The evolution of genomic anatomy

Laurence D. Hurst

Just as Darwin applied his theory of natural selection to understand the details of natural history, so others have applied the idea to almost every aspect of biology from morphology to medicine. Can we similarly comprehend the rapidly accumulating details of the natural history of genomes or is selection not that strong a force? Recent case histories indicate that selection can affect everything from minuscule details, such as codon usage, to broader scale patterns, such as the linkage arrangement of genes, their chromosomal position and copy number. Although we should not assume that the structure of genomes is exclusively the result of history and chance, few generalities are presently possible because evidence is largely restricted to case-by-case analyses.

Laurence Hurst is at the Dept of Biology and Biochemistry, University of Bath, UK BA2 7AY (l.d.hurst@bath.ac.uk).

Genomic details, such as where on a chromosome genes reside, next to which others, how many copies of the gene exist and on which chromosomes, are accumulating at an unprecedented rate. With this growth in knowledge, it seems appropriate to ask whether there is any prospect of making evolutionary sense of these details of genomic anatomy. By evolutionary sense, I mean whether genomic anatomy is the result of selection or whether it is a result of chance events.

One's null expectation might be that genomic anatomy is the result of chance. One might suppose: (1) that most features of genomic anatomy have no obvious adaptive utility, or (2) that even if variants in genomic anatomy did have selection acting on them, that selection is too weak to be effective, or (3) that even if selection is strong enough, variation in features such as chromosomal location is likely to be limiting. The first supposition is largely correct. Unlike the eye, for example, features of genomic anatomy rarely have much evident and complex 'design'. However, not all that is the result of selection has apparent design, and so the first supposition is equivalent to noting that it is necessary to use statistical tests to identify deviation from null expectations. The second and third objections concern the limits of natural selection.

Drift and the limits of selection

Alleles will change frequency either because of chance events (drift) or because of selection (acting either on the variant or on a linked variant). Both forces can in principle give rise to patterns observed in genomes, such as the linkage of gene duplications. Selection can give rise to nonrandom distributions because it is a deterministically biased process: if selection favours duplicated genes to be linked, then duplicates typically will end up linked (here a duplication is treated as an allele, the same being possible for any form of genomic variation). Drift does not have such a bias, but can result in biased distributions if the process of mutation is biased. If duplication of a gene is effectively neutral and if linked duplicates are often generated, then by chance one would expect to find an excess of linked duplicate genes. The null hypothesis, against which a selectionist model must compete, is then one in which the trait is selectively neutral and patterns are the result of biased mutational processes. Understanding the form and magnitude of such biases is, therefore, critical to progress in evolutionary genomics.

Additionally, if the mutation rate (μ) is high ($\mu > 1/N_e$, where N_e is the effective population size), such as for some microsatellites, mutation alone can cause significant changes in allele frequency that are independent of the action of either drift or selection. For the most part, however, mutation alone is probably not powerful enough to have any major effect.

When should drift be more important than selection? It is unlikely that many changes are precisely neutral, because even tiny differences in genomic anatomy could have effects on fitness. Importantly, however, when the selective difference between two alleles is very small, random fluctuations in frequency of the two alternative types will tend to swamp the effect of selection – that is, the mutations will be effectively neutral. How small must the selective effects be to be drowned out by noise and what sort of genomic phenotypes are likely to be of such low selective value? At least three important parameters have been identified for the former:

Recombination rate

Genomic regions of low recombination, and nonrecombining populations, are thought to suffer from at least four disadvantages[1]:

- They cannot reverse the accumulation of deleterious mutations caused by drift and mutation.
- Linked deleterious alleles can interfere with the spread of advantageous ones.
- If a favourable mutation goes to fixation other favourable variants on the competing chromosome segments are lost.
- Their genetic loads can be relatively high when deleterious mutations have synergistic effects (i.e. every additional deleterious mutation has proportionally greater impact on fitness).

Recombination, in contrast, allows small segments of the genome to be individually viewed by selection with minimal interference from surrounding alleles.

In experimental lines, a nonrecombining chromosomal segment rapidly accumulates deleterious mutations[2]. The degeneration of Y chromosomes is expected for comparable reasons[2]. Similarly, reduced recombination has been shown to inhibit the accumulation of advantageous mutations[3,4] (but see Ref. 5), and recombination rate is positively correlated with the rate of adaptation[6–8].

Population size

The dynamics of a deleterious allele are effectively those of a perfectly neutral allele if the reduction in fitness associated with the allele is 'much less' than $1/2N_e$. More generally, as populations become small, advantageous mutations are more likely to be lost and slightly deleterious ones are more likely to become fixed by chance[9]. Features, such as population structure, that affect N_e will hence tend to affect the efficiency of both positive and stabilizing selection. Similarly, the effect of both hitchhiking and background selection (Box 1) can be considered equivalently, as reductions in N_e.

Controlling for the time of divergence, intracellular bacteria (with very small N_e) have more highly diverged proteins than their free-living relatives (with relatively large population sizes)[10]. Given additional evidence indicating that these substitutions are probably deleterious, this is consistent with an increasing substitution rate of slightly deleterious mutations as populations become small, although variance in the levels of recombination might also contribute.

The number of excess progeny

Selection might be more efficient in species that produce many progeny. In a stable population, sexual parents on average produce more than two progeny but

 0169-5347/99/$ – see front matter PII: S0169-5347(98)01501-8

leave only two that survive to reproduce, the others being the 'opportunity for selection'. As each population has only so many excess individuals, there are only so many polymorphic sites in a genome that can be undergoing selection simultaneously[11]. There is no experimental verification of this logic.

It is likely, therefore, that large sexual populations will be nearer adaptive optima than small asexual ones. Similarly, segments of the genome with low recombination rates are less likely to be near an evolutionary optima (although the low recombination rate might be an adaptation). This is a useful generalization for evolutionary genomics, because covariance between the bias in a trait and local recombination rate and/or N_e can be used as correlatory evidence consistent with the activity of selection.

Small-scale genome structure and the power of selection

These results, although helpful, shed no light on the problem of whether small changes in genomic anatomy will be noticed by selection, even in large sexual populations. In the absence of any theory for the sizes of selection coefficients, this is an empirical issue.

Consider, then, the problem of a minuscule change, such as a mutation changing one codon to another coding for the same amino acid. Within the set of tRNAs specifying the same amino acid, some are more abundant than others. If selection were to favour rapid translation of a gene, then codons for the relatively abundant tRNAs would be selectively advantageous (see also Ref. 12). All that we can say about the size of the selection coefficient is that it is probably extremely small[13]. But is it so small that drift always overpowers it?

In *Drosophila*, there is good evidence for biased codon usage[14], which mostly cannot be accounted for under a null hypothesis of mutational bias allied with drift: there is little evidence for regional mutational bias, mutation bias is probably in the opposite direction to that observed, and not all amino acids have the preference for the same nucleotide at the wobble position[14] (i.e. the third 'variable' position of a codon). In contrast, the alternative hypothesis of selection finds good support: highly expressed genes tend to be highly biased, preferred codons in highly biased genes optimally bind the most abundant relevant tRNAs (Ref. 15) and amino acids whose major isoaccepting tRNAs change concentration during development do not show as strong a bias as those with developmentally unchanged tRNA pools[15]. Therefore, it is likely that much codon-usage bias in *Drosophila*, as in *Escherichia coli*, *Saccharomyces* and *Caenorhabditis*[16], is in part the result of selection[14,17]. A lack of codon bias could also be due to selection. In *E. coli*, synonymous codon usage is less biased at the start of genes. The rate of synonymous substitution (Box 1) is substantially reduced, indicating stabilizing selection, possibly acting to avoid mRNA secondary structure[18].

A role for selection is consistent with evidence suggesting a link between the degree of bias and population size. First, mutational bias[19] appears to explain most mammalian codon-usage bias, which is consistent with mammals having small populations. Second, selection on codon usage is less efficient in *D. melanogaster* than *D. simulans*, the former being thought to have a much smaller population size than the latter[17]. However, this evidence is limited and a multispecies comparative analysis is necessary. The view that recombination rate is important is supported by correlative evidence from *Drosophila*[20].

If we suppose selection to have shaped codon usage, it still remains unclear precisely what selection is acting on. It is possible that a trait such as codon usage (or GC content, Box 1) of the gene as a whole might be under selection, while any given third-site mutation might be effectively neutral[21]. Given that selection can affect codon usage, it is not surprising that nonsynonymous mutations (Box 1) in genes are typically subject to stabilizing selection[22]. It is unclear whether the few nonsynonymous substitutions in genes are mostly slightly deleterious mutations that have evaded stabilizing selection or advantageous ones favoured by selection.

Box 1. Glossary

Background selection: the loss of deleterious mutations by selection can drag linked alleles out of the population. This background selection will reduce the effective population size of the chromosomal region with the disadvantageous mutation and also result in lower levels of neutral polymorphism.
Centromere: the region of the chromosome where mitotic spindles attach. This region is typically associated with low recombination rates. Genes near the centromere are said to be centromeric.
GC content: species differ in the usage of the four bases. This is often measured in terms of the percentage of bases that are G (guanine) or C (cytosine) in noncoding regions. Additionally, in mammals, it is known that the genome is divided into fairly discrete regions (**isochores**) of more or less consistent GC content. In an isochore, the GC content of the third sites in exons (i.e. those that are mostly synonymous) strongly correlates with that in introns and with that in spacer DNA.
Hitchhiking: like background selection of deleterious mutations, an allele in linkage disequilibrium with an advantageous allele can spread as a result of being dragged along with the advantageous allele. This is known as hitchhiking and can also reduce the effective population size.
Linkage disequilibrium: this is the nonrandom association between alleles at different loci. Imagine two loci each with two alleles (A/a and B/b). If there is linkage equilibrium, then the probability of finding a haplotype that is AB will be equal to the frequency of A times the frequency of B (i.e. the distribution is random). If this is not so, then there is linkage disequilibrium. This might arise as a consequence of linkage. Imagine a population polymorphic for A/a but monomorphic for B. If B mutates to b, this mutant allele must either initially be associated with either A or a. If the two loci are linked, then it could take many generations before this initial association is randomized by recombination. Linkage is a sufficient but not a necessary prerequisite for linkage disequilibrium.
Nonsynonymous mutation: a mutation in an exon that changes the amino acid composition of the protein is said to be nonsynonymous. If a nonsynonymous mutation reaches fixation then it is said to be a **nonsynonymous substitution**.
Synonymous mutation: a mutation in an exon that does not change the amino acid composition of the protein is said to be silent or synonymous. If a synonymous mutation reaches fixation then it is said to be a **synonymous substitution**.
Telomere: the end of the chromosome. In mammals, this is associated with high recombination rates but in *Drosophila* it is associated with low rates. Genes near the telomere are said to be telomeric.

Larger-scale genome structure

The patterns of codon usage indicate that selection is more powerful than often supposed. In large sexual species, it is more likely that limited variation might prevent genomes from achieving a position of optimal organization. Although there is potentially abundant variation in codon usage, intraspecific variation in, for example, gene location, is often limited. But is the amount of variation limiting and can selection affect which chromosome a gene is on, its copy number, linkage and the chromosomal position? In this context, the expectation that selection will be a minor player is most easily defensible.

Which chromosome should a gene be on?

Genes with sex differences in fitness effects are favoured to be on chromosomes with differences in their transmissibility through the sexes[23]. This might be either because selection will favour such genes moving to a 'more appropriate' chromosome, or because selection acts as a sieve for spontaneous mutations. Sexually antagonistic genes are a special subclass that are beneficial to one sex but disadvantageous to the other.

Mutations that occur in maternally transmitted genomes, such as mitochondrial ones, that are very deleterious in females but advantageous in males will not typically be found at appreciable

frequencies, but the converse (i.e. mitochondrial mutations advantageous to females but not to males) will[24]. Evidence for such a filter exists: in plants, spontaneous mutations inducing male sterility are more common in nuclear genes than cytoplasmic ones, whereas those found at appreciable frequencies in natural populations are more commonly cytoplasmic than nuclear (A. Burt, pers. commun.).

Rice has provided experimental evidence for the rapid accumulation, in appropriate genomic locations, of genes that are deleterious to the sex through which they are not transmitted[25]. Using artificial selection, he forced part of the *Drosophila* genome to be transmitted exclusively from mother to daughter. Over the 29 generations of this treatment, the chromosomal segment had an increasingly negative effect on male fitness. Similarly, a chromosome transmitted exclusively down the male line should tend to accumulate mutations beneficial to males. In experiments in which 99% of a haploid genome of *D. melanogaster* are forced to segregate like a male-limited Y chromosome (so eliminating potential counter selection in females), the synthetic Y chromosomes rapidly accumulated genetic variation that increased male fitness and decreased female fitness[26].

Such sexual antagonism can explain why some genes are Y-linked, although for the most part the data is anecdotal. Winge[27], for example, found 18 major segregating loci for male-specific display traits of guppies, of which 17 were within two centimorgans from the male-determining region. Some, however, recombine with the X chromosome and have suppressed expression in females. The reason for linkage of these to the sex-determining region is unclear. Notably, in humans, mice and fruitflies, genes involved in spermatogenesis are found on the Y chromosome. Similarly, the human Y chromosome contains genes necessary for postnatal growth[28], which is probably a sexually antagonistic trait. In humans and other primates, the Y chromosome also contains gene(s) responsible for tooth growth[28]. In primates, tooth size is probably a sexually antagonistic trait; the baring of teeth being a display trait.

There is no good evidence for sexually antagonistic genes on any X chromosome, although there is an excess of sex- and reproduction-related ones on the human X (Ref. 51). More generally, although we now have good evidence that mutations with different fitness effects in the two sexes are common, we have no good indication of the proportion of genes on each chromosomal type that are optimally suited to being where they are.

The evolution of copy number

One might suppose that a duplication, or a genome-wide increase in ploidy, would be favoured by natural selection because it can mask the effects of deleterious mutations. Such explanations, however, fail to explain the variance in ploidy[29], and models show that a duplication can invade only if it provides a direct advantage to the organism[30]. The most obvious advantage of having two copies of a gene is that a higher dosage of the gene product is easily attainable. For example, the multiple copy nature of rDNA repeats, even in organisms such as *Neurospora* with no other multicopy genes, is most likely a consequence of selection for high dosage.

Direct support for the possibility that copy number can be under selection for dosage comes from segregation patterns of the X-linked multicopy gene *Stellate* (*Ste*) in *D. melanogaster*. *Stellate* has been hypothesized to be an X-chromosome meiotic driver, which has evolved in a dosage sensitive 'arms race' with its Y-linked suppressor. Meiotic-drive genes inhibit gametes from the same individual not containing the drive 'gene'. Analysis[31] of segregation data from males with no Y-linked suppressor but with low *Stellate* copy number revealed meiotic drive, the strength of which is proportional to *Stellate* copy number[31]. This proportionality is direct evidence that increased dosage can be favoured by selection.

Generally, if a gene family (including satellite repeats) were under selection for copy number, covariance of copy number with some fitness components (not necessarily segregation rate) would be expected[32]. However, fitnesses that we can experimentally assess are vastly less subtle than those that selection will affect in large sexual populations[21]. Hence, an absence of an effect of dosage need not be proof of an absence of selection. Indeed, murine genes whose knockout have no phenotype still have normal (low) ratios of nonsynonymous to synonymous rates (L. Hurst, unpublished), indicating stabilizing selection.

The evolution of linkage

There are numerous indications that linkage need not be random. The finding of functionally interacting but genetically distinct genes is especially suggestive. Linkage of *LMP7*, *LMP2* and *Tap*, class II genes in the mammalian major histocompatibility complex (MHC), is one example[33]. Perhaps the best described examples are in *Caenorhabditis*[34]. Two different enzymes are needed for trimerizing collagen and these are encoded within one operon, as are *des-2* and *deg-3* (the two subunits of the acetylcholine receptor) and *lin-15B* and *lin-15A* (unrelated proteins that collaborate in vulva development). The frequently observed clustering of eukaryotic genes coding for metabolically related enzymes is also suggestive[35].

Further evidence is claimed to come from phylogenetic analysis[36] of nine gene families that have members on both human chromosomes 6 and 9. Although block duplication is a possible explanation for the existence of these two clusters, phylogenetic evidence suggests this should be rejected as an overall explanation, because some genes appear to have duplicated at different times over at least 1.6 billion years. However, a model evoking block duplication and gene loss can also account for the pattern, so a special selective explanation is not necessarily warranted. Better evidence comes from *Drosophila* in which maltase clusters do appear to be a consequence of independent evolutions of linkage[37].

These examples, although possibly suggestive of a role for selection, are of uncertain importance in the absence of null models for the regularity of such coincidences. Nonetheless, we can still ask why selection might favour linkage. There exist two categories of explanation: those that require physical chromosomal proximity for mechanistic reasons and those that point to advantages of maintaining linkage disequilibrium (Box 1). Coordinated gene expression explains the evolution of numerous prokaryotic operons. The relationship between gene order and expression order seen in Hox clusters and globin clusters is indicative of similar effects.

The problem of the maintenance of linkage disequilibrium is the same as the problem of the evolution of sex. Sex and recombination might be beneficial because they break down linkage disequilibrium and in so doing permit the spread of advantageous alleles or enhance the removal of deleterious ones[1]. For example, the mutational deterministic model[38] for the evolution of sex proposes that if mutations are deleterious and interact synergistically, selection will favour recombination between them[38]. These models generally point to the disadvantages of having genes clustered and to the build up of linkage disequilibrium. Consequently, they predict that genes should generally be dispersed throughout the genome. This is hard to test.

The opposite conditions provide the conditions favouring linkage. If extra deleterious mutations have proportionally declining effects on fitness, then clustering and an absence of recombination can be favoured. If these genes belong to a multigene family, then both clustering and gene conversion are favoured[39]. For

advantageous alleles, the case is slightly more complex. Sex permits two mutually beneficial alleles to be put more easily into the same genome because they are most likely to occur in different individuals. However, once they are in the same genome, selection should then favour their linkage[40].

The best data for selectively driven maintenance of linkage disequilibrium comes from meiotic-drive genes and supergene complexes. The two best described examples of meiotic drive are *Segregation Distorter* (*SD*) in *D. melanogaster* and *t-complex* in mice[41]. Drive 'genes' usually involve two types of locus[41]. Putatively, one codes for a toxin that is shared by all sperm, the other is the antidote whose effects are restricted to the sperm containing it. As a consequence the wild-type sperm die, whereas those carrying the antidote survive. Invasion of drive is only possible if there is tight linkage of the two loci[41]. Similarly, genes enhancing the effect of the driver must also be linked to the drive locus for them to invade, whereas suppressors must be unlinked. No exception to these rules of linkage is known[41].

As for meiotic drive genes, supergene complexes are closely linked clusters of coadapted genes. Examples include the numerous genes controlling heterostyly in *Primula*, those for Batesian mimicry in butterflies, colour genes in grouse-locusts (*Acrydium arenosum*), and for snail shell patterns[35]. The best described is the mating-type locus of *Chlamydomonas*, in which genes controlling organelle inheritance and mating-type are held in linkage by inversions, deletions and rearrangements[42].

Although these cases illustrate the ability of selection to affect linkage patterns, it remains unclear whether the clustering of interacting genes resulting from selection is the exception or the rule. Data from bacteria suggest clustering to be the exception: in four species only 16 clusters are conserved, possibly because these genes require close proximity for coordinated gene expression[43]. Conservation of linkage groupings above null expectations can generally be considered a means to examine a role of selection. Defining null expectations will be difficult because different regions can have different rates of rearrangement and translocation. Again, we need to know more about mutational biases.

The evolution of chromosomal location

Chromosomal locations differ in at least two important respects. First, different regions can have different rates of transcription. These differences probably explain the positioning on rDNA genes near to the origins of transcription and/or replication in both eubacteria and on the vertebrate mitochondrial genome. Second, different regions can also have different recombination rates. Typically, centromeres have low recombination rates. In mammals (but not in Drosophilids), telomeres have high rates (Box 1). When selection favours linkage, we might also expect selection for genes to be close to the centromere. When selection favours an absence of linkage, it might favour genes to be telomeric (at least in mammals).

Although meiotic drive genes are often in inversions, they are also usually in close proximity to the centromere[41]. However, we must ask to what extent selection for low recombination rates will result in centromeric localization and whether centromeric location need be the result of selection for low recombination rates.

The first question can be answered by analysis of the location of supergene complexes. Unfortunately, in only two cases is the chromosomal location known: the heterostyly supergene of *Primula* is close to the centromere[44] but the mating-type locus of *Chlamydomonas* is not[42]. Similarly, the distorting inversion on chromosome 1 in mice is telomeric[45]. This location might be adaptive. Distortion might be occurring in meiosis II in oogenesis. Were this so, recombination between the centromere and the inversion is necessary to ensure that the oocyte is heterozygous[45]. Centromeric location might then be sufficient to reduce recombination rates, but it is not necessary.

At least in a few special circumstances, selection can favour centromeric linkage for reasons other than reduction of recombination. Meiotic drive genes in the pseudohomothallic fungi, *Podospora anserina*[46] and the mating-type locus in *Neurospora tetrasperma*[47] are centromeric and ensure appropriate segregation patterns. To ensure self-fertility, meiosis in these species is adapted to allow delivery of two nuclei of opposite mating type into the same ascospore[47]. In *P. anserina*, this involves a single reciprocal crossover in the mating-type centromere interval. Selection in this case favours a noncentromeric location. Unlike mating-type loci, for meiotic drive to occur, both nuclei in an ascospore must contain the drive genes. Centromeric location is necessary to ensure this (it is therefore surprising that some *P. anserina* distorters are not centromeric[46]). In contrast, *N. tetrasperma* has a form of meiosis that requires both mating-type alleles to be centromeric, which they are[46].

It is reasonable to conclude that the chromosomal location of certain genes is the product of selection. Some of these are cases in which selection favours low recombination rates. If, however, selection generally favours dissolution of linkage disequilibrium, we might not be able to predict the location of a gene with any accuracy, but we might be able to make predictions concerning the covariance of gene density and recombination rates and hence covariance of gene density and chromosomal position. Selection, either to allow the spread of advantageous mutations or to eliminate deleterious ones, should result in recombination being common in regions of high gene density, because it is here that the interference between linked mutations is potentially most intense. This pattern is observed: regions of the human genome with high recombination rates typically have high GC content[48] and the GC-richest component of the human genome is at least 17 times as gene-rich as the GC-poor regions[49]. The GC content is also highest towards telomeres, suggesting a correlate of gene density, recombination rate and GC content with proximity to the telomere. Although these details are suggestive of a selective effect, the mutational-bias alternative has not been eliminated. For example, if recombination induces deletions then the compact nature of GC-rich isochores might be a simple consequence of recombination.

Rather than asking whether selection to favour linkage disequilibrium is more common than selection to dissolve it, it might be more helpful to consider the sexual genome as varying from parts that are more asexual (i.e. regions of low recombination) to others that are more sexual (i.e. regions of high recombination). Selection might favour different classes of gene to be found in one rather than the other[39]. An approach from metabolic control theory[50] seems the best prospect for making predictions about the sort of gene in each class.

Conclusions, problems and prospects

The evidence discussed here has indicated the following: first, that selection, especially in large sexual populations, is strong enough to affect even minuscule details of genomic anatomy; second, that variation in features such as chromosomal location and linkage need not be limiting; and third, that features with no obvious adaptive use can nonetheless be the result of selection. It is unknown whether variation is usually limiting and whether the selective coefficients associated with variations in genomic anatomy are usually adequately large. However, there certainly need be no reason not to hypothesize

about the influence of selection on genomic structures, whether it concern the problems described here or one of the numerous related ones, such as the variation between organisms in the number of genes, the variation in genome size, the size and number of introns, the distribution of transposable elements, or chromosome number.

One limitation for the field at present is the necessity to work on a case by case basis. However, some problems, such as variations in GC content, are more amenable to systematic analyses and these might be especially instructive as regards the limits of selection. Like a microscope with ever higher magnification, we can ask whether selection affects variation in GC content between genomes, between isochores, between genes and within genes. That GC content at synonymous sites is higher than GC content of the flanking spacer[21] is suggestive of selection, even on this small scale.

With the prospect of sample sizes ordinarily unimaginable to evolutionionists, evolutionary genomics has the potential to provide evolutionary biology with its most rigorous testing ground. Unfortunately, to make sense of much of the data, we will need more than sequenced genomes. Most especially, for null models, we will need details of genomic processes to understand mutational biases, as well as knowledge of intragenomic variation in the recombination rate and the mutation rate, etc. For the appropriate comparative analyses, we will need good estimates of N_e from very many species, as well as good phylogenies. To understand linkage better, we will need to understand both epistasis and gene regulation. Assuming that these difficulties can be addressed, evolutionary genomics will show us just where selection has it limits.

Acknowledgements

I am grateful to A. Kondrashov, W. Rice, L. Partridge, P. Harvey, N. Smith, A. Eyre-Walker, L. Kruuk, A. Pomiankowski, G. Hurst, G. McVean and four anonymous referees for comments. I am grateful to S. Jones, J. Richards, J. Mallet, W. Rice, A. Burt, R. Hoekstra and U. Goodenough for help and access to unpublished data.

References

1 Hurst, L.D. and Peck, J.R. (1996) **Recent advances in understanding the evolution and maintenance of sex,** *Trends Ecol. Evol.* 11, 46–52
2 Rice, W.R. (1994) **Degeneration of a non-recombining chromosome,** *Science* 263, 230–232
3 McPhee, C.P. and Robertson, A. (1970) **The effect of suppressing crossing-over on the response to selection in *Drosophila melanogaster*,** *Genet. Res.* 16, 1–16
4 Markow, T.A. (1975) **A genetic analysis of phototactic behavior in *Drosophila melanogaster*,** *Genetics* 79, 527–534
5 Thompson, V. (1977) **Recombination and response to selection in *Drosophila melanogaster*,** *Genetics* 85, 125–140
6 Burt, A. and Bell, G. (1987) **Mammalian chiasma frequencies as a test of two theories of recombination,** *Nature* 326, 803–805
7 Korol, A.B. and Iliadi, K.G. (1994) **Recombination increase resulting from directional selection for geotaxis in *Drosophila*,** *Heredity* 72, 64–68
8 Flexon, P.B. and Rodell, C.F. (1982) **Genetic-recombination and directional selection for DDT resistance in *Drosophila melanogaster*,** *Nature* 298, 672–674
9 Ohta, T. (1992) **The nearly neutral theory of molecular evolution,** *Annu. Rev. Ecol. Syst.* 23, 263–286
10 Moran, N.A. (1996) **Accelerated evolution and Muller's rachet in endosymbiotic bacteria,** *Proc. Natl. Acad. Sci. U. S. A.* 93, 2873–2878
11 Haldane, J.B.S. (1957) **The cost of natural selection,** *J. Genet.* 55, 511–524
12 Akashi, H. (1994) **Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy,** *Genetics* 136, 927–935
13 Li, W-H. (1987) **Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons,** *J. Mol. Evol.* 24, 337–345
14 Powell, J.R. and Moriyama, E.N. (1997) **Evolution of codon usage bias in *Drosophila*,** *Proc. Natl. Acad. Sci. U. S. A.* 94, 7784–7790
15 Moriyama, E.N. and Powell, J.R. (1997) **Codon usage bias and tRNA abundance in *Drosophila*,** *J. Mol. Evol.* 45, 514–523
16 Sharp, P.M. *et al.* (1995) **DNA-sequence evolution: the sounds of silence,** *Philos. Trans. R. Soc. London Ser. B* 349, 241–247
17 Akashi, H. (1995) **Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA,** *Genetics* 139, 1067–1076
18 Eyre-Walker, A. and Bulmer, M. (1993) **Reduced synonymous substitution rate at the start of enterobacterial genes,** *Nucleic Acids Res.* 21, 4599–4603
19 Eyre-Walker, A. (1991) **An analysis of codon usage in mammals: selection or mutation bias?** *J. Mol. Evol.* 33, 442–449
20 Kliman, R.M. and Hey, J. (1993) **Reduced natural-selection associated with low recombination in *Drosophila melanogaster*,** *Mol. Biol. Evol.* 10, 1239–1258
21 Gillespie, J.H. (1991) *The Causes of Molecular Evolution*, Oxford University Press
22 Li, W-H. (1997) *Molecular Evolution*, Sinauer
23 Rice, W.R. (1984) **Sex-chromosomes and the evolution of sexual dimorphism,** *Evolution* 38, 735–742
24 Frank, S.A. and Hurst, L.D. (1996) **Mitochondria and male disease,** *Nature* 383, 224
25 Rice, W.R. (1992) **Sexually antagonistic genes: experimental evidence,** *Science* 256, 1436–1439
26 Rice, W.R. (1998) **Male fitness increases when females are eliminated from gene pool: implications for the Y chromosome,** *Proc. Natl. Acad. Sci. U. S. A.* 95, 6217–6221
27 Winge, O. (1927) **The location of eighteen genes in *Lebistes reticulata*,** *J. Genet.* 18, 1–43
28 Schafer, A. (1994) **Genes and phenotypes of the human Y chromosome,** *Reprod. Med. Rev.* 3, 77–95
29 Bell, G. (1997) **The evolution of the life cycle of brown seaweeds,** *Biol. J. Linn. Soc.* 60, 21–38
30 Clark, A.G. (1994) **Invasion and maintenance of a gene duplication,** *Proc. Natl. Acad. Sci. U. S. A.* 91, 2950–2954
31 Hurst, L.D. (1996) **Further evidence consistent with *Stellate*'s involvement in meiotic drive,** *Genetics* 142, 641–643
32 Wu, C.I., True, J.R. and Johnson, N. (1989) **Fitness reduction associated with the deletion of a satellite DNA array,** *Nature* 341, 248–251
33 Hughes, A.L. and Yeager, M. (1997) **Molecular evolution of the vertebrate immune system,** *BioEssays* 19, 777–786
34 Blumenthal, T. (1998) **Gene clusters and polycistronic transcription in eukaryotes,** *BioEssays* 20, 480–487
35 Korol, A.B., Preigel, I.A. and Preigel, S.I. (1994) *Recombination Variability and Evolution*, Chapman & Hall
36 Hughes, A.L. (1998) **Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1,** *Mol. Biol. Evol.* 15, 854–870
37 Vieira, C.P., Vieira, J. and Hartl, D.L. (1997) **The evolution of small gene clusters: evidence for an independent origin of the maltase gene cluster in *Drosophila virilis* and *Drosophila melanogaster*,** *Mol. Biol. Evol.* 14, 985–993
38 Kondrashov, A. (1988) **Deleterious mutations and the evolution of sexual reproduction,** *Nature* 336, 435–440
39 Hurst, L.D. and Smith, N.G.C. (1998) **The evolution of concerted evolution,** *Proc. R. Soc. London Ser. B* 265, 121–127
40 Bodmer, W.F. and Parsons, P.A. (1962) **Linkage and recombination in evolution,** *Adv. Genet.* 11, 1–100
41 Lyttle, T.W. (1991) **Segregation distorters,** *Annu. Rev. Genet.* 25, 511–557
42 Ferris, P.J. and Goodenough, U.W. (1994) **The mating-type locus of *Chlamydomonas reinhardtii* contains highly rearranged DNA sequences,** *Cell* 76, 1135–1145
43 Siefert, J.L. *et al.* (1997) **Conserved gene clusters in bacterial genomes provide further support for the primacy of RNA,** *J. Mol. Evol.* 45, 467–472
44 Lewis, D. and Jones, D.A. (1993) **The genetics of heterostyly,** in *Evolution and Function of Heterostyly* (Barrett, S.C.H., ed.), pp. 129–150, Springer-Verlag
45 Agulnik, S.I., Sabantsev, I.D. and Ruvinsky, A.O. (1993) **Effect of sperm genotype on chromatid segregation in female mice heterozygous for aberrant chromosome 1,** *Genet. Res.* 61, 97–100
46 Raju, N.B. (1996) **Meiotic drive in fungi: Chromosomal elements that cause fratricide and distort genetic ratios,** *J. Genet.* 75, 287–296
47 Raju, N.B. and Perkins, D.D. (1994) **Diverse programs of ascus development in pseudohomothallic species of *Neurospora*, *Gelasinospora*, and *Podospora*,** *Dev. Genet.* 15, 104–118
48 Eyre-Walker, A. (1993) **Recombination and mammalian genome evolution,** *Proc. R. Soc. London Ser. B* 252, 237–243
49 Zoubak, S., Clay, O. and Bernardi, G. (1996) **The gene distribution of the human genome,** *Gene* 174, 95–102
50 Szathmáry, E. (1993) **Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer,** *Genetics* 133, 127–132

Reference added in proof

51 Saifi, G.M. and Chandra, H.S. **An apparent excess of sex-related and reproduction-related genes on the human X-chromosome,** *Proc. R. Soc. London Ser. B* (in press)