutation rate (u_c) when the population size (N), the degree of dominance of mutations (h) and the recombination rate between the duplicate genes (c) are all sufficiently large ($Nu_c > 1$, $h > 0{\cdot}1$ and $c > u_c$). The preservation probability declines rapidly with h and becomes 0 when $h = 0$ (haplo-sufficiency). We infer that masking deleterious loss-of-function mutations give duplicate genes an immediate selective advantage and, together with effects of increased gene dosage, would predominantly determine the fates of the duplicate genes in the early phase of their evolution.

1. Introduction

Segmental duplications have received growing attention in the last few years and, indeed, they are enriched within the human and other mammalian genomes (e.g. Bailey *et al.*, 2002, 2004; Gu *et al.*, 2002; Samonte & Eichler, 2002; She *et al.*, 2004; Cheng *et al.*, 2005). There is also emerging evidence that copy-number variation, generated by duplications and deletions of DNA segments that are 1 kb or larger in size, is abundant throughout human and Drosophila genomes (e.g. Iafrate *et al.*, 2004; Li *et al.*, 2004; Sebat *et al.*, 2004; Perry *et al.*, 2006; Dopman & Hartl, 2007; Graubert *et al.*, 2007; Wong *et al.*, 2007; Emerson *et al.*, 2008). A potential consequence of segmental duplications is gene duplication.

For a duplicate gene to be evolutionarily preserved, it must go through three steps. At the time of origination, a new duplicate gene is carried by a single individual in a population in heterozygous condition (origination step). The majority of new duplications would be lost soon after their appearance in the population, unless they are strongly advantageous (Kondrashov *et al.*, 2002; Kondrashov & Koonin, 2004). Only a small fraction increases its frequency and subsequently becomes fixed in the population. This fixation step has largely been neglected in the preceding literature. By focusing on duplicate genes created by whole genome duplication, previous work has mainly studied their evolutionary trajectories, starting from a population where the duplicate genes are already fixed (e.g. Haldane, 1933; Nei &

* Corresponding author. Department of Population Genetics, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan. Tel: +81-55-981-6781. Fax: +81-55-981-6785. e-mail: totakano@lab.nig.ac.jp

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Roychoudhury, 1973; Bailey *et al.*, 1978; Kimura & King, 1979; Takahata & Maruyama, 1979; Li, 1980; Watterson, 1983; Force *et al.*, 1999; Walsh, 2003; Xue & Fu, 2009). However, for duplicate genes created by segmental duplication, most of their fates are determined in this fixation step. Subsequently after this period, the fates of the fixed duplicate genes will finally be resolved (resolution step), through non-functionalization, neofunctionalization, subfunctionalization, or other selection processes. Nonfunctionalization refers to the process whereby one member of the duplicate pair is completely silenced by degenerative mutation(s). The population then returns to the original single-gene state. Alternatively, both genes may be indefinitely preserved by gaining a beneficial novel function (neofunctionalization; Ohno, 1970), by partitioning multiple functions of the ancestral gene via complementary loss-of-function mutations (subfunctionalization; Force *et al.*, 1999) or by positive selection for increased amount of gene product (Kondrashov *et al.*, 2002). Because each of these processes may well proceed even in the prefixation period, the resolution step may sometimes be completed before the termination of the fixation step.

Here, we investigate another possibility, namely, enhanced preservation of the duplicate genes simply by the direct effect of gene duplication that masks deleterious loss-of-function mutations (Fisher, 1935). In two previous studies (Clark, 1994; Lynch *et al.*, 2001), this masking effect was only of minor importance in finite populations. Clark (1994) reported that the masking effect of gene duplication does not significantly affect the equilibrium frequency and that a duplicate gene actually behaves like a neutral mutation in his simulations. Lynch *et al.* (2001) found only a two-fold increase in the fixation probability of a completely linked duplicate gene in large populations, although either one of the two duplicate genes is silenced in the early phase of evolution. However, there are several factors that were not fully explored yet, such as the dominance of deleterious mutations (haplo-insufficiency), the strength of mutation pressure and recombination between duplicate genes. Recently, it was found that, for genes with dominant lethal effect when their copy number is halved in diploid organisms (i.e. haplo-insufficient genes), the masking effect can retard the non-functionalization of a fixed duplication (Xue & Fu, 2009; see also Takahata & Maruyama, 1979). This finding further raises a question of whether the masking effect also increases the fixation probability of a newly arisen duplicate gene.

Both Clark (1994) and Lynch *et al.* (2001) used the double-null recessive model, whereby all two-locus genotypes have an equal fitness, except for double-null honozygotes that completely lack a gene function and therefore are lethal (haplo-sufficiency). Actually, in addition to a small number of haplo-insufficient genes, most, if not all, of loss-of-function mutations are slightly deleterious in heterozygous condition. The average degree of dominance of lethal mutations is indeed estimated to be about 0·02 in *Drosophila melanogaster* (Simmons & Crow, 1977).

In this paper, we explored whether such a small degree of dominance, together with high mutation pressure and recombination, is sufficient to become of evolutionary significance. It is demonstrated that the fixation of a newly arisen duplicate gene is substantially facilitated by loss-of-function mutations for a wide range of parameter values. The effect of masking deleterious mutations serves as an alternative mechanism to preserve both duplicate genes for long periods, which could increase the chance for neofunctionalization. To this end, we consider two models, single- and two-function models, with loss-of-function mutations.

2. Single-function model

Table 1 summarizes the abbreviations and parameters used to describe the fixation and resolution steps. Throughout we assume that right after the duplication event, both duplicate genes maintain the original function of the ancestral gene, with no intrinsic advantage or disadvantage to duplications.

Initially, we consider a single locus under mutation–selection–drift balance in a panmictic population of N diploids. A new duplicate gene is then created, with only a single chromosome carrying a duplicate gene (at an initial frequency of $1/(2N)$). Unless otherwise stated, the gene is 'essential'. Namely, the relative fitness (viability) of individuals harbouring no functional allele is 0 (i.e. selection coefficient $s=1$). Individuals carrying a single functional allele have a relative fitness of $1-h$, and all other individuals carrying two or more functional alleles have a relative fitness of unity (Fig. 1*a*). We allowed three different degrees of dominance: $h=0$ (for double-null recessive genes), $h=0{\cdot}02$ (for partially recessive genes) and $h=1$ (for haplo-insufficient genes). We also studied non-essential genes assuming $s=0{\cdot}1$, with five different degrees of dominance ($h=0$, 0·002, 0·02, 0·2, or 1).

Duplicate genes are either completely linked to each other ($c=0$, where c denotes recombination rate per generation between the two loci) or freely recombining ($c=0{\cdot}5$). We also consider another case of $c=10^{-4}$, which represents the average recombination rate between adjacent genes of *D. melanogaster* (Lindsley & Zimm, 1992).

In the single-function model, loss-of-function mutations that completely disrupt the function occur at a rate of $u_c=10^{-3}$ per locus per generation. It is

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Table 1. *Parameters used for describing the fixation and resolution steps*

N	Effective (and actual) population size.
u_c	Rate of loss-of-function mutations that completely disrupt a gene.
u_r	Rate of regulatory mutations that eliminate a subfunction in the two-function model ($2u_r$ per gene).
s	Selection coefficient of mutations.
h	Degree of dominance of mutations.
c	Recombination rate between duplicate genes.
Scaled probability of fixation	Probability that a newly arisen duplicate gene is fixed in a population, regardless of whether it is functional or nonfunctional, divided by $1/(2N)$.
Scaled probability of preservation of a new duplicate gene	Probability that a newly arisen duplicate gene is permanently preserved due to non-functionalization of the original gene or both duplicate genes are functionally preserved for $100N$ generations, divided by $1/(2N)$. This corresponds to Θ in Lynch *et al.* (2001).
Probability of functional fixation	Probability that a newly arisen duplicate gene is fixed while keeping both duplicate genes functional.
Scaled time of fixation	Time until fixation of a newly arisen duplicate gene, divided by $4N$.

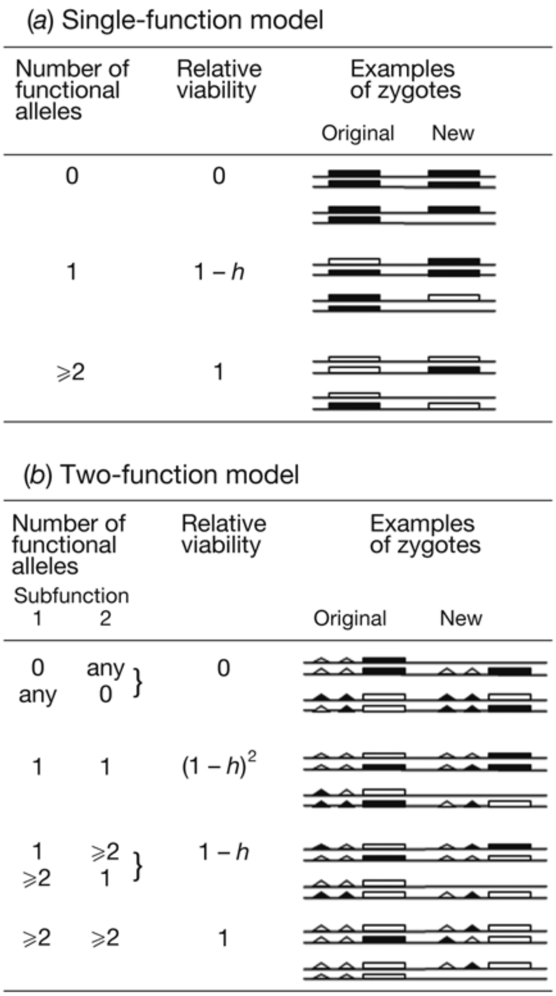

Fig. 1. Fitness scheme in (*a*) the single-function and (*b*) two-function models. A square and a triangle denote a protein coding region and a *cis*-regulatory region, respectively. Functionally intact regions are indicated in white, while degenerated regions with loss-of-function mutations are indicated in black.

assumed that the mutations are unidirectional and that backward mutations do not occur. We further ignore advantageous mutations that lead to neo-functionalization.

To study the evolutionary fates after gene duplication (fixation and resolution steps), we performed stochastic simulations based on a gamete-based model (Lynch & Force, 2000) over a range of population size ($N=50$–10^5). Given the frequencies of gametes in the previous generation, we first calculated the expected frequencies of zygotes after mutations, random mating and viability selection. Based on these expectations, the actual zygote frequencies are obtained by sampling N individuals using the improved pseudo-sampling method (Kimura & Takahata, 1983). Finally, the expected frequencies of gametes after recombination are determined for the next generation. In this gamete-based model, mutation, selection and recombination are all treated as deterministic processes (Lynch & Force, 2000).

To investigate the fixation probability and fixation time of a newly arisen duplicate gene in the fixation step, the above cycle is repeated until the duplicate gene reaches fixation or is lost from the population, irrespective of the functional state of the duplicate gene. At least 100 fixation events were simulated for each set of parameter values. To investigate the evolutionary fates in the resolution step, the simulation cycle is further continued until one member of a duplicate pair becomes silenced (non-functionalization). If functional alleles are preserved for $100N$ generations at both loci, the simulation run is halted and the next run is initiated.

(i) *Fixation step in the single-function model*

For a newly arisen duplicate of an essential gene (with $s=1$), the fixation probability is given in Table 2, together with the mean time to fixation. In the table, the results are scaled in units of neutral expectations ($1/(2N)$ for the fixation probability, or $4N$ generations for the fixation time). When $Nu_c \leqslant 0{\cdot}1$, both fixation probability and time are not much different from their neutral expectations, irrespective of s, h and c values. However, when $Nu_c \geqslant 0{\cdot}5$, the fixation probability is substantially increased and, concomitantly, the

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Table 2. *Scaled probability and time of fixation of a newly arisen duplicate gene in the single-function model with* $u_c = 10^{-3}$ *and* $s = 1$

	$c=0$			$c=10^{-4}$			$c=0{\cdot}5$		
N	$h=0$	$h=0{\cdot}02$	$h=1$	$h=0$	$h=0{\cdot}02$	$h=1$	$h=0$	$h=0{\cdot}02$	$h=1$
Fixation probability									
50	1·1	1·0	1·3	1·0	1·1	1·3	1·1	1·1	1·4
100	1·1	1·3	1·3	1·2	1·1	1·3	1·2	1·4	1·7
500	1·4	1·7	2·3	1·4	1·6	2·3	2·0	2·5	3·2
1000	1·6	2·1	2·6	1·6	2·0	2·9	2·7	3·8	5·7
5000	2·0	2·8	5·5	2·6	4·2	8·7	4·5	11·5	20·9
10 000	2·1	3·3	6·9	2·9	4·8	15·9	6·9	18·9	44·8
50 000	2·0	3·2	16·1	5·5	18·0	78·8	15·8	96·0	226·2
100 000	1·9	4·0	22·3	6·6	39·1	150·0	18·6	178·5	339·4
Median time to fixation									
50	0·75 (0·37–1·91)	0·81 (0·40–2·00)	0·80 (0·37–1·99)	0·81 (0·38–1·90)	0·87 (0·40–1·92)	0·81 (0·35–1·84)	0·76 (0·37–1·99)	0·86 (0·36–2·21)	0·80 (0·32–1·69)
100	0·90 (0·34–2·36)	0·80 (0·36–2·24)	0·89 (0·38–1·77)	0·82 (0·37–1·79)	0·79 (0·36–1·87)	0·82 (0·39–1·84)	0·86 (0·38–2·17)	0·80 (0·31–1·74)	0·87 (0·41–2·05)
500	0·82 (0·34–1·99)	0·85 (0·35–1·78)	0·73 (0·32–1·89)	0·84 (0·39–1·76)	0·81 (0·38–1·96)	0·96 (0·47–2·09)	0·65 (0·34–1·65)	0·69 (0·33–1·78)	0·59 (0·29–1·23)
1000	0·78 (0·37–1·85)	0·74 (0·33–1·76)	0·64 (0·36–1·66)	0·78 (0·30–1·72)	0·83 (0·36–1·93)	0·88 (0·40–1·90)	0·66 (0·31–1·85)	0·60 (0·28–1·30)	0·46 (0·25–0·90)
5000	0·75 (0·40–1·82)	0·71 (0·30–2·08)	0·65 (0·27–1·72)	0·73 (0·36–1·73)	0·69 (0·37–1·76)	0·60 (0·30–1·32)	0·43 (0·20–1·44)	0·31 (0·18–1·12)	0·17 (0·10–0·29)
10 000	0·77 (0·33–1·74)	0·80 (0·31–1·73)	0·60 (0·19–2·31)	0·69 (0·38–1·67)	0·72 (0·33–1·61)	0·48 (0·20–1·20)	0·43 (0·17–1·42)	0·21 (0·12–0·70)	0·11 (0·07–0·14)
50 000	0·80 (0·39–1·65)	0·86 (0·32–1·83)	0·62 (0·13–1·64)	0·64 (0·29–1·73)	0·34 (0·18–1·16)	0·15 (0·09–0·23)	0·41 (0·10–1·81)	0·06 (0·04–0·09)	0·03 (0·02–0·04)
100 000	0·82 (0·34–1·88)	0·80 (0·37–1·86)	0·54 (0·12–1·78)	0·57 (0·24–1·81)	0·23 (0·15–0·97)	0·09 (0·06–0·13)	0·47 (0·07–1·52)	0·04 (0·02–0·05)	0·02 (0·01–0·02)

90 % interval of fixation time is represented in the parentheses.

fixation time is decreased. This result implies that under sufficiently high mutation pressure, a duplicate gene becomes selectively advantageous by masking the deleterious effect of recurrent loss-of-function mutations. While this masking effect was more evident with larger h values, even an h value as small as 0·02 had a significant impact. Recombination also has an important effect on the evolution of duplicate genes. When the new duplicate gene is completely linked to the original gene ($c = 0$), its advantage was not particularly noticeable, except for haplo-insufficient genes ($h = 1$) under high mutation pressure ($Nu_c > 1$). By contrast, for an unlinked copy ($c = 0{\cdot}5$), substantial increase in the fixation probability and decrease in fixation time were observed when $Nu_c \geqslant 0{\cdot}5$, irrespective of the degree of dominance. Although less intense in its magnitude, the same tendency was detected even for the recombination rate as small as $c = 10^{-4}$.

We analysed the joint effects of h and s more in depth under $c = 0{\cdot}5$, and obtained the following two findings as summarized in Table 3. First, the fixation probability and time did not much differ between $s = 1$ and 0·1 except for the case of $h = 0{\cdot}02$; by contrast, the degree of dominance (h) of mutations had stronger effects. Second, when hs was kept constant, the selective advantage of a duplicate gene was more evident for larger h (and smaller s); for instance, for $hs = 0{\cdot}02$, the deviation from neutrality was more substantial when $(h, s) = (0{\cdot}2, 0{\cdot}1)$ than when $(h, s) = (0{\cdot}02, 1)$. Likewise, for $hs = 0{\cdot}002$, the effect of selection was more obvious when $(h, s) = (0{\cdot}02, 0{\cdot}1)$ than when $(h, s) = (0{\cdot}002, 1)$.

(ii) *Resolution step in the single-function model*

For the resolution step, we focus on essential genes with $s = 1$. In the single-function model, non-functionalization is usually inevitable and one of the duplicate genes will be silenced sooner or later. Indeed, our simulations demonstrated that either when $h = 0$ or $c = 0$, non-functionalization was always completed within $100N$ generations unless $Nu_c = 0{\cdot}05$ (Figs 2*a*–*d*, *g* and 3*a*). Non-functionalization occurred with an approximately equal frequency at either of the two loci when $c = 0$ (Fig. 2*a*–*c*), while it happened mostly at the new locus when $h = 0$ and $c = 10^{-4}$ or 0·5 (Fig. 2*d* and *g*). By contrast, when $h > 0$ and $c = 0{\cdot}5$, functional alleles were largely preserved at both loci even after $100N$ generations if the mutation pressure is sufficiently high (Fig. 2*h* and *i*). Although larger Nu_c values are required, the same tendency was seen with $c = 10^{-4}$ (Fig. 2*e* and *f*).

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Table 3. *Scaled probability and time of fixation of a newly arisen duplicate gene in the single-function model when $N = 50\,000$, $u_c = 10^{-3}$ and $c = 0{\cdot}5$*

	h				
s	0	0·002	0·02	0·2	1
Fixation probability					
1	15·8	20·4	96·0	195·9	226·2
0·1	16·0	17·9	43·4	170·1	209·7
Median time to fixation					
1	0·413 (0·105–1·814)	0·211 (0·092–1·431)	0·063 (0·044–0·091)	0·032 (0·024–0·046)	0·030 (0·022–0·044)
0·1	0·359 (0·118–1·682)	0·306 (0·099–1·783)	0·115 (0·071–0·538)	0·043 (0·034–0·057)	0·033 (0·026–0·048)

90 % interval of fixation time is represented in the parentheses.

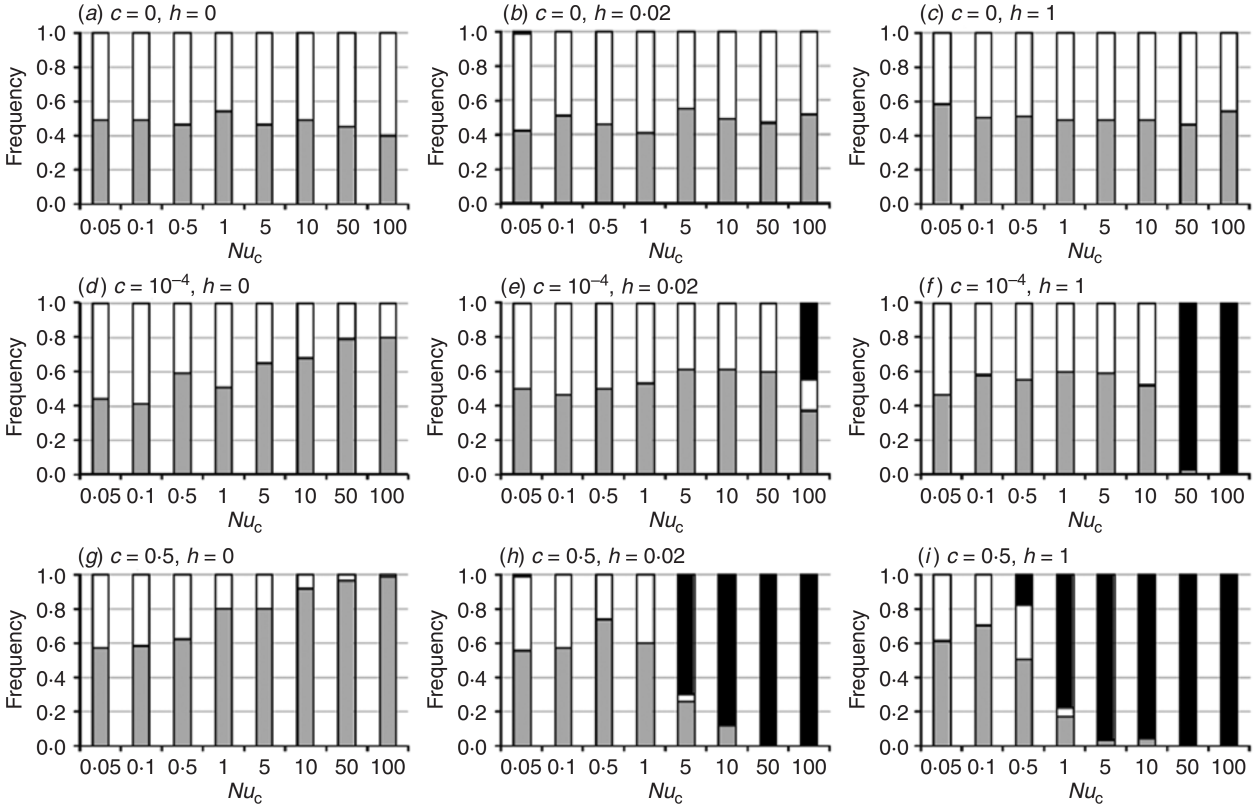

Fig. 2. Evolutionary fates of duplicate genes after $100N$ generations in the single-function model. Results for three different recombination rates ($c = 0$, 10^{-4}, or 0·5) and three different degrees of dominance ($h = 0$, 0·02, or 1) are illustrated; $u_c = 10^{-3}$ and $s = 1$ are assumed throughout. For each combination of parameter values, simulations were performed with nine different population sizes ($N = 50$–10^5). The figure shows the relative frequencies of three possible outcomes: non-functionalization at the new locus (grey), non-functionalization at the original locus (white) and preservation of both loci (black).

Time course of non-functionalization is illustrated in Fig. 3 for the case of $c = 0{\cdot}5$ and $Nu_c = 5$. When $h = 0$, most non-functionalization events occurred within $10N$ generations, particularly at the new locus (Fig. 3*a*). When $h = 0{\cdot}02$, non-funcitonalization occurred gradually after $10N$ generations, but both genes were still functional in more than 60 % of simulation runs even when $100N$ generations have elapsed since the appearance of a new duplicate gene (Fig. 3*b*). When $h = 1$, non-functionalization is almost completely prevented from occurring (Fig. 3*c*). Indeed, the frequencies of functional alleles after $100N$

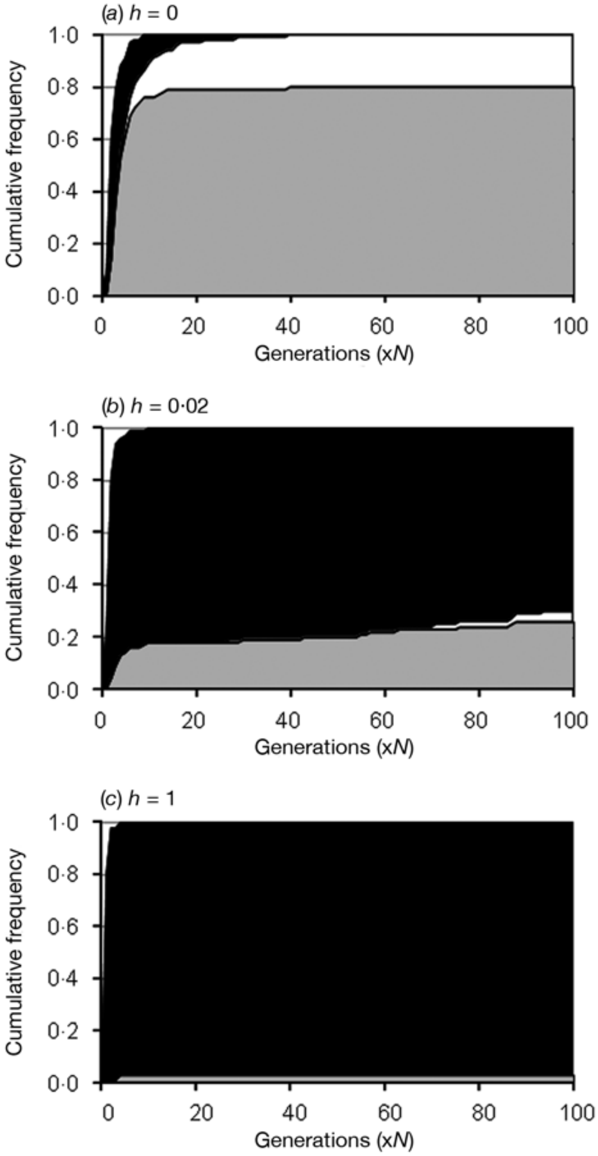

Fig. 3. Temporal increase in non-functionalization in the single-function model. Parameter values are $N=5000$, $u_c=10^{-3}$, $c=0{\cdot}5$ and $s=1$, with (*a*) $h=0$, (*b*) $h=0{\cdot}02$, or (*c*) $h=1$. The figure shows the cumulative frequencies of three possible outcomes (conditional on the ultimate fixation of the new duplicate gene): non-functionalization at the new locus (grey), non-functionalization at the original locus (white) and preservation of both loci (black).

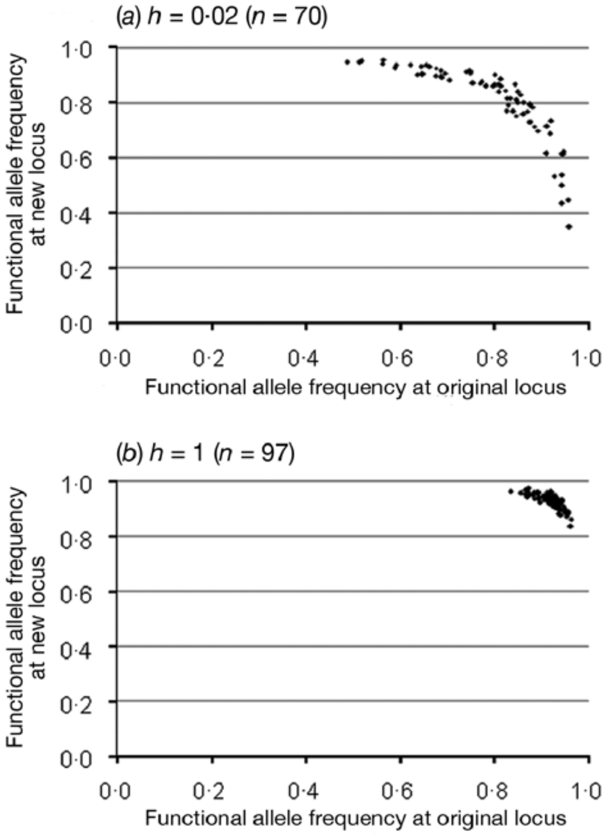

Fig. 4. Frequencies of functionally intact alleles at the two loci after $100N$ generations in the single-function model. Parameter values are $N=5000$, $u_c=10^{-3}$, $c=0{\cdot}5$ and $s=1$, with (*a*) $h=0{\cdot}02$ or (*b*) $h=1$. The value of n in the parenthesis indicates the observed number of simulation runs in which both loci remained polymorphic for $100N$ generations.

generations were kept higher than 0·8 at both loci when $h=1$ (Fig. 4*b*), while asymmetry in allele frequency between the two loci was stronger for $h=0{\cdot}02$ (Fig. 4*a*).

So far, we have analysed the fixation and resolution steps separately. To contrast the present observations with the results of Lynch *et al.* (2001), we here consider the probability of preservation of a newly arisen duplicate gene after $100N$ generations. There are two possibilities: non-functionalization of the original gene (and therefore permanent preservation of the newly arisen duplicate gene) and preservation of both original and new duplicate genes. Table 4 gives the combined probability scaled in units of the neutral expectation ($=1/(2N)$). Lynch *et al.* (2001) found that irrespective of the recombination rate ($c=0$ or $0{\cdot}5$), the scaled combined probability was $\sim 0{\cdot}5$ in small populations ($Nu_c<0{\cdot}1$), assuming haplo-sufficient essential genes ($h=0$ and $s=1$). In large populations, while the scaled probability was kept almost constant ($\sim 0{\cdot}5$) for freely recombining loci, it increased up to unity under complete linkage (see Fig. 3 in Lynch *et al.*, 2001). In the present analysis, we found much higher increase in the combined probability for $h>0$, especially when there was a nonzero opportunity of recombination between the two loci (Table 4). Importantly, this increase is largely contributed by long preservation of both duplicate genes, which cannot be seen under conditions studied in Lynch *et al.* (2001). These results also highlight the importance of the degree of dominance in promoting the preservation of a new duplicate gene under high mutation pressure.

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Table 4. *Scaled probability of preservation of a new duplicate gene under* $u_c = 10^{-3}$ *and* $s = 1$

	$c=0$			$c=10^{-4}$			$c=0{\cdot}5$		
N	$h=0$	$h=0{\cdot}02$	$h=1$	$h=0$	$h=0{\cdot}02$	$h=1$	$h=0$	$h=0{\cdot}02$	$h=1$
Single-function model									
50	0·5	0·6	0·5	0·7	0·5	0·7	0·4	0·5	0·4
	(0·4–0·6)	(0·3–0·6)	(0·3–0·6)	(0·5–0·8)	(0·4–0·6)	(0·4–0·7)	(0·3–0·5)	(0·4–0·7)	(0·3–0·5)
100	0·6	0·5	0·6	0·6	0·7	0·6	0·4	0·6	0·5
	(0·4–0·7)	(0·3–0·5)	(0·4–0·8)	(0·4–0·6)	(0·4–0·7)	(0·4–0·7)	(0·3–0·5)	(0·4–0·7)	(0·3–0·6)
500	0·7	0·9	1·0	0·5	0·8	1·0	0·7	0·6	1·5
	(0·5–0·8)	(0·6–1·0)	(0·7–1·2)	(0·4–0·6)	(0·6–1·0)	(0·8–1·4)	(0·5–0·9)	(0·3–0·7)	(1·1–1·9)
1000	0·8	1·3	1·3	0·9	1·0	1·0	0·5	1·5	4·6
	(0·5–1·0)	(0·8–1·4)	(1·0–1·6)	(0·7–1·1)	(0·7–1·2)	(0·7–1·2)	(0·3–0·7)	(1·1–1·9)	(3·6–5·3)
5000	1·1	1·4	3·1	0·8	1·6	4·0	1·0	9·8	20·5
	(0·7–1·2)	(1·0–1·7)	(2·3–3·7)	(0·6–1·0)	(1·2–2·1)	(3·0–5·1)	(0·6–1·4)	(7·5–11·3)	(15·8–22·6)
10 000	1·1	1·9	3·7	0·9	2·3	6·7	0·5	19·2	40·7
	(0·8–1·3)	(1·4–2·3)	(2·8–4·5)	(0·6–1·2)	(1·7–2·9)	(5·2–8·5)	(0·4–0·6)	(14·4–21·0)	(31·8–45·4)
50 000	1·2	1·9	8·2	1·1	7·5	74·9	0·6	86·2	235·8
	(0·8–1·3)	(1·4–2·3)	(5·9–9·5)	(0·7–1·5)	(5·6–9·7)	(57·8–82·7)	(0·4–0·7)	(65·2–93·3)	(198·4–278·5)
100 000	1·1	2·0	10·3	1·4	21·5	174·3	0·3	169·6	455·7
	(0·8–1·3)	(1·5–2·5)	(7·9–13·1)	(0·9–2·0)	(16·3–25·3)	(146·6–205·7)	(0·2–0·4)	(142·6–200·2)	(383·4–538·2)
Two-function model ($u_r = u_c = 10^{-3}$)									
50	0·7	0·7	0·8	0·6	0·7	0·9	0·7	0·7	0·8
	(0·5–0·8)	(0·5–0·7)	(0·6–1·0)	(0·4–0·8)	(0·5–0·8)	(0·6–1·1)	(0·5–0·8)	(0·4–0·7)	(0·6–1·0)
100	0·8	0·9	0·9	0·6	1·0	1·1	0·7	1·0	1·0
	(0·6–0·9)	(0·6–0·9)	(0·6–1·1)	(0·4–0·7)	(0·7–1·1)	(0·8–1·4)	(0·5–0·9)	(0·7–1·1)	(0·7–1·2)
500	1·1	1·3	2·3	0·8	1·3	2·3	0·6	1·2	5·5
	(0·7–1·2)	(0·9–1·4)	(1·5–2·7)	(0·6–1·0)	(0·9–1·5)	(1·6–2·9)	(0·4–0·7)	(0·8–1·6)	(3·7–5·9)
1000	1·0	1·6	2·1	1·2	1·8	2·7	0·6	1·6	12·4
	(0·7–1·1)	(1·2–1·9)	(1·5–2·9)	(0·8–1·3)	(1·2–2·0)	(2·0–3·7)	(0·4–0·7)	(1·0–2·3)	(9·7–14·2)
5000	1·1	2·6	6·8	1·0	1·9	8·9	0·7	24·5	59·6
	(0·7–1·2)	(1·7–2·8)	(4·6–8·2)	(0·7–1·3)	(1·4–2·5)	(6·5–11·9)	(0·4–0·8)	(20·1–29·5)	(47·5–68·0)
10 000	1·1	2·4	8·9	1·1	3·2	19·3	0·7	49·2	130·0
	(0·7–1·2)	(1·8–3·0)	(5·8–10·3)	(0·8–1·5)	(2·3–4·3)	(13·9–26·2)	(0·5–0·9)	(39·9–58·4)	(103·6–148·2)
50 000	0·8	2·0	20·7	1·0	12·7	216·6	0·5	252·7	515·9
	(0·5–0·9)	(1·5–2·5)	(15·3–27·4)	(0·6–1·3)	(9·3–16·9)	(176·2–263·4)	(0·3–0·7)	(187·0–273·6)	(394·7–564·3)
100 000	1·0	2·6	30·0	1·0	57·5	538·1	0·5	486·7	1025·2
	(0·7–1·2)	(2·0–3·3)	(22·2–39·7)	(0·7–1·3)	(46·2–67·5)	(437·8–654·3)	(0·3–0·7)	(392·3–560·9)	(784·3–1121·3)

95 % confidence limit of the probability based on Poisson statistics (Gehrels, 1986) is represented in the parentheses.

3. Two-function model

We here extend the single-function model developed in the preceding section and consider an ancestral gene that has two independently mutable subfunctions. In this two-function model, each of the two duplicate genes is subject to two distinct classes of degenerative mutations: regulatory mutations that eliminate only one of the two subfunctions and coding mutations that disrupt the entire gene functions simultaneously. The former class occurs at a rate $u_r = 10^{-3}$ per locus per generation for each subfunction, and the latter also occurs at a rate $u_c = 10^{-3}$. We here focus on essential genes; the relative fitness values are set to be 0 for individuals carrying no functional allele for either subfunctions, $(1-h)^2$ for those carrying only a single functional allele for each subfunction, $1-h$ for those carrying one functional allele for one of the subfunctions together with two or more functional alleles for the other subfunction, and 1 for those carrying two or more functional alleles for both subfunctions (multiplicative fitness model, Fig. 1*b*).

As in the single-function model, an initial population was assumed to be in mutation–selection–drift equilibrium. Each run of simulations was started by introducing a single copy of haplotype with two fully functional alleles at both loci.

Because subfunctionalization may occur in the two-function model (Force *et al.*, 1999), we repeated at lease 100 simulation runs, each leading to either non-functionalization, subfunctionalization, or preservation of functional alleles at both loci (after 100N generations).

(i) *Fixation step in the two-function model*

Because u_c and u_r are all set to be 10^{-3}, the total mutation rate is three times as large as in the single-function model. This entails even higher pressure for degenerative mutations, further leading to greater probabilities for the fixation of a new duplicate gene (Table 5). Otherwise, the results were essentially the same as in the single-function model.

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Table 5. *Scaled probability and time of fixation of a newly arisen duplicate gene in the two-function model with* $u_c = u_r = 10^{-3}$ *and* $s = 1$

	$c=0$			$c=10^{-4}$			$c=0{\cdot}5$		
N	$h=0$	$h=0{\cdot}02$	$h=1$	$h=0$	$h=0{\cdot}02$	$h=1$	$h=0$	$h=0{\cdot}02$	$h=1$
Fixation probability									
50	1·4	1·3	1·3	1·2	1·3	1·3	1·5	1·2	2·0
100	1·4	1·5	2·0	1·5	1·7	1·9	1·6	1·7	2·6
500	1·9	2·6	3·9	1·8	2·3	3·6	3·8	4·7	7·5
1000	1·9	2·5	5·3	1·9	3·5	5·1	4·7	8·2	15·8
5000	2·0	4·7	9·5	3·0	6·2	22·7	11·3	34·4	66·2
10 000	2·4	4·3	15·6	3·0	8·7	38·6	16·8	66·0	106·3
50 000	1·9	4·8	37·1	5·7	31·7	249·2	35·6	271·4	592·0
100 000	2·2	5·4	55·8	7·1	67·9	434·0	61·8	528·0	1240·0
Median time to fixation									
50	0·80 (0·40–1·84)	0·83 (0·41–1·89)	0·89 (0·40–2·14)	0·82 (0·39–1·74)	0·76 (0·38–1·82)	0·81 (0·35–1·59)	0·80 (0·35–1·90)	0·77 (0·34–1·45)	0·69 (0·28–1·57)
100	0·71 (0·37–1·81)	0·75 (0·33–1·83)	0·81 (0·38–1·94)	0·72 (0·40–1·55)	0·76 (0·32–1·64)	0·80 (0·39–1·92)	0·75 (0·34–1·66)	0·76 (0·40–1·78)	0·64 (0·34–1·31)
500	0·72 (0·34–1·64)	0·65 (0·29–1·79)	0·66 (0·32–1·83)	0·77 (0·32–2·00)	0·73 (0·34–1·95)	0·67 (0·33–1·77)	0·54 (0·27–1·37)	0·53 (0·25–1·11)	0·36 (0·21–0·78)
1000	0·74 (0·33–1·58)	0·71 (0·37–1·97)	0·57 (0·26–1·51)	0·72 (0·32–1·71)	0·78 (0·28–1·73)	0·54 (0·26–1·35)	0·52 (0·21–1·60)	0·43 (0·22–1·46)	0·25 (0·15–0·63)
5000	0·72 (0·29–1·61)	0·74 (0·33–1·93)	0·47 (0·17–1·61)	0·65 (0·29–1·58)	0·64 (0·26–1·43)	0·42 (0·16–1·12)	0·38 (0·13–1·49)	0·17 (0·09–0·79)	0·08 (0·06–0·27)
10 000	0·76 (0·32–1·81)	0·79 (0·37–2·22)	0·53 (0·15–1·49)	0·66 (0·35–1·23)	0·60 (0·27–1·48)	0·28 (0·13–0·94)	0·37 (0·10–1·56)	0·11 (0·06–0·82)	0·05 (0·03–0·22)
50 000	0·79 (0·38–1·99)	0·77 (0·30–1·95)	0·45 (0·08–1·56)	0·59 (0·23–1·79)	0·34 (0·17–1·74)	0·08 (0·05–0·12)	0·41 (0·05–1·81)	0·03 (0·02–0·38)	0·01 (0·01–0·02)
100 000	0·79 (0·43–2·00)	0·82 (0·35–1·83)	0·55 (0·06–1·71)	0·57 (0·17–1·82)	0·18 (0·11–0·32)	0·04 (0·03–0·06)	0·39 (0·03–1·45)	0·02 (0·01–0·36)	0·01 (0·01–0·01)

90 % interval of fixation time is represented in the parentheses.

(ii) *Resolution step in the two-function model*

As in the previous study (Lynch & Force, 2000; Lynch *et al*., 2001), subfunctionalization was observed only for small Nu_c values, say $Nu_c \leqslant 0{\cdot}5$, irrespective of h values (Fig. 5). The probability of subfunctionalization was not much affected by recombination rate. As in the single-function model, joint preservation of fully functional alleles at both loci was facilitated under high mutation pressure so long as $h \geqslant 0{\cdot}02$ and $c \geqslant 10^{-4}$. The transition from non-functionalization to preservation occurred in a narrow range of Nu_c values.

4. Probability of fixation of functional duplications

As shown above, when $c = 0{\cdot}5$, $h > 0$ and $Nu_c > 1$, both members of a duplicate pair can functionally be preserved during and after the fixation of the newly arisen duplicate gene (Fig. 2*h* and *i*). Here, we refer to the probability that a new duplicate gene is fixed while keeping both genes functional as the 'functional fixation' probability. To obtain the probability of functional fixation for arbitrary s and h values, consider selection acting on a rare duplication in the single-function model. Assume that a duplication occurs in a sufficiently large population at mutation–selection equilibrium. Let q_e denote the equilibrium frequency of the non-functional allele at the original locus. For large populations ($Nu_c > 1$), q_e is given by

$$q_e = \frac{\sqrt{h^2 s^2 (1+u_c)^2 + 4u_c s(1-2h)} - hs(1+u_c)}{2s(1-2h)} \quad \text{for } h \neq 1/2,$$

and

$$q_e = \frac{2u_c}{s(1+u_c)} \quad \text{for } h = 1/2 \tag{1}$$

(Crow & Kimura, 1970).

The expected change in the frequency (x) of a rare, unlinked duplicate gene per generation is given by

$$\begin{aligned}\Delta x &= \frac{x}{\bar{w}}[(1-u_c)\{1-q^2(1-x)hs\} - \bar{w}] \\ &= \frac{x(1-x)}{\bar{w}}\left[(1-x)\{2(1-q)qhs + q^2(1-h)s\}\right. \\ &\quad \left. + xq^2hs - u_c\left\{\frac{1}{1-x} - q^2hs\right\}\right],\end{aligned} \tag{2}$$

where $\bar{w}$ is the mean fitness of the population and q is the frequency of the non-functional allele at the

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

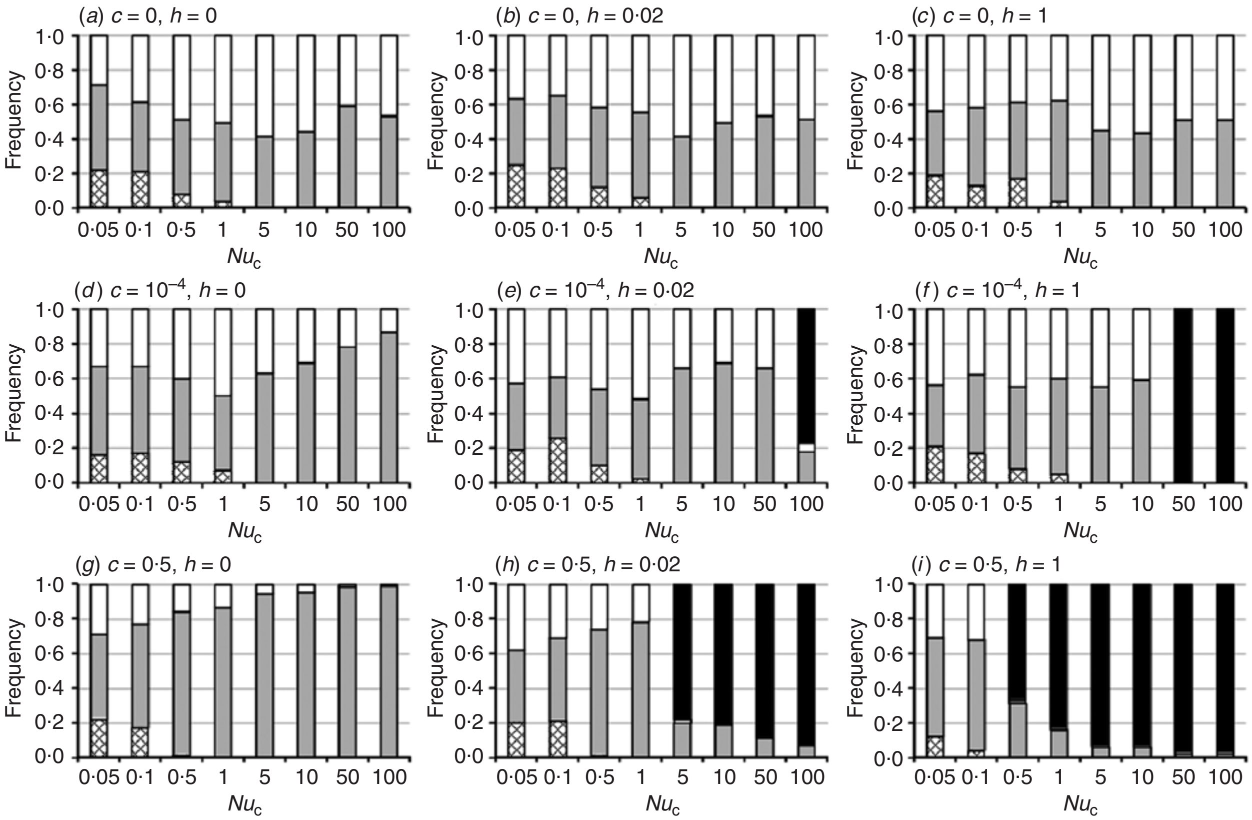

Fig. 5. Evolutionary fates of duplicate genes after $100N$ generation in the two-function model. Results for three different recombination rates ($c = 0$, 10^{-4}, or 0·5) and three different degrees of dominance ($h = 0$, 0·02, or 1) are illustrated; $u_c = u_r = 10^{-3}$ and $s = 1$ are assumed throughout. For each combination of parameter values, simulations were performed with nine different population sizes ($N = 50$–10^5). The figure shows the relative frequencies of four possible outcomes: subfunctionalization (cross-hatched), non-functionalization at the new locus (grey), non-functionalization at the original locus (white) and preservation of both loci (black).

original locus. Unlike in the standard derivation, the effect of recurrent mutation cannot be neglected here, because the selective advantage of the new duplicate gene is of the same order of magnitude as the mutation rate. When x is small, we may replace q by q_e. Then, we see an increase of x ($\Delta x > 0$) when $0 < h \leqslant 1$, implying the selective advantage of the new duplicate gene at low frequencies. This advantage in large populations can account for the enhanced fixation and preservation of duplicate genes under large Nu_c and $c = 0{\cdot}5$ (Table 2, Fig. 2*h* and *i*).

The fixation probability of a mutation can be approximated by twice the selective advantage of the heterozygote (Kimura, 1957; Gale, 1990). This may hold true for a duplicate gene. In the present case, this selective advantage may be obtained by taking the limit $x \to 0$ in the right-hand side of eqn (2) and then replacing q by q_e. This yields approximately the functional fixation probability (P_{ff}) as

$$P_{\mathrm{ff}} = 2\{2(1-q_e)q_e hs + q_e^2(1-h)s - u_c\}. \qquad (3)$$

As shown graphically in Fig. 6, the predicted probability (3) increases from 0 to $2u_c$ rapidly as h increases. The prediction is in close agreement with the simulated probabilities of functional fixation (Table 6). When $h = 0$, the selective advantage of a new duplicate gene becomes ~ 0. Therefore, loss-of-function mutations accumulate on *neutral* duplicate genes immediately after the origination, leading to non-functionalization predominantly at the new locus.

For small populations, q_e becomes smaller than the equilibrium frequency given by the formulae (1) due to the purging effect (Kirkpatrick & Jarne, 2000; Glémin, 2003). This reduction in the frequency of non-functional alleles decreases the selective advantage of a new duplicate gene, which, in turn, reduces the probability of functional fixation.

When the two loci are completely linked ($c = 0$), functional fixation as defined above may be considered as fixation of the functional two-copy allele (designated *ff*, where *f* refers to a functional allele at a single locus). The expected change per generation in the frequency (y) of the *ff* allele is given by

$$\Delta y = \frac{y}{\bar{w}}(1 - 2u_c - \bar{w}). \qquad (4)$$

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

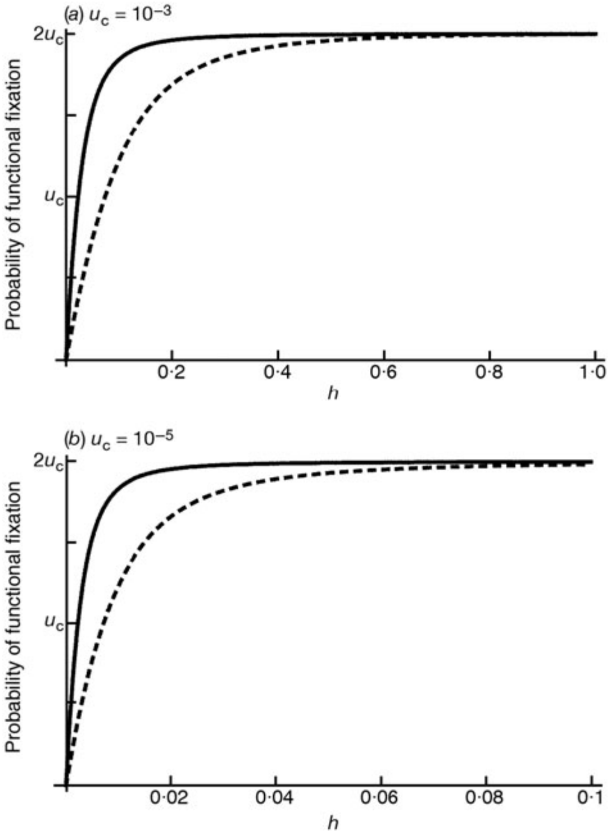

Fig. 6. The predicted probability of functional fixation as a function of h. (*a*) $u_c = 10^{-3}$ and (*b*) $u_c = 10^{-5}$, with $s = 1$ (solid line) or $s = 0{\cdot}1$ (dotted line).

This equation has been obtained by Lynch *et al.* (2001) for the double-null recessive model ($h = 0$). In eqn (4), the mutation rate u_c is multiplied by two because mutation at either of the two loci destroys an *ff* allele. Therefore, the effect of mutation that hinders functional fixation is twice as strong as in the free recombination case ($c = 0{\cdot}5$; see eqn 2). Consequently, in the absence of recombination, the selective advantage of a duplication cannot be significant enough to overcome the counteracting effect of loss-of-function mutation. Indeed, our simulations have found a complete lack of functional preservation of both loci when $c = 0$ (Fig. 2*a–c*). Because an *ff* allele is converted to *f*0 or 0*f* allele (0 refers to a non-functional allele) with equal probability, non-functionalization occurs equally at the original and new loci for fixed duplications.

We have also seen in the above simulations that reduced recombination decreases the selective advantage of a new duplicate gene (compare Fig. 2*e* and *f* with *h* and *i*, respectively). Roughly speaking, recombination greater than the mutation rate ($c > u_c$) is needed for ample opportunities of functional fixation (Table 6).

In conclusion, with a sufficient amount of recombination ($c > u_c$), the probability of functional fixation approaches $2u_c$ in large populations ($Nu_c > 1$) as h increases, and the transition of the probability from 0 to $2u_c$ occurs in a narrow range of h, especially when u_c is small.

5. Discussion

Most theoretical studies of gene duplication have been concerned with the evolutionary consequences of ancient whole-genome duplications, focusing mainly on the resolution process leading to non-functionalization, starting from a population where the duplicate genes are already fixed (e.g. Haldane, 1933; Nei & Roychoudhury, 1973; Bailey *et al.*, 1978; Kimura & King, 1979; Takahata & Maruyama, 1979; Li, 1980; Watterson, 1983; Force *et al.*, 1999; Walsh, 2003; Xue & Fu, 2009). On the other hand, for duplicate genes created by segmental duplication, the fixation process is more important because the majority of duplications would be lost or silenced before reaching a significant frequency in a population (Kondrashov *et al.*, 2002).

Here, we showed that purifying selection against loss-of-function mutations increases the fixation probability of a new duplicate gene (Tables 2, 3 and 5) and enhances the preservation of functional alleles at both duplicate loci (Figs 2, 3 and 5, Table 6). In large populations ($Nu_c > 1$), the probability that a new duplicate gene is fixed while preserving both genes functional increases from 0 to $2u_c$ rapidly as h increases from 0 to 1. Indeed, the transition from 0 to $2u_c$ occurs in a narrow range of h: for example, when $u_c = 10^{-5}$, it occurs in the range 0–0·02 (Fig. 6). Although recombination is also important for a new duplicate gene to be selectively advantageous, the required amount of recombination is small (of the same order of magnitude as the mutation rate, $> u_c$). In sum, the fixation of a newly arisen duplicate gene can be enhanced under a wide range of reasonable conditions.

A duplicate gene is equivalent to a modifier that reduces the level of dominance of mutations (Fisher, 1928; Wright, 1929), in the sense that both restore the fitness of mutant heterozygotes to the same optimum of the wild-type. Indeed, the selective advantage of u_c can be applied to a dominance modifier that gives complete dominance to the wild-type allele (the maximum case, Fisher, 1929; Haldane, 1930; Wright, 1934; the selective advantage becomes $2u_c$ in these papers, ignoring mutations at the modifier locus). While these authors focused on the rate of frequency change of the modifier, we showed here that the probability of functional fixation becomes $2u_c$.

Our findings account for the much less enhancement of the fixation of a new duplication in Clark (1994) and Lynch *et al.* (2001). Both Clark (1994) and Lynch *et al.* (2001) are based on the double-null

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Table 6. *Probability of functional fixation and its theoretical prediction* (P_{ff}) *in the single-function model with* $u_{\mathrm{c}} = 10^{-3}$

			h				
s	*N*	*c*	0	0·002	0·02	0·2	1
1	P_{ff}		0	0·00012	0·00092	0·00196	0·00200
	50 000	0·5	0·00005	0·00013	0·00095	0·00185	0·00199
	10 000	0·5	0·00015	0·00019	0·00085	0·00211	0·00257
	10 000	0·002	0·00010	0·00013	0·00062	0·00146	0·00159
	10 000	0·001	0·00013	0·00011	0·00044	0·00129	0·00121
	10 000	0·0001	0·00005	0·00007	0·00013	0·00055	0·00072
	10 000	0	0	0	0	0·00004	0·00005
	5000	0·5	0·00024	0·00027	0·00076	0·00202	0·00240
	5000	0·002	0·00023	0·00027	0·00055	0·00152	0·00157
	5000	0·001	0·00018	0·00019	0·00040	0·00121	0·00115
	5000	0·0001	0·00008	0·00007	0·00002	0·00047	0·00057
	5000	0	0	0	0	0·00007	0·00012
	1000	0·5	0·00071	0·00083	0·00114	0·00216	0·00270
	1000	0·002	0·00066	0·00070	0·00111	0·00159	0·00166
	1000	0·001	0·00044	0·00053	0·00081	0·00147	0·00135
	1000	0·0001	0·00019	0·00031	0·00038	0·00079	0·00055
	1000	0	0·00011	0·00012	0·00022	0·00050	0·00054
0·1	P_{ff}		0	0·00004	0·00033	0·00169	0·00200
	50 000	0·5	0·00005	0·00006	0·00033	0·00168	0·00204
	10 000	0·5	0·00015	0·00014	0·00036	0·00180	0·00174
	10 000	0·002	0·00015	0·00014	0·00025	0·00133	0·00161
	10 000	0·001	0·00013	0·00011	0·00023	0·00106	0·00143
	10 000	0·0001	0·00008	0·00007	0·00009	0·00041	0·00071
	10 000	0	0	0	0	0·00002	0·00006
	5000	0·5	0·00023	0·00021	0·00040	0·00178	0·00191
	5000	0·002	0·00018	0·00022	0·00022	0·00114	0·00155
	5000	0·001	0·00018	0·00024	0·00026	0·00099	0·00130
	5000	0·0001	0·00008	0·00009	0·00010	0·00040	0·00062
	5000	0	0	0	0	0·00007	0·00010
	1000	0·5	0·00072	0·00087	0·00117	0·00176	0·00252
	1000	0·002	0·00067	0·00071	0·00085	0·00161	0·00167
	1000	0·001	0·00054	0·00067	0·00066	0·00118	0·00147
	1000	0·0001	0·00026	0·00023	0·00027	0·00058	0·00071
	1000	0	0·00012	0·00011	0·00012	0·00035	0·00064

recessive model ($h = 0$). In addition, Clark (1994) assumed no recombination between the duplicate genes ($c = 0$) as well as low mutation pressure ($Nu_{\mathrm{c}} \leqslant 0{\cdot}1$). Especially under the condition of Clark (1994), we found almost no increase in the fixation probability (Table 2), which is consistent with his result. In any case, when $h = 0$, recurrent deleterious mutations do not enhance the preservation of functional copies (Table 6; Fig. 6).

Recent sequencing of mutation accumulation lines in several model organisms has provided estimates of mutation rate per site per generation of $0{\cdot}3$–21×10^{-9} (Denver *et al.*, 2004; Haag-Liautard *et al.*, 2007; Lynch *et al.*, 2008). Assuming a mutational target size of ~1 kb, the mutation rate per locus would then be approximately 10^{-5}–10^{-6}. These figures appear compatible with previous estimates from specific locus tests (10^{-4}–10^{-6}; see for review, Woodruff *et al.*, 1983; Drake *et al.*, 1998). Given these mutation rates, it is not unexpected to see the Nu_{c} values as high as 10 in certain species. As we have seen above, the masking effect of duplication could have important consequences for genome evolution in these species. When $u_{\mathrm{c}} = 10^{-5}$, the average degree of dominance of mutations ($h \sim 0{\cdot}02$) estimated in *Drosophila* (Simmons & Crow, 1977) is just enough to reach the maximum level of the functional fixation probability ($\sim 2u_{\mathrm{c}}$; see Fig. 6*b*).

In this study, we did not consider potential disadvantage of gene duplication caused by imbalanced gene dosage. Indeed, segmental duplications in the human genome are often associated with diseases,

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

which, together with other structural changes, are called genomic disorders (Stankiewicz & Lupski, 2002). There is further evidence for deleterious effects associated with segmental duplications. Segmental duplications are created by non-allelic homologous recombination (NAHR; ectopic recombination) between repeated sequences (e.g. Goldberg *et al.*, 1983; Roeder, 1983; Chance *et al.*, 1994). The occurrence rate of NAHR was estimated to be $0{\cdot}4\text{–}170\times10^{-6}$ per gene per generation in *Drosophila* (Gelbart & Chovnick, 1979; Shapira & Finnerty, 1986; Watanabe *et al.*, 2009). This rate is more than 400 times larger than the origination rate of gene duplication estimated from the genome sequence analysis ($0{\cdot}001\times10^{-6}$; Lynch, 2007), implying that the majority of duplications are deleterious and rapidly eliminated by purifying selection before reaching fixation.

While neofunctionalization and subfunctionalization have possibly been involved in the retention of duplicate genes in the later stages of evolution, the selective advantage of gene duplication via its masking effect must have played a more important role, together with its direct disadvantage, in the early stage before fixation. If different classes of genes are characterized by distinct levels of heterozygous fitness effects (*hs*), then this could be the primary reason for the non-random distribution of duplicate genes, where certain types of genes, namely those associated with immunity and defense, membrane surface interactions, drug detoxification and growth/development, are overrepresented (Bailey *et al.*, 2002; Nguyen *et al.*, 2006; Perry *et al.*, 2006; Dopman & Hartl, 2007; Graubert *et al.*, 2007). However, the distribution of fitness effects of loss-of-function mutations and spontaneous duplications remains largely undetermined.

There is evidence that duplicate genes are more enriched for haplo-insufficient than haplo-sufficient genes (Kondrashov & Koonin, 2004; see Qian & Zhang, 2008, for a contrasting view). While Kondrashov *et al.* (2002) proposed an increased protein dosage as the primary factor promoting the persistence of duplicate genes (see also Kondrashov & Koonin, 2004), the masking effect of gene duplication is an alternative explanation for the differential preservation of duplicate genes between the two classes of genes with distinct heterozygous fitness effects (*hs*). A long-term persistence of duplicate genes due to the masking effect may increase the chance for neofunctionalization to occur in the future generations.

We thank Hideki Innan for help with simulation procedures and Fyodor Kondrashov for suggestions. This work was supported by a Grant-in-Aid for Scientific Research (C) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (TT-S).

References

Bailey, G. S., Poulter, R. T. M. & Stockwell, P. A. (1978). Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proceedings of the National Academy of Sciences of the USA* **75**, 5575–5579.

Bailey, J. A., Gu, Z., Clark, R. A., Reinert, K., Samonte, R. V., Schwartz, S., Adams, M. D., Myers, E. W., Li, P. W. & Eichler, E. E. (2002). Recent segmental duplications in the human genome. *Science* **297**, 1003–1007.

Bailey, J. A., Church, D. M., Ventura, M., Rocchi, M. & Eichler, E. E. (2004). Analysis of segmental duplications and genome assembly in the mouse. *Genome Research* **14**, 789–801.

Chance, P. F., Abbas, N., Lensch, M. W., Pentao, L., Roa, B. B., Patel, P. I. & Lupski, J. R. (1994). Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Human Molecular Genetics* **3**, 223–228.

Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K., Church, D., DeJong, P., Wilson, R. K., Pääbo, S., Rocchi, M. & Eichler, E. E. (2005). A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**, 88–93.

Clark, A. G. (1994). Invasion and maintenance of a gene duplication. *Proceedings of the National Academy of Sciences of the USA* **91**, 2950–2954.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Minneapolis: Burgess Publishing Company.

Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. (2004). High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682.

Dopman, E. B. & Hartl, D. L. (2007). A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **104**, 19920–19925.

Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.

Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O. & Long, M. (2008). Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* **320**, 1629–1631.

Fisher, R. A. (1928). The possible modification of the response of the wild type to recurrent mutations. *American Naturalist* **62**, 115–126.

Fisher, R. A. (1929). The evolution of dominance; reply to professor Sewall Wright. *American Naturalist* **63**, 553–556.

Fisher, R. A. (1935). The sheltering of lethals. *American Naturalist* **69**, 446–455.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-L. & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545.

Gehrels, N. (1986). Confidence limits for small numbers of events in astrophysical data. *Astrophysical Journal* **303**, 336–346.

Gale, J. S. (1990). *Theoretical Population Genetics*. London: Unwin Hyman.

Gelbart, W. M. & Chovnick, A. (1979). Spontaneous unequal exchange in the rosy region of *Drosophila melanogaster*. *Genetics* **92**, 849–859.

Glémin, S. (2003). How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* **57**, 2678–2687.

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

Goldberg, M. L., Sheen, J.-Y., Gehring, W. J. & Green, M. M. (1983). Unequal crossing-over associated with asymmetrical synapsis between nomadic elements in the *Drosophila melanogaster* genome. *Proceedings of the National Academy of Sciences of the USA* **80**, 5017–5021.

Graubert, T. A., Cahan, P., Edwin, D., Selzer, R. R., Richmond, T. A., Eis, P. S., Shannon, W. D., Li, X., McLeod, H. L., Cheverud, J. M. & Ley, T. J. (2007). A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genetics* **3**, e3.

Gu, X., Wang, Y. & Gu, J. (2002). Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nature Genetics* **31**, 205–209.

Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Charlesworth, B. & Keightley, P. D. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445**, 82–85.

Haldane, J. B. S. (1930). A note on Fisher's theory of the origin of dominance, and on a correlation between dominance and linkage. *American Naturalist* **64**, 87–90.

Haldane, J. B. S. (1933). The part played by recurrent mutation in evolution. *American Naturalist* **67**, 5–19.

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W. & Lee, C. (2004). Detection of large-scale variation in the human genome. *Nature Genetics* **36**, 949–951.

Kimura, M. (1957). Some problems of stochastic process in genetics. *Annals of Mathematical Statistics* **28**, 882–901.

Kimura, M. & King, J. L. (1979). Fixation of a deleterious allele at one of two ''duplicate'' loci by mutation pressure and random drift. *Proceedings of the National Academy of Sciences of the USA* **76**, 2858–2861.

Kimura, M. & Takahata, N. (1983). Selective constraint in protein polymorphism: study of the effectively neutral mutation model by using an improved pseudosampling method. *Proceedings of the National Academy of Sciences of the USA* **80**, 1048–1052.

Kirkpatrick, M. & Jarne, P. (2000). The effects of a bottleneck on inbreeding depression and the genetic load. *American Naturalist* **155**, 154–167.

Kondrashov, F. A. & Koonin, E. V. (2004). A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends in Genetics* **20**, 287–291.

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biology* **3**, research0008.1-0008.9.

Li, J., Jiang, T., Mao, J.-H., Balmain, A., Peterson, L., Harris, C., Rao, P. H., Havlak, P., Gibbs, R. & Cai, W.-W. (2004). Genomic segmental polymorphisms in inbred mouse strains. *Nature Genetics* **36**, 952–954.

Li, W.-H. (1980). Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* **95**, 237–258.

Lindsley, D. L. & Zimm, G. G. (1992). *The Genome of Drosophila melanogaster*. San Diego: Academic Press, Inc.

Lynch, M. (2007). *The Origins of Genome Architecture*. Massachusetts: Sinauer Associations, Inc. Publishers.

Lynch, M. & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473.

Lynch, M., O'Hely, M., Walsh, B. & Force, A. (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804.

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., Dickinson, W. J., Okamoto, K., Kulkarni, S., Hartl, D. L. & Thomas, W. K. (2008). A genome-wide view of spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the USA* **105**, 9272–9277.

Nei, M. & Roychoudhury, A. K. (1973). Probability of fixation of nonfunctional genes at duplicate loci. *American Naturalist* **107**, 362–372.

Nguyen, D.-Q., Webber, C. & Ponting, C. P. (2006). Bias of selection on human copy-number variants. *PLoS Genetics* **2**, 198–207.

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer-Verlag.

Perry, G. H., Tchinda, J., McGrath, S. D., Zhang, J., Picker, S. R., Cáceres, A. M., Iafrate, A. J., Tyler-Smith, C., Scherer, S. W., Eichler, E. E., Stone, A. C. & Lee, C. (2006). Hotspots for copy number variation in chimpanzees and humans. *Proceedings of the National Academy of Sciences of the USA* **103**, 8006–8011.

Qian, W. & Zhang, J. (2008). Gene dosage and gene duplicability. *Genetics* **179**, 2319–2324.

Roeder, G. S. (1983). Unequal crossing-over between yeast transposable elements. *Molecular and General Genetics* **190**, 117–121.

Samonte, R. V. & Eichler, E. E. (2002). Segmental duplications and the evolution of the primate genome. *Nature Reviews Genetics* **3**, 65–72.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., Zetterberg, A. & Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528.

Shapira, S. K. & Finnerty, V. G. (1986). The use of genetic complementation in the study of eukaryotic macromolecular evolution: rate of spontaneous gene duplication at two loci of *Drosophila melanogaster*. *Journal of Molecular Evolution* **23**, 159–167.

She, X., Jiang, Z., Clark, R. A., Liu, G., Cheng, Z., Tuzun, E., Church, D. M., Sutton, G., Halpern, A. L. & Eichler, E. E. (2004). Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**, 927–930.

Simmons, M. J. & Crow, J. F. (1977). Mutations affecting fitness in *Drosophila* populations. *Annual Review of Genetics* **11**, 49–78.

Stankiewicz, P. & Lupski, J. R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends in Genetics* **18**, 74–82.

Takahata, N. & Maruyama, T. (1979). Polmorphism and loss of duplicate gene expression: A theoretical study with application to tetraploid fish. *Proceedings of the National Academy of Sciences of the USA* **76**, 4521–4525.

Walsh, B. (2003). Population-genetic models of the fates of duplicate genes. *Genetica* **118**, 279–294.

Watanabe, Y., Takahashi, A., Itoh, M. & Takano-Shimizu, T. (2009). Molecular spectrum of spontaneous *de novo* mutations in male and female germline cells of *Drosophila melanogaster*. *Genetics* **181**, 1035–1043.

Watterson, G. A. (1983). On the time for gene silencing at duplicate loci. *Genetics* **105**, 745–766.

Wong, K. K., deLeeuw, R. J., Dosanjh, N. S., Kimm, L. R., Cheng, Z., Horsman, D. E., MacAulay, C., Ng, R. T., Brown, C. J., Eichler, E. E. & Lam, W. L. (2007). A comprehensive analysis of common copy-number

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press

variations in the human genome. *American Journal of Human Genetics* **80**, 91–104.

Woodruff, R. C., Slatko, B. E. & Thompson, J. N. Jr (1983). Factors affecting mutation rates in natural populations. In *The Genetics and Biology of Drosophila*, vol. 3c, pp. 37–124. London: Academic Press.

Wright, S. (1929). Fisher's theory of dominance. *American Naturalist* **63**, 274–279.

Wright, S. (1934). Physiological and evolutionary theories of dominance. *American Naturalist* **68**, 24–53.

Xue, C. & Fu, Y. (2009). Preservation of duplicate genes by originalization. *Genetica* **136**, 69–78.

https://doi.org/10.1017/S0016672309000196 Published online by Cambridge University Press