Primer

Genetic linkage and molecular evolution

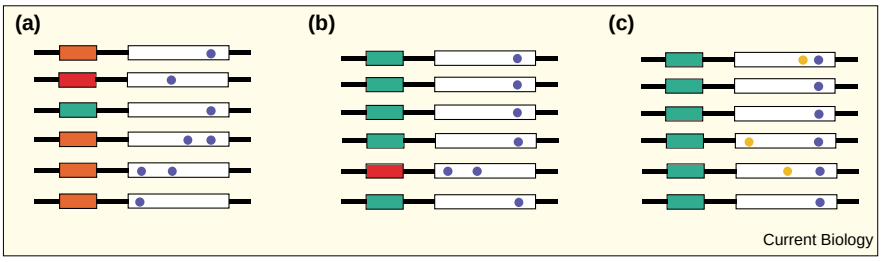
Isabel Gordo and Brian Charlesworth

The rate of genetic recombination varies among species, and across different regions of the genome within a species. In the extreme case, all genes in an asexual or self-fertilizing species are effectively completely linked, in contrast to the substantial opportunities for recombination in outbreeding sexual species. Several key observations on the relation between patterns of evolution and levels of genetic recombination have been made. First, asexual or highly self-fertilizing species tend to be young in terms of the evolutionary timescale, suggesting that they become extinct more rapidly than their recombining relatives. Second, organelle and clonally transmitted genomes in several taxa show reduced levels of adaptation with respect to RNA and protein sequences. In *Drosophila*, regions where the recombination rate is low, such as near centromeres and telomeres, show reduced levels of codon bias — the non-random use of alternative codons encoding the same amino acid — suggesting a reduction in the ability of selection to maintain this aspect of molecular adaptation. Finally, in a number of different species, genomic regions with restricted recombination show lower levels of DNA sequence variation. One of the best examples is provided by non-recombining Y chromosomes, or neo-Y chromosomes formed by fusions between autosomes and sex chromosomes, which show degeneration of gene function and reduced genetic variability.

Why should there be a relation between the amount of recombination and levels of variation and adaptation? As Fisher and Muller pointed out, the dynamics of a given gene are influenced by the evolutionary forces acting on the gene itself, as well as by forces acting at linked loci. Thus the predictions of single-locus population genetics must be modified when selection is acting on sets of linked loci. Here we describe some of the processes that may shape evolution when recombination is restricted over a large genomic region, and which may explain the above observations. These processes all reflect a general effect first quantified by Hill and Robertson: a locus linked to another locus under directional selection experiences a reduced effective population size, N_e. The extent of random fluctuations in allele frequencies due to finite population size, genetic drift, is inversely related to N_e; thus selective differences at one locus tend to enhance the effects of drift at a linked locus.

According to the neutral theory of molecular evolution, mutation and genetic drift have a major influence on DNA sequence variation within species and on differences between species. Mutation creates new neutral variants, with no significant fitness effects, and genetic drift causes random changes in their frequencies until fixation or loss. Kimura showed that the level of variation within a population at a neutral locus is proportional to the product of N_e and the neutral mutation rate, μ, and that the rate of sequence evolution is equal to μ. A mutation with a selection coefficient, s, which measures the reduction or increase in the fitness of its carriers relative to that of the rest of the population, is effectively neutral if $N_e s << 1$. A strongly deleterious mutation, for which $N_e s >> 1$, will be rapidly eliminated, but a weakly deleterious mutation, with $N_e s < 1$, can persist and even become fixed in the population through genetic drift. Similarly a favourable mutation will have almost the same chance of loss from the population as a neutral mutation if $N_e s < 1$. Changes in N_e caused by different kinds of selection at linked loci can thus greatly affect both genetic variability and the efficacy of selection.

Figure 1



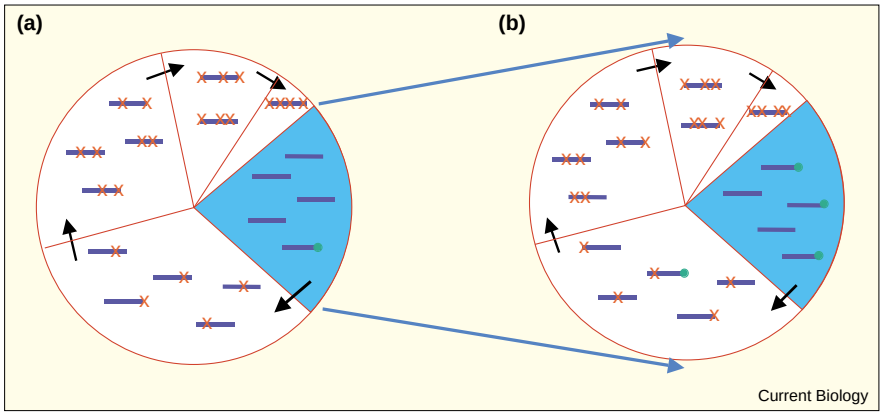
(a) In a population with a large amount of standing variation (blue) at a neutral locus (white), an advantageous mutation arises at a completely linked locus (green). (b) Because of its higher fitness, it will increase in frequency in the population. As it spreads through the population, it wipes out variation at the neutral locus. As the mutation becomes fixed, all variation at the locus is lost. (c) Some time after fixation, neutral variation starts to build up in the population through new mutations (yellow). Since the new neutral variants are relatively young, their frequencies in the population are very low. If we take a sample from the population and sequence alleles of the neutral locus, we will tend to observe mutations that appear only once in the sample. This pattern is very different from that expected in a population at statistical equilibrium under mutation and genetic drift, where some variants will occur at intermediate frequencies.

Selective sweeps

Consider the effect of an advantageous mutation on the level of variation at a completely linked

neutral locus (Figure 1). A peculiar footprint is left just after such a mutation has swept through the population to fixation: variation is drastically reduced, and any variants that can be observed occur at low frequencies in the population. From the point of view of the neutral locus, it is as if the population went through a bottleneck of one individual, and then expanded to its normal size. If an advantageous mutation occurs in a region where there is a small amount of recombination, neutral variation will not be completely lost, but will still be reduced. The level of variation in a particular region of a chromosome depends on the rate of adaptive evolution and on the ratio of the strength of selection to the recombination rate in that region. If selective sweeps continually occur throughout the genome of a sexual species, there will be a correlation between neutral variability and the local recombination rate. Recurrent selective sweeps will also cause very low levels of variation in an asexual species, even if it has a large population size. Sweeps can also result in the fixation of linked deleterious alleles, and this has been proposed as a possible factor in the degeneration of Y chromosomes.

Figure 2



(a) Consider a non-recombining population of size N_e, with ***m*** mutable loci. Each locus undergoes recurrent mutations to deleterious alleles (x), at a rate ***u*** per locus per generation. If the deleterious effect, ***s***, of a mutation is such that $N_e s >> 1$, deleterious alleles stay at low frequencies, and back-mutation is negligible. Thus, a balance between the rate of production of new mutations and of their elimination by natural selection is established. The population is divided into various classes: the least-loaded class, free of deleterious mutations (chromosomes in the blue zone), and other classes having some deleterious mutations (in the red zone, chromosomes have 1, 2 etc. mutations). All the chromosomes with deleterious mutations are eventually eliminated by selection, and new ones are derived from chromosomes free of such mutations (as indicated by black arrows). The least-loaded class is the ultimate source of all future lineages in the population. With independent effects of different loci on fitness, its frequency is $f_0 = \exp(-U/s)$, where $U = mu$ is the deleterious mutation rate for the entire chromosome. Suppose now that a neutral or weakly selected variant arises (in green). If this variant is on a chromosome with one or more strongly deleterious mutations, it will soon be eliminated. The only way it can survive is if it arises in the least-loaded class – this effect is known as background selection (**b**).

Background selection

Any natural population is subject to a continual rain of deleterious mutations at loci throughout the genome. If such mutations are strongly selected against, they are constantly being eliminated. Such purifying selection can have a remarkable effect when linkage is complete (Figure 2). Because every chromosome with a deleterious mutation is destined to be lost sooner or later, the ancestry of future generations is derived from the fraction, f_0, of the current population that is free of deleterious mutations. The population thus has its effective size and level of neutral variation reduced by a factor of f_0. The rate of neutral evolution is independent of N_e, and so remains unchanged. Now consider the effect of background selection when there is a certain frequency of recombination, r, between a neutral variant and a deleterious mutation with which it is associated. The neutral variant can now unhitch itself from the deleterious mutation through recombination. As the mean time spent by a deleterious allele in a large population is roughly s^{-1}, the neutral variant has a chance of survival if it unhitches itself during this time. Thus r/s is an important ratio in determining the effect of background selection. Mathematical analysis shows that this effect can be indeed approximated by a reduction in N_e that involves r/s, and yields an expression relating the level of variation in a region of the genome to the local rate of recombination. This can largely explain the *Drosophila* data mentioned at the start.

If the sites concerned are not evolving neutrally but are very weakly selected, as with synonymous changes to codons, the effect of background selection can be quantified as though N_e is reduced in the same manner as for neutral sites. Because the rate of evolution at weakly selected sites depends on N_e, the rate of fixation of weakly deleterious mutations is increased, and that of advantageous mutations is decreased, if background selection is operating. In regions of the genome where linkage is very tight, such as the Y chromosome, levels of variability and adaptation should both be reduced. Because background selection impairs the ability of an asexual species to adapt, it may also accelerate the extinction of asexual populations.

Muller's ratchet

The background selection model assumes that the frequency of the least-loaded class is stable. But this need not be the case. When f_0 is small, Muller pointed out that genetic drift could then cause the loss of the least-loaded class — a 'click' of the ratchet. After this click, the class with one deleterious mutation becomes the new least-loaded class. But this class can also be lost in the same way as the previous one, so that successive clicks occur, leading to a continual accumulation of deleterious mutations to the detriment of the population's mean fitness — a process now known as Muller's ratchet. As a consequence of each click, fixation of a deleterious allele occurs in the entire population; such fixations would effectively be impossible in the absence of the ratchet. As with background selection, the ratchet can cause a strong reduction in the level of variability at neutral loci.

Muller argued that recombination is the only way to recreate the fittest individuals, and suggested that avoidance of the ratchet provides an advantage to recombination. The ratchet may also be involved in Y chromosome degeneration and the extinction of asexual populations. The frequency at which the ratchet clicks is a very important determinant of its evolutionary effects. Theoretical work shows that the ratchet clicks faster with smaller population size, higher mutation rates, and smaller selection coefficients. Very large asexual species with small genomes are probably immune to the ratchet, but species of smaller sizes with large genomes and no recombination will probably suffer from its consequences.

Weak Hill-Robertson effects

Suppose now that many sites in a tightly linked genome are undergoing mutations to very weakly selected alleles, for which $N_e s$ is around one. Such mutations segregate at much higher frequencies than those involved in the previous processes, and there is now a non-negligible chance of back-mutation. An example of this is provided by synonymous mutations affecting codon usage, which may affect the efficiency or accuracy of translation. Models of genomes with many sites subject to such weakly selected mutations show that, as linkage increases, both intra-species variability and the mean level of adaptation — measured by the frequency with which optimal codons are used — decrease. This effect results from the cumulative effect of numerous polymorphic selected sites on N_e, and is larger as more sites are involved. This process may be of importance in the evolution of any sizeable genome where linkage is complete, and in genomic regions with low levels of recombination.

The effect of balancing selection

There is one situation, however, where selection at a locus can *increase* the level of variability at linked neutral sites. This is when two or more alternative alleles at the selected locus are maintained in the population by balancing selection, for far longer than they would persist if they were neutral. Here, the different alleles act as separate sub-populations, which diverge under drift and mutation, leading to increased variability in the population as a whole. Recombination between a neutral locus and the selected locus is analogous to migration, and opposes this divergence. The size of the increase in variability is proportional to $(N_e r)^{-1}$. We thus expect increased variation in the neighbourhood of a target of such balancing selection, falling off with the genetic map distance from the target; this is seen in the MHC complex in humans, and self-incompatibility loci in plants.

Relating models to observations

The different models make similar predictions which makes it difficult to ascribe an observation to just one process. Are there any observations that allow one to identify which process is operating? It turns out that the distribution of frequencies of neutral variants in a sample of sequences (the frequency spectrum) is model-dependent. A large distortion of the frequency spectrum towards rare variants is more likely with selective sweeps (Figure 1). And when there is some recombination, so that initially rare variants are not necessarily swept to complete fixation, a transient signature of a sweep is provided by the occurrence of an excess of 'derived' variants at a high frequency within a sample. (One can infer if a variant is derived or ancestral from the sequence of a closely related species.) Other features of genetic variation, such as patterns of linkage disequilibrium, may also be useful in testing alternative models. Perhaps the best way to quantify the relative importance of these processes is to get solid estimates of the rate at which deleterious mutations occur, and of the distribution of their effects on fitness. Although this is a simple question to ask, it is hard to answer. But given such information, and using the increasing amount of information on DNA sequence variation and evolution, one can perhaps try to answer an even harder question: what is the rate at which advantageous mutations occur, and what are their effects on fitness?

Key references

Kimura M: *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press; 1983.

Charlesworth B, Charlesworth D: The degeneration of Y chromosomes. *Phil Trans R Soc Lond B* 2000, 355:1563-1572.

Charlesworth B, Guttman DS: Reductions in genetic variation in ***Drosophila*** and ***E. coli*** caused by selection at linked sites. *J Genet* 1996, 75:49-61.

Felsenstein J: The evolutionary advantage of recombination. *Genetics* 1974, 78:737-756.

Kreitman M: Methods to detect selection in populations with applications to the human. *Annu Rev Genomics Hum Genet* 2000, 1:539-559.

Lynch M, Blanchard JL: Deleterious mutation accumulation in organelle genomes. *Genetica* 1998, 102/103:29-39.

Address: Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.