The conservation of redundancy in genetic systems: effects of sexual and asexual reproduction

J A MORRIS* and R D MORRIS

Department of Pathology, Royal Lancaster Infirmary, Lancaster, LA1 4RP, UK

**Corresponding author (Fax,* 1524-583798*; Email, Jim.A.Morris@rli.mbht.nhs.uk)*

The relationship between probability of survival and the number of deleterious mutations in the genome is investigated using three different models of highly redundant systems that interact with a threatening environment. Model one is a system that counters a potentially lethal infection; it has multiple identical components that act in sequence and in parallel. Model two has many different overlapping components that provide threefold coverage of a large number of vital functions. The third model is based on statistical decision theory: an ideal detector, following an optimum decision strategy, makes crucial decisions in an uncertain world. The probability of a fatal error is reduced by a redundant sampling system, but the chance of error rises as the system is impaired by deleterious mutations. In all three cases the survival profile shows a synergistic pattern in that the probability of survival falls slowly and then more rapidly. This is different than the multiplicative or independent survival profile that is often used in mathematical models. It is suggested that a synergistic profile is a property of redundant systems.

Model one is then used to study the conservation of redundancy during sexual and asexual reproduction. A unicellular haploid organism reproducing asexually retains redundancy when the mutation rate is very low (0·001 per cell division), but tends to lose high levels of redundancy if the mutation rate is increased (0·01 to 0·1 per cell division). If a similar unicellular haploid organism has a sexual phase then redundancy is retained for mutation rates between 0·001 and 0·1 per cell division. The sexual organism outgrows the asexual organism when the above mutation rates apply. If they compete for finite resources the asexual organism will be extinguished. Variants of the sexual organism with increased redundancy will outgrow those with lower levels of redundancy and the sexual process facilitates the evolution of more complex forms. There is a limit to the extent that complexity can be increased by increasing the size of the genome and in asexual organisms this leads to progressive accumulation of mutations with loss of redundancy and eventual extinction. If complexity is increased by using genes in new combinations, the asexual form can reach a stable equilibrium, although it is associated with some loss of redundancy. The sexual form, by comparison, can survive, with retention of redundancy, even if the mutation rate is above one per generation.

The conservation and evolution of redundancy, which is essential for complexity, depends on the sexual process of reproduction.

[Morris J A and Morris R D 2003 The conservation of redundancy in genetic systems: effects of sexual and asexual reproduction *J. Biosci.* **28** 671–681]

1. Introduction

The dictionary definition of redundant is "superabundant, superfluous or in excess". This is the way the term is used in common speech. Redundancy also has a technical meaning in information theory: a redundant system is one that reduces error in information processing (Tautz 1992; Morris 2001a). The common and the technical definition

Keywords. Asexual reproduction; complexity, deleterious mutations; genetic system; redundancy; sexual reproduction; synergistic interaction

are different but they are rather more closely related than what appears at first sight. Systems that reduce error often have an excess of components, more than the minimum required for the task in hand. The system as a whole is redundant in the technical sense in that it minimises the risk of error or failure, it could also be argued that some of the components are redundant in the common sense in that they are present in excess. In published works in biology and genetics both the definitions are used and this leads to confusion (Morris 1997; Nowak *et al* 1997; Krakauer and Nowak 1999; Krakauer and Plotkin 2002).

There are conserved genes in yeast and in mice that when deleted have no measurable or observable effect on phenotype (Erickson 1993; Lewin 1997). This was a form of redundancy that came as a surprise to those who use the common definition (Oliver 1996). It led to a number of publications concerned with the question "how can a gene survive long term if deletion has no measurable effect?" (Nowak *et al* 1997; Krakaeur and Nowak 1999; Krakaeur and Plotkin 2002).

Those who use the technical definition of redundancy have a different set of expectations. Redundancy reduces error, and is an essential part of a complex system. Thus the expectation is that redundancy will be found in all aspects of biological function including the genome, the proteome and in morphology and physiology. According to this concept, there will be genes in the genome that code for vital functions but are present in excess and deletion will cause no discernible change. The theoretical problem is to explain how a component can survive, given that it is specified by a number of genes, deletion of any one gene will delete the component but that deletion has no measurable effect. It is obviously a different and a larger problem to maintain a set of genes rather than a single gene in the face of random mutation.

Extensive literature exists on the accumulation of deleterious mutations in the genome; the relationship of the mutational load to fitness and survival; and the effects of sexual and asexual reproduction on this process (Haldane 1937; Muller 1950; Kimura and Marauyama 1966; King 1966; Kondrashov 1982, 1988, 1994; Redfield 1994). In many models it is assumed that the deleterious mutations act independently to decrease fitness. This leads to a multiplicative or independent survival profile i.e. the probability of survival falls progressively less rapidly as the number of deleterious mutations increases. The assumption that the genes are independent means that the models are concerned with single genes rather than genetic systems. On the otherhand, in a synergistic survival profile the probability of survival falls progressively more rapidly as the number of mutations increases. A synergistic survival profile confers advantage on organisms that reproduce sexually, and allows for an equilibrium mutational load when the new mutation rate exceeds one per generation. Population size is also an important consideration, since an equilibrium in a large population can be lost in a small population due to random genetic drift causing the overall fitness to decline (Lynch and Gabriel 1990; Gabriel *et al* 1993). This applies in particular to asexual organisms, as the lightest mutational load cannot be regained once it has been lost. In most of the above mentioned theoretical work, on the dynamical relationship between mutation rate and mutational load, there is no reference to the concept of redundancy; this is because the importance of redundancy for genetic organization has only recently been realized (Tautz 1992). Models published more recently on redundancy and mutation tend to assume an independent survival profile rather than a synergistic profile (Nowak *et al* 1997; Krakauer and Nowak 1999; Krakauer and Plotkin 2002).

Mechanisms of DNA repair are conserved in evolution and the error rate per base pair per cell generation is similar in unicellular and multicellular organisms (Drake *et al* 1998). The most reliable measurements have been made in bacteria, where the mean deleterious mutation rate per cell generation is 0·002. Bacteria have in the region of 5000 genes, yeast have 6000 genes, flies have 13000 genes, worms have 18000 genes, plants have 26000 genes, and the latest estimate for mice and men is 30000 genes in the haploid set i.e. 60000 in the diploid set (Baltimore 2001). Thus, extrapolating from bacteria, the probability of a deleterious mutation per cell division in man will be approximately 0·024. This estimate is of the right order, since the new deleterious mutation rate per human generation exceeds one (Eyre-Walker and Keightley 1999) and the number of cell generations per human generation is variously estimated at 50 to 200 (Crow 1997; Morris 1999).

In this article we explore the properties of redundant systems that reduce error and use models to determine the relationship between probability of survival and the number of accumulated deleterious mutations. One of the models is then used to examine conservation of redundancy in genetic systems in both sexual and asexual reproduction. The models have mutation rates per cell division between 0·001 and 0·1 to cover the range seen in nature.

2. Models of redundancy

2.1 *Model one*

A biological organism (organism A) has a redundant system to protect itself against a potentially lethal parasitic infection. The redundant system has m components, but i of them are inactivated by deleterious mutations. Organism A is attacked by q parasites and the probability that

one parasite can evade any one active component is p. If a parasite evades all (m–i) active components, organism A is killed. It is assumed that each component is specified by a set of genes, and not by a single gene, and that a deleterious mutation in any one of the genes will delete the component.

The probability that organism A survives the parasitic attack is $(1–p^{m–i})^q$. Figure 1 shows how the probability of survival falls with increasing values of i. In this example it is assumed that $m = 7$ and $p = 0{\cdot}1$. The survival curves are for $q = 1000$ and $q = 100{,}000$. When q is 1000 there is very little difference in survival between $i = 0$, $i = 1$ and $i = 2$. The system in this case is highly redundant. When $q = 100{,}000$ the survival profile falls more rapidly. Thus the degree of redundancy varies with environmental pressure and a system with a superabundance of components in one setting might be just adequate in another.

The components, in this model, could act sequentially or in parallel or some combination of the two. For ease of computation it is assumed that the components are identical, but there is no change in principle if the components are different i.e. have different values of p.

2.2 *Model two*

This model concerns overlapping components each specified by a bank or set of genes. Once again, a deleterious mutation in any one gene will delete the component. In this model there are n^2 different functions to perform, and there are $3n$ different components each specifying n functions. The system is arranged so that every function is covered by 3 different components. The organism (organism alpha) will survive if every function is covered by at least one active component.

A square "n by n units" contains n^2 square units each representing a single vital function. The square is completely covered by n components each "n units by one unit" arranged in parallel. A second layer is formed by a further n components, each "n units by one unit", which are laid at right angles to the first layer. The third layer is formed by n components each "(0·5) n units by 2 units". Thus every function is covered three times and every component covers a different set of n functions (see figure 2a).

Let us assume that deleterious mutations occur at random, a single mutation will delete an entire component, and the components have an equal number of genes and are at equal risk of deletion. Organism alpha will survive if only one or two components are deleted but it is at risk of death if three or more are inactivated. The relationship between the probability of survival and the number of deleterious mutations is shown in figure 2b. In this example it is assumed that $n = 10$, and the calculation is as follows: Suppose there are j deleterious mutations with b in the first layer, c in the second and d in the final layer, so $b + c + d = j$; where the values b, c and d are non-negative integers. It is important to note that more than one mutation may lie in the same component.

If one component is deleted from each layer the probability that a function is not covered is 1 over n. If b deleterious mutations are present in the first layer, c in the second and d in the third, then the probability that no function is uncovered is:

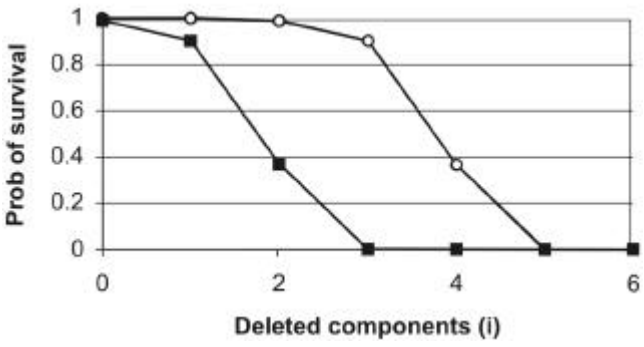

Figure 1. The probability of survival of organism A is defined by $(1 – p^{m\,-\,i})^q$. In this example, $p = 0{\cdot}1$, $m = 7$ and q is 1000 (○) or 100,000 (■).

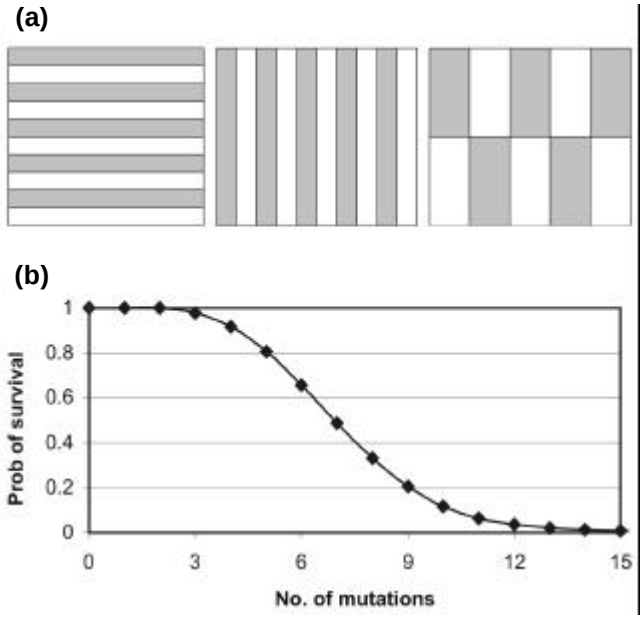

Figure 2. (**a**) A square, "n by n units", has n^2 unit squares, and is covered by three layers. In this example, $n = 10$. Each layer has 10 components; each component covers a different set of 10 unit squares; each unit square represents a vital function and is covered three times. (**b**) The probability of survival of organism alpha.

$$\left(1-\frac{1}{n}\right)^{bcd},$$

since the positions of different mutations are independent events. Hence the probability of survival of an organism with j deleterious mutations is the sum for non-negative integers b, c and d such that $b + c + d = j$, of the following function:

$$\frac{\binom{j}{b}\times\binom{j-b}{c}\times\left(1-\frac{1}{n}\right)^{bcd}}{3^j}.$$

2.3 *Model three*

This model is based on statistical decision theory. Figure 3a shows a classical problem of setting a decision threshold to distinguish to states *h*-0 and *h*-1 given that both are random variables. The frequency distributions of the intensity of some stimulus, shown in arbitrary units, overlap and therefore there is always a finite chance of error in any attempt to make a distinction. One way to reduce error is to take multiple samples and compute the mean. The variance of the sample mean is inversely proportional to the number of samples obtained, thus more samples reduce the variance of the distributions in figure 3a and thereby reduce overlap and error.

Consider a biological system (organism beta) capable of obtaining m samples in parallel through m independent components each specified by a different set of genes. Deleterious mutations inactivate the components reducing the number of samples obtained to $m–i$. Let us assume that the distinction between *h*-0 and *h*-1 is crucial and the errors respond *H*-0 given *h*-1 and respond *H*-1 given *h*-0 are both fatal. The decision threshold, placed approximately at value 10 in figure 3a, is x standard deviations from the mean of both distributions. If $m–i$ samples are obtained and the mean computed the decision threshold will be x times the square root of $m–i$ standard errors of the mean from the true means of the distributions. This follows because the standard error of the sample mean equals the standard deviation of the initial distribution divided by the square root of the number of samples obtained. In the example shown in figure 3b, $x = 1$ and the decision is required q times per individual lifetime. The probability of a correct response can be read from standard tables, raised to the power q and plotted as in figure 3b. Once again a system can appear redundant when the environment is less demanding, but the extra components are needed when conditions are increasingly harsh.

3. Model one: reproduction

3.1 *Organism A (asexual reproduction)*

Let us assume that organism A, described above, is unicellular and has a haploid genome. It also has a highly redundant system with which to resist a potentially lethal parasitic infection. The system has m components, each specified by y genes. A deleterious mutation in one gene will inactivate a component. The number of components deleted by mutation is i and the number of deleterious mutations is j. The probability that a parasite evades an active component is p. If the parasite evades all $m–i$ active components then organism A is killed. If organism A survives attacks by q parasites it then divides by mitosis. The probability of a new deleterious mutation in a daughter cell at division is b. In the models the probability of two new mutations in one daughter cell is ignored, as is the remote possibility of a mutation reverting a deleted gene to an active state.

The probability of survival for a cell with i deleted components is $s_i = (1 - p^{m-i})^q$. The expected number of cells with $i + 1$ deleted components following the division of a

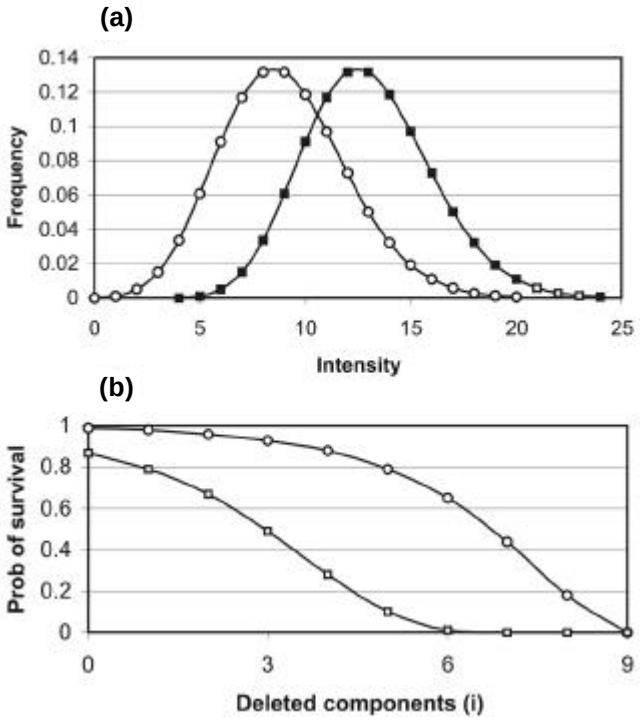

Figure 3. (**a**) The frequency distributions h0 (○) and h1 (■) overlap. Attempts to distinguish h0 and h1 based on the intensity level will lead to possibility of error. The frequency distributions of the sample means, however, will show less variance; therefore there is less overlap and the probability of error is reduced. (**b**) The probability of survival for organism beta for the values $q = 10$ (○) and $q = 100$ (□).

single cell with i deleted components is $2b(m-i)\ (m^{-1})$. Let Y_i be the expected number of cells with i deletions following the next division of a single daughter cell with i deletions, so $Y_i = 2s_i[1 - b(m-i)(m^{-1})]$.

Let X_i be the expected number of cells with $i+1$ deletions following the next division of a single daughter cell with i deletions, so $X_i = 2s_i\, b(m-i)(m^{-1})$.

A computer program, written in Pascal, has been used to follow the fate of a large number of cells in which the initial value of i is 0. In this model it is assumed that $m = 7$, $p = 0{\cdot}1$ and $q = 1000$. Tables 1–3 show representative results. Table 3 gives the per centage of cells for different values of i, and three values of b, after 100,000 generations. The population is close to equilibrium after 10,000 generations and thereafter there is little change (table 2).

When $b = 0{\cdot}1$, the seven-fold ($i = 0$) and the six-fold ($i = 1$) redundant cells die out but the five-fold redundant cells survive. The seven-fold and six-fold redundant cells grow in absolute numbers but fall as a percentage of the total and in a finite population they will eventually disappear (Muller's ratchet) (Muller 1964). If the parasitic load rose from 1000 to 100,000, this asexual population would be wiped out. When $b = 0{\cdot}01$, the seven-fold redundant cells are lost but the six-fold survive. With $b = 0{\cdot}001$, there is no loss of redundancy. If $Y_i < Y_{i+1}$, then those organisms with $i+1$ mutations will always outgrow those with i mutations. The latter will gradually decrease in per centage terms and eventually become extinct. If $Y_i > Y_{i+1} + X_i$, then those organisms with $i+1$ mutations cannot outgrow those with i mutations in the long term, and the former will be part of a stable equilibrium. This inequality implies the following:

$$s_i - s_{i+1} > b[(m-i)(m^{-1})\ (2s_i - s_{i+1}) + (s_{i+1})\ (m^{-1})].$$

The right side of the inequality takes values from $0{\cdot}25\ b$ to $2\ b$ (approximate), although in most cases the range is $(0{\cdot}25)b$ to b. Thus, in practice, if one additional deletion causes a decrease in the probability of survival that exceeds the mutation rate, then the former cell is unlikely to be outgrown and will survive long term. If the decrease in probability of survival is less than the mutation rate, then a stable equilibrium is less likely to form.

If $Y_{i+1} < Y_i < Y_{i+1} + X_i$, the situation is more complicated. If Y_i is towards the lower end of the interval, then that level of redundancy is liable to disappear; if it is towards the upper end, it is more likely to survive and form part of a stable equilibrium.

3.2 *Organism B (asexual)*

This is a unicellular organism with a haploid genome. It has N systems, each with m components, and each component is specified by y genes. Each system deals with a separate external threat equivalent to the parasitic infection. The values of m, p and q are the same for each of the N systems and are equal to the values for organism A. In this case $b = 0{\cdot}001$ for each system, but the mutation rate per cell per division is Nb.

Consider an extreme case in which $N = 100$, then $Nb = 0{\cdot}1$ and the cell will lose seven-fold and six-fold redundancy from all N systems. This occurs because each deletion from a seven-fold or six-fold system causes a small decrease in the probability of survival, which is much less than 0·1. If all 100 systems have five-fold redundancy, then the probability of survival is $0{\cdot}99^{100} \approx 0{\cdot}37$, and the cell will die out.

This process has been followed using the computer model. If the probability of survival falls very slowly with each additional mutation then an equilibrium is reached around a mean survival equal to one minus the mutation rate. Thus if $Nb = 0{\cdot}1$, at equilibrium the mean survival is 0·9 and this applies irrespective of the rate of fall of survival to that point. This is a demonstration of the long established principle that in a large population an equilibrium mutational load exists that is independent of the mutational effect (Haldane 1937; Muller 1950). The equilibrium, however, is not stable in an asexual population. In the computer model, if cells below 0·01% of the population die out, then mean survival of the population progressively falls and the population as a whole dies out. This is an illustration of Muller's ratchet. This unstable equilibrium must be contrasted with the stable equilibrium in which one additional mutation or component deletion causes a fall in survival that exceeds the mutation rate.

3.3 *Organism C (asexual)*

This organism is also unicellular with a haploid genome. It has N systems performing N different functions; all the

Table 1. Values of Y_i and X_i are shown for $i = 0$ to $i = 5$ and $b = 0{\cdot}1$, 0·01 and 0·001. In this example $m = 7$, $p = 0{\cdot}1$ and $q = 1000$.

	$i = 0$		$i = 1$		$i = 2$		$i = 3$		$i = 4$	
b	Yi	Xi	Yi	Xi	Yi	Xi	Yi	Xi	Yi	Xi
0·1	1·800	0·200	1·827	0·171	1·839	0·141	1·706	0·103	0·704	0·032
0·01	1·980	0·020	1·981	0·017	1·966	0·014	1·799	0·010	0·732	0·003
0·001	1·998	0·002	1·996	0·002	1·979	0·001	1·809	0·001	0·735	0·000

Table 2. The percentage of cells with varying levels of redundancy ($i = 0$ to $i = 5$) after 10 to 100,000 generations of asexual growth when $b = 0{\cdot}01$. (As in table 1, $m = 7$, $p = 0{\cdot}1$ and $q = 1000$.)

Generations	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
0	100	0	0	0	0
10	90·49	9·16	0·34	0·00	0·00
100	41·40	42·96	14·70	0·94	0·00
1000	3·42	45·90	47·02	3·63	0·03
10,000	0·01	44·69	51·28	4·00	0·03
100,000	0·00	44·68	51·29	4·00	0·03

Table 3. The percentage of cells for levels of redundancy from $i = 0$ to $i = 5$ after 100,000 generations of asexual growth when $b = 0{\cdot}1$, $0{\cdot}01$ and $0{\cdot}001$.

b	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
0·1	0·00	0·00	46·14	49·29	4·49	0·08
0·01	0·00	44·68	51·29	4·00	0·03	0·00
0·001	40·96	54·18	4·85	0·00	0·00	0·00

functions are essential for survival**.** Each system has m components and each component is specified by y genes. Each gene, however, contributes to z components. Thus the total number of genes is $Nmyz^{-1}$. The mutation rate per cell per generation is Nbz^{-1}. Furthermore, the decrease in survival with each additional mutation is increased approximately z times.

This organism can, in theory, reach stable equilibria without loss of redundancy. Consider the example in which $m = 7$, $p = 0{\cdot}1$, $q = 1000$, $b = 0{\cdot}001$, $N = 100$ and $z = 10$. The decrease in the probability of survival with each deleted component is increased ten-fold (since $z = 10$), as is the mutation rate (since $Nz^{-1} = 10$); thus the organism will retain a similar level of redundancy as organism A when $b = 0{\cdot}001$.

There is no direct evidence to show that genes act in this way but there is indirect evidence to indicate that it is likely. There is only a six-fold increase in the number of genes (haploid) from bacteria to mice and men (Baltimore 2001). The increase in complexity, at least from our perspective, seems much greater. The rate of mutation places a limit on the extent to which complexity can be increased by increasing the number of genes; but there is no limit to the number of ways in which genes can be used in new combinations to produce new functions and capabilities.

3.4 *Organism D (asexual)*

This organism is specified in the same way as organism C, but is multicellular. In a multicellular organism, the N functions are performed by the whole organism and selection acts at that level – there is no effective selection for most of the functions at the cellular level. Thus the mutation rate, between bouts of selection, is $Nbz^{-1}k$, where k is the number of cell generations between successive gametes.

Thirty mitotic divisions, without cell loss, produce 1 g of tissue, forty generations produce 1 kg of tissue, and fifty generations produces 1000 kg of tissue. Thus in practice, k will take values between 30 and 50 generation. In these models this means some loss of redundancy, but it is possible for a multicellular asexual organism to reach a stable equilibrium in a stable environment and survive long term. It will, however, lose some redundancy and be at a risk of extinction if the environment becomes more threatening.

3.5 *Organism E (sexual reproduction)*

Organism E is unicellular and haploid. Its specification is similar to organism A, but after every ten asexual generations there is a sexual phase in which the cells fuse at random to form diploid cells and these undergo a meiotic division to form a new generation of haploid cells. It is assumed that this leads to a Poisson distribution of deleterious mutations in the progeny. Strictly the variance will be slightly less than Poisson because of the possibility of close positioning of deleterious mutations on the same or opposing chromosomes (Bulmer 1980).

Let c = the mean number of mutations per cell, and let $d = cm^{-1}$ be the mean number of mutations per component at the onset of the sexual phase (meiosis), and after 10 rounds of asexual division (mitosis). The probability that a specific component has no mutations is e^{-d}, and the probability that it has at least one mutation is $1 - e^{-d}$. The probability that a post-meiotic cell has $m–i$ functioning components can be calculated using binomial theory. The appropriate function is:

$$\binom{m}{i}(e^{-d})^{m-1}(1-e^{-d})^{i}.$$

The post-meiotic cells then undergo 10 rounds of mitosis; this uses the computer program as in organism A, but with modifications as below. After each mitotic division the mean number of mutations per component and per surviving cell is calculated. There are three possible pathways leading to a daughter cell with i deletions: a mutation may occur in a cell with $i - 1$ deletions, no mutation may occur in a cell with i deletions; or a mutation may occur in a deleted component of a cell with i deletions. The expected number of mutations for each surviving cell is calculated after each generation based on the expectations in the above three pathways. The final values c and d are then calculated for the surviving cells after

10 mitotic generations and the process of sexual reproduction is repeated.

Table 4 shows the equilibrium distribution for $b = 0{\cdot}1$, $b = 0{\cdot}01$ and $b = 0{\cdot}001$. If tables 3 and 4 are compared, it can be seen that sexual reproduction preserves redundancy. Thus when $b = 0{\cdot}1$, the asexual organism loses seven-fold and six-fold redundancy but this is preserved in the sexual organism. When $b = 0{\cdot}01$, seven-fold redundancy is lost by the asexual organism but remains at the modal level in the sexual phase.

It is possible, using the computer model, to run organism A and organism E in parallel (table 5 and figure 4). It is assumed that they compete for resources in a large but finite environment. If $b = 0{\cdot}1$, organism A forms 75% of the total population after 60 asexual generations, but falls to 10% of the population after 230 generations and becomes extinct (less than 0·01%) after 480 generations. If $b = 0{\cdot}01$, the asexual population forms 70% after 300 generations, but falls to 10% after 1300 generations and dies out after 3000 generations. Even when $b = 0{\cdot}001$ the sexual form outgrows the asexual in the long term. Thus the asexual forms 74% of the total population after 1000 generations, but falls to 10% after 10,000 generations and dies out after 28,000 generations.

Table 4. The percentage of cells after 10,000 mitotic generations (1000 sexual generations) for $i = 0$ to $i = 5$ and $b = 0{\cdot}1$, 0·01 and 0·001.

b	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
0·1	8·20	24·64	31·72	22·68	9·73	2·51
0·01	42·19	38·75	15·25	3·34	0·44	0·03
0·001	75·14	21·92	2·74	0·19	0·01	0·00

Consider two variants of organism E: one in which $m = 7$ and $b = 0{\cdot}1$, and one in which $m = 8$ and $b = 0{\cdot}1 \times 8 \div 7 \approx 0{\cdot}14$ (table 6). The increased mutation rate reflects the increased number of components. The new variant with $m = 8$ can be considered an evolutionary advance with more complexity and more redundancy. If the variants are run in parallel the one with $m = 8$ outgrows and replaces $m = 7$. In the same way $m = 9$ will outgrow $m = 8$, and $m = 10$ outgrows $m = 9$, but $m = 11$ is outgrown by $m = 10$ (in all cases $p = 0{\cdot}1$, $q = 1000$ and the value of b is adjusted to match the increased number of components). This comparison demonstrates the power of sexual reproduction to maintain redundancy; when $m = 10$ the probability of survival is 0·9999999, when $m = 9$ it is 0·999999. Sexual reproduction can maintain this minute survival advantage, but asexual reproduction cannot.

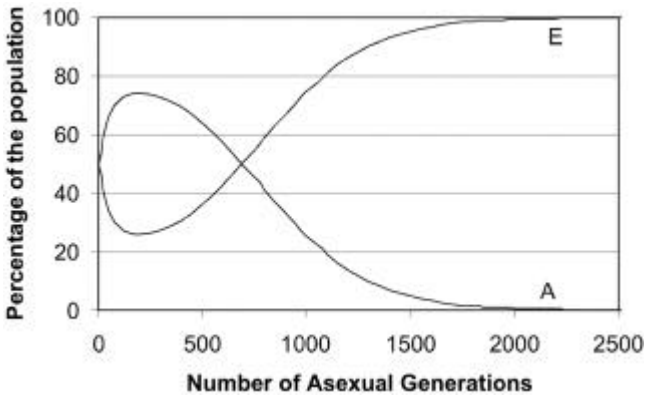

Figure 4. The graph shows the relative sizes of the sexual and asexual populations over the first 2500 asexual generations when organisms A and E are grown together (as in table 5) and $b = 0{\cdot}01$. (A, organism A; E, organism E).

Table 5. Organism A (asexual) and organism E (sexual) are grown together in the same environment and compete for finite resources. In each case ($b = 0{\cdot}1$, 0·01 and 0·001), organism E outgrows and eventually displaces organism A.

$b = 0{\cdot}1$		Asexual generations					
		50	100	150	200	250	500
	org A	75·71	66·50	42·55	18·35	5·80	0·01
	org E	24·29	33·50	57·45	81·65	94·20	99·99
$b = 0{\cdot}01$		Asexual generations					
		100	200	500	1000	2000	5000
	org A	71·60	74·25	63·95	25·52	0·62	0·00
	org E	28·40	25·75	36·05	74·48	99·38	100·00
$b = 0{\cdot}001$		Asexual generations					
		500	1000	2000	5000	10000	25000
	org A	73·66	73·72	69·41	43·73	9·66	0·03
	org E	26·34	26·28	30·59	56·27	90·34	99·97

3.6 *Organism F (sexual)*

This organism is the sexual equivalent of organism D. It is multicellular and has N functions. The mutation rate between zygotes is $Nbz^{-1}k$. It is assumed that deleterious mutations follow a Poisson distribution in zygotes.

The expected number of mutations per component is $r = zc(mN)^{-1}$, where c is the number of deleterious mutations per cell. The probability that a component has at least one deleterious mutation is $1 - e^{-r}$; the probability that the component is active is e^{-r}.

Let $P(i)$ be the probability of i deletions in any one system, so

$$P(i) = \binom{m}{i} (e^{-r})^{m-i} (1 - e^{-r})^{i}.$$

The probability of survival for a system with i deletions is s_i, as before.

Let t_i be the number of systems with i deletions. The probability of survival of the organism is then

$$\prod_i S_i^{t_i},$$

where the product is over all possible values of i. The model estimates this probability using $NP(i)$, which is the expected number of systems with i deletions, rather than t_i. The product is for all values of i from 0 to $m - 1$: this is to avoid multiplying by 0, as $s_i = 0$ when $i = m$.

Figure 5 shows the distribution of deleterious mutations in zygotes and in survivors, assuming that $N = 60$, $z = 5$, $b = 0{\cdot}001$, $p = 0{\cdot}1$, $q = 1000$, $m = 7$, and $Nbz^{-1}k = 1$. The Poisson distribution in zygotes has a mean of 7·75 mutations per cell. The distribution in survivors has a mean of 6·75. The computer model goes through a series of iterations to produce this equilibrium point. In this example 65% of the population of zygotes survive.

The multicellular sexual organism can, therefore, survive long term, even with a mutation rate of 1 per generation. It maintains a high level of redundancy in the population as a whole and a small per centage of the organisms have few or no mutations.

4. Discussion

In using the concept of redundancy there is a problem of definition. It is possible to emphasize the positive aspects of redundancy, as in this article: to consider complex systems undertaking vital functions and to be reassured that there is more than the minimum number of components so that the risk of error is reduced. It is also possible to emphasize the negative aspects: if components are present in excess, then some are inessential and surplus to requirements.

The negative view of redundancy has predominated in recent publications (Nowak *et al* 1997; Krakauer and Nowak 1999; Krakauer and Plotkin 2002), but there are exceptions (Tautz 1992). The negative view leads to surprise that there is redundancy in the genome and a need to explain it away. It leads to the idea that true redundancy or perfect redundancy is the condition in which genotype AB has no advantage over A or B (Nowak *et al* 1997). It leads to concentration on the actions of single genes rather than genes in systems, and to the idea that gene duplications are a form of redundancy. It also leads to the prediction that redundant genes act independently and the fitness profile will be proportional to $(1 - s)^j$, where s is the deleteriousness of each mutation and j is the number of mutations (Krakauer and Plotkin 2002).

Table 6. Two variants of organism E, one with $m = 7$ and one with $m = 8$ are grown together in the same environment and compete for finite resources. The mutation rate per gene per generation is the same, so b is slightly greater for $m = 8$ than for $m = 7$. In each case $m = 8$ outgrows and eventually displaces $m = 7$.

		Sexual generations					
		10	20	50	100	200	500
$b = 0{\cdot}1$	$m = 7$	35·82	32·29	23·20	12·36	2·98	0·03
$b = 0{\cdot}14$	$m = 8$	64·18	67·71	76·80	87·64	97·02	99·97
		Sexual generations					
		50	100	200	300	500	1000
$b = 0{\cdot}01$	$m = 7$	37·08	29·06	16·53	8·74	2·19	0·06
$b = 0{\cdot}014$	$m = 8$	62·92	70·94	83·47	91·26	97·81	99·94
		Sexual generations					
		250	500	750	1000	2000	5000
$b = 0{\cdot}001$	$m = 7$	34·12	22·86	14·49	8·84	1·03	0·00
$b = 0{\cdot}0014$	$m = 8$	65·88	77·14	85·51	91·16	98·97	100·00

With this profile the rate of fall is progressively less as j increases (a multiplicative or independent profile).

The concept of redundancy as error reduction leads to the prediction that it will be essential for complex systems and will be found in all aspects of biological function including genetic structure. In order to build models it is necessary to concentrate on genetic systems rather than single genes, and in particular develop models that interact with a changing environment.

The three models of redundancy, described in the first part of this article, show a survival profile that falls slowly and then more rapidly as i or j is increased. This is a synergistic fitness profile rather than the independent profile described above. This synergistic profile is intermediate between the quadratic and truncated viability profiles described by Redfield (1994). It is sometimes referred to as quasi-truncation selection. The first and the third models interact with the environment and it is possible to show that when external conditions are more threatening the degree of redundancy apparently decreases i.e. more components are required than in normal circumstances and the degree of excess is diminished. The second model is designed to show that even when all components have different functions, but the functions overlap, then the same synergistic profile is produced. Although all models of complex biological processes are abstractions and none is perfect, the fact that three different models of redundant systems predict similar synergistic curves indicates that this is a property of redundancy.

This synergistic profile, however, creates a number of theoretical problems. How can high levels of redundancy survive in the face of recurrent mutation? If there is little survival difference between $i = 0$ and $i = 1$, then recurrent mutation will lead to loss of a component. This is a bigger problem with a genetic system than with a single gene as the risk of mutation is proportionately greater. It is also important to note that redundancy must be preserved in times when the environment is less threatening in order to survive through periods when conditions are harsher.

There has been considerable debate about the relative advantages and disadvantages of sexual and asexual reproduction. Females who reproduce sexually pass on half as many genes to the next generation as those who reproduce asexually – there must be a major advantage to sexual reproduction to compensate for this two-fold disadvantage. One advantage of the sexual state is that gene mixing impairs the co-evolution of parasites and protects against lethal infection (Hamilton *et al* 1990). A second advantage, and the one considered in this article, is that sexual reproduction distributes deleterious mutations unequally in the next generation (Kondrashov 1988, 1994). This means that progeny can have fewer mutations than their parents and the progressive build up of deleterious mutations can be limited.

The relationship between survival, the accumulation of deleterious mutations and the mode of reproduction is complex. In general, if the new mutation rate per generation exceeds one and there is synergistic interaction bet-

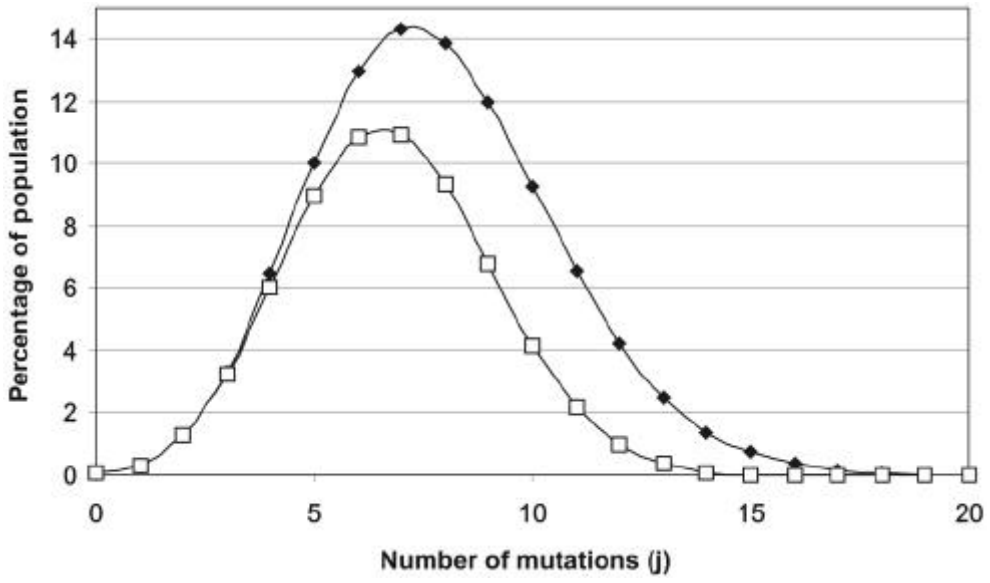

Figure 5. The frequency distribution of number of mutations (j) in zygotes (♦) and in the germ line of adult (□) survivors. At equilibrium the difference in the mean values of the two distributions equals the mean of the new mutation rate per generation.

ween mutations then sexual reproduction is advantageous (Kondrashov 2001). If the mutation rate is less than one and mutations act independently then the advantage of the sexual process is less clear cut.

The simplest way to compare sexual and asexual reproduction is to consider a unicellular haploid organism growing by mitosis, with or without a diploid phase and a reduction division with maximum recombination. Organism A grows asexually, in the process it loses redundancy by mutation. If $b = 0{\cdot}1$ then seven-fold and six-fold redundancy are lost. If $b = 0{\cdot}01$, seven-fold redundancy is lost. Only if $b = 0{\cdot}001$ is seven-fold redundancy retained in some of the cells. This organism is at risk of extinction if the environment changes and the parasite load increases. Organism E is the sexual equivalent of organism A, it grows by mitosis but has a sexual phase. In the sexual organism an equilibrium is reached in which the redundancy is retained. Thus the organism would survive even if the parasitic load were increased markedly. In addition organism E outgrows organism A at all values of b. This occurs because a seven-fold redundant cell produces over 1000 progeny in 10 generations, a five-fold or a four-fold cell, by comparison, produces far fewer because the survival per cell is reduced. If organism E and organism A compete in the same environment for finite resources the latter will be extinguished.

Variants of organism E with increased values of m will outgrow those with lower values. This means that once the sexual process is established the evolution of complexity through increased redundancy is possible.

Increasing complexity by increasing genome size poses problems for asexual organisms. The mutation rate is directly proportional to the size of the genome and this means that some redundancy will be lost. Furthermore, if survival falls in small steps with each mutation, the organism will not reach a stable equilibrium and the progressive accumulation of mutations will lead to extinction. If complexity can be increased, however, by using genes in different combinations to perform different functions then the rise in the mutation rate is less than the rise in complexity. In addition decrements in survival with each mutation will be greater and a stable equilibrium, in theory, is possible. Under these circumstances asexual organisms could survive long term, but there would be some loss of redundancy and the organism would be at risk if the environment changed.

The increase in the number of functioning genes in the genome between bacteria and men is much less than expected given the apparent increase in complexity. It is therefore an attractive idea to suggest that during the evolution of complexity, genes can be used in new combinations to perform new tasks. There are countless ways of combining 30000 genes to form new genetic systems and therefore no limit to the complexity that can be obtained. There is no direct evidence that this is how the genome is organized, but there is a large number of RNA transcripts that do not directly code for proteins and there are suggestions that these are used in genomic control (Mattick 1994). The complexity of the proteome exceeds that of the genome, since 30000 genes code for over 250000 proteins (Banks *et al* 2000). This is achieved by differential splicing of RNA products and requires a high level of regulation of functioning genes. If this analysis is correct, it implies that evolutionary steps are likely to arise by using old genes in new combinations rather than through the emergence of new genes. It also implies that evolution is concerned with genetic systems rather than single genes.

The synergistic survival profile of redundant systems enables multicellular sexual organisms to survive even when the mutation rate per generation exceeds one. The organisms that fail to survive remove several mutations from the population and those that survive carry only a few deleterious mutations as a result. There is some loss of redundancy in the process but there are always a few organisms in which the mutational load is very small. Selective mating, rather than random mating, can, in theory, reduce the load even further (Morris *et al* 2002).

The diploid state in multicellular sexual and asexual organisms is another complication. It is a form of redundancy in that there are two genes for each protein, and if both genes are absent recessive disease occurs. Consideration of the rate of recessive disease in the children of cousin marriages and the progeny of sibling incest indicate that the mean number of deleterious mutations in humans is between 4 and 12 (Morris 2001b). Given that at least one new mutation arises in each human generation (Eyre-Walker and Keightley 1999) there must be selection against the heterozygote deletion or the number of deleterious mutations would be much greater than 12. This observation is consistent with the fact that lethal recessives in *Drosophila* show fitness decrements in heterozygotes (Simmons and Crow 1977). Even an identical allelic pair of genes must be used differently at some stage of development to allow differential selection. There are ways this could occur using methylation and differential regulation. Given differential selection it is reasonable to build models, as in this article, using the haploid state.

In conclusion, there is a tendency for redundancy to be lost in asexual reproduction. In the absence of a stable equilibrium this loss will lead eventually to extinction. Even in the presence of a stable equilibrium loss of redundancy means that the organism risks extinction if the environment becomes more threatening. Sexual reproduction, by comparison, preserves redundancy. The simplest form of unicellular haploid organism gains an advantage from a sexual phase; the conservation of redundancy enables it outgrow and displace its asexual rela-

tion. The models also indicate that sexual organisms with higher levels of redundancy will outgrow those with lower levels even though the difference in survival is too small to measure. These models show that redundancy in the genome is not about superfluous components; on the contrary it concerns essential components: redundancy maintains integrity, preserves function, and confers survival advantage. The conservation and evolution of redundancy is only possible because the otherwise steady erosion of complexity by mutation is reversed by the stochastic process of sexual reproduction.

References

Baltimore D 2001 Our genome unveiled; *Nature* (*London*) **409** 814–816

Banks R E, Dunn M J, Hochstrasser D F, Sanchez J-C, Blackstock W, Pappin D J and Selby P J 2000 Proteomics: new perspectives, new biomedical opportunities; *Lancet* **356** 1749–1756

Bulmer M G 1980 *The mathematical theory of quantitative genetics* (Oxford: Oxford University Press)

Crow J F 1997 The high spontaneous mutation rate: is it a health risk?; *Proc. Natl. Acad. Sci. USA* **94** 8380–8386

Drake J W, Charlesworth B, Charlesworth D and Crow J F 1998 Rates of spontaneous mutation; *Genetics* **148** 1667–1686

Eyre-Walker A and Keightley P D 1999 High genomic deleterious mutation rates in hominoids; *Nature* (*London*) **397** 344–347

Erickson H P 1993 Gene knockouts of c-src, transforming growth factor beta-1 and tenascin suggests superfluous, nonfunctional expression of proteins; *J. Cell Biol.* **120** 1079–1081

Gabriel W, Lynch M and Burger R 1993 Muller's rarchet and mutational meltdown; *Evolution* **47** 1744–1757

Haldane J B S 1937 The effects of variation on fitness; *Am. Nat.* **71** 337–349

Hamilton W D, Axelrod R and Tanese R 1990 Sexual reproduction as an adaptation to resist parasites (a review); *Proc. Natl. Acad. Sci. USA* **87** 3566–3573

Kimura M and Maruyama T 1966 The mutation load with epistatic gene interactions in fitness; *Genetics* **54** 1337–1351

King J L 1966 The gene interaction component of the genetic load; *Genetics* **53** 403–413

Kondrashov A S 1982 Selection against harmful mutations in large sexual and asexual populations; *Genet, Res. Camb.* **40** 325–332

Kondrashov A S 1988 Deleterious mutations and the evolution of sexual reproduction; *Nature* (*London*) **336** 435–440

Kondrashov A S 1994 Muller's ratchet under epistatic selection; *Genetics* **136** 1469–1473

Kondrashov A S 2001 Sex and U; *Trends Genet.* **17** 75–77

Krakauer D C and Nowak M A 1999 Evolutionary preservation of redundant duplicated genes; *Cell Dev. Biol.* **10** 555–559

Krakauer D C and Plotkin J B 2002 Redundnacy, antiredundnacy, and the robustness of genomes; *Proc. Natl. Acad. Sci. USA* **99** 1405–1409

Lewin B 1997 *Genes V1* (Oxford: Oxford University Press)

Lynch M and Gabriel W 1990 Mutational load and the survival of small populations; *Evolution* **44** 1725–1737

Mattick J S 1994 Introns: evolution and function; *Curr. Opinion Genet. Dev.* **4** 823–831

Morris J A 1997 Genetic control of redundant systems; *Med. Hypotheses* **49** 159–164

Morris J A 1999 The kinetics of epithelial cell generation; its relevance to cancer and ageing; *J. Theor. Biol.* **199** 87–95

Morris J A 2001a Information theory: a guide in the investigation of disease; *J. Biosci.* **26** 15–23

Morris J A 2001b How many deleterious mutations are there in the human genome?; *Med. Hypotheses* **56** 646–652

Morris R D, Morris K L and Morris J A 2002 The mathematical basis of sexual attraction; *Med. Hypotheses* **59** 475–481

Muller H J 1950 Our load of mutations; *Am. J. Hum. Genet.* **2** 111–176

Muller H J 1964 The relation of recombination to mutational advance: *Mutational advance*; *Mutat. Res.* **1** 2–9

Nowak M A, Boerlijst M C, Cooke J and Maynard-Smith J 1997 Evolution of genetic redundnacy; *Nature* (*London*) **388** 167–171

Oliver S G 1996 From DNA sequence to biological function; *Nature* (*London*) **379** 597–600

Redfield R J 1994 Male mutation rates and the cost of sex for females; *Nature* (*London*) **369** 145–147

Simmons M J and Crow J F 1977 Mutations affecting fitness in *Drosophila* populations; *Annu. Rev. Genetics* **11** 49–78

Tautz D 1992 Redundancies, development and the flow of information; *Bioessays* **14** 263–266

MS received 6 June 2003; accepted 22 October 2003

Corresponding editor: VIDYANAND NANJUNDIAH