

REVIEW

How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination

R Frankham^{1,2}

Levels of genetic diversity in finite populations are crucial in conservation and evolutionary biology. Genetic diversity is required for populations to evolve and its loss is related to inbreeding in random mating populations, and thus to reduced population fitness and increased extinction risk. Neutral theory is widely used to predict levels of genetic diversity. I review levels of genetic diversity in finite populations in relation to predictions of neutral theory. Positive associations between genetic diversity and population size, as predicted by neutral theory, are observed for microsatellites, allozymes, quantitative genetic variation and usually for mitochondrial DNA (mtDNA). However, there are frequently significant deviations from neutral theory owing to indirect selection at linked loci caused by balancing selection, selective sweeps and background selection. Substantially lower genetic diversity than predicted under neutrality was found for chromosomes with low recombination rates and high linkage disequilibrium (compared with 'normally' recombining chromosomes within species and adjusted for different copy numbers and mutation rates), including W (median 100% lower) and Y (89% lower) chromosomes, dot fourth chromosomes in *Drosophila* (94% lower) and mtDNA (67% lower). Further, microsatellite genetic and allelic diversity were lost at 12 and 33% faster rates than expected in populations adapting to captivity, owing to widespread selective sweeps. Overall, neither neutral theory nor most versions of the genetic draft hypothesis are compatible with all empirical results.

Heredity (2012) **108**, 167–178; doi:10.1038/hdy.2011.66; published online 31 August 2011

Keywords: background selection; effective population size; genetic diversity; neutral theory; recombination; selective sweeps

INTRODUCTION

Loss of genetic diversity is a major concern in conservation and evolutionary biology, as genetic diversity is the raw material upon which natural selection acts to produce adaptive evolutionary change. Further, it is related to inbreeding and loss of reproductive fitness in random mating populations and ultimately to elevated extinction risks.

Neutral theory (Kimura, 1983) is widely used to predict changes in genetic diversity over time and to predict equilibrium levels for populations and species. However, Maynard Smith and Haigh (1974) and Gillespie (2000) proposed that linked selection (hitchhiking or genetic draft) may be a more important stochastic factor than genetic drift in natural populations. Thus, quantitative analyses of how well neutral theory and genetic draft predict genetic diversity across a range of circumstances are necessary.

I review recent insights into loss of genetic diversity in finite populations, concentrating on the magnitude of deviations from neutral predictions. First, I review data on correlations between genetic diversity and population size. Second, I evaluate the magnitude of deviations from neutral predictions for loss of genetic diversity in small populations over short durations (~50 generations). Third, I evaluate the effects on equilibrium genetic diversity of selection on linked loci for chromosomes with low versus relatively normal recombination. To allow quantitative comparisons for different chromosomes and genomes, data on the latter issue are converted to a common currency, the ratio of observed diversity to that predicted by

neutral theory (adjusted for copy number and mutation rate difference, as described below) where the neutral expectation is 1.

If there is widespread non-neutral behaviour of coding, non-coding and synonymous sites, as the results indicate, it has widespread implications, including the following:

- Prediction based on neutral theory in conservation and evolutionary biology, and animal and plant breeding may often be inaccurate.
- Estimates of effective population sizes may vary for different regions of genomes, especially those with different recombination rates (Charlesworth, 2009).
- Estimates of migration rates may vary for different regions of the genome.
- Phylogenies inferred from sequence data on non-recombining chromosomes (for example, mitochondrial (mtDNA) and chloroplast DNA (cpDNA)) may be distorted by selection.
- Molecular clock dating may be incorrect.
- Tests for selection based on the use of non-coding loci or synonymous substitutions may miss selected loci and provide inaccurate estimates of selection coefficients.

Predictions of neutral theory

For neutral loci, the proportion of initial genetic diversity (expected heterozygosity) retained after t generations (H_0/H_t) in a diploid

¹Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia and ²Australian Museum, Sydney, New South Wales, Australia
Correspondence: Professor R Frankham, Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia.
E-mail: richard.frankham@mq.edu.au

Received 19 December 2010; revised 21 June 2011; accepted 27 June 2011; published online 31 August 2011

population is predicted to be related to effective population size (N_e), generations (t) and the inbreeding coefficient at generation t (F_t) in random mating populations, as follows (see Wright, 1969; Falconer and Mackay, 1996):

$$H_t/H_0 = [1 - 1/(2N_e)]^t = 1 - F_t \quad (1)$$

In the long term, the predicted equilibrium genetic diversity (H_e) due to the balance between neutral mutations and random genetic drift for the infinite alleles model is (Crow and Kimura, 1970):

$$H_e = 4N_e u / [(4N_e u + 1)] \quad (2)$$

where u is the neutral mutation rate. The factor 4 in equation (2) is altered to 1 for Y and W chromosomal loci, chloroplast DNA (cpDNA) and mtDNA, and to 3 for loci on X and Z chromosomes in species with separate sexes (assuming random variation in offspring numbers, no separate sexes and transmission of single mtDNA and cpDNA genomes per female gamete).

Predicted impacts of selection at linked loci on genetic diversity

The equations above are based upon single locus models that ignore the impacts of selection at linked loci. However, there is now substantial theoretical and empirical evidence that linked selected loci often affect genetic diversity of nearby inherently neutral loci, owing to associative balancing selection (see Latter, 1998; Charlesworth, 2006), selective sweeps (hitchhiking) (see Maynard Smith and Haigh, 1974) and background (purifying) selection (see Charlesworth *et al.*, 1993).

Directional selection

When a new favourable mutation goes to fixation (positive selection) it will remove initial genetic diversity for all loci on a non-recombining chromosome within a closed population. This effect was first designated as periodic selection in asexual bacteria (Atwood *et al.*, 1951; Cohan, 2005) and later referred to in eukaryotes as hitchhiking (Maynard Smith and Haigh, 1974), selective sweeps (Berry *et al.*, 1991) or genetic draft (Gillespie, 2000). The impact of selective sweeps on long-term effective population size (N_l), and thus on genetic diversity (through equations (1) and (2)) in a non-recombining segment of chromosome, is predicted to be (Gillespie, 2000):

$$N_l \sim N_e / (1 + 2N_e \delta) \quad (3)$$

where N_e is the effective population size in the absence of selective sweeps and δ is the rate of selective sweeps. Selective sweeps also reduce genetic diversity in regions with recombination but to a lesser degree, as recombination breaks down linkage disequilibrium (Gillespie, 2000). There are many examples of reduced genetic diversity in regions flanking loci subject to directional selection (see Nurminsky, 2005). For example, reduced diversity occurs around the *tb1* locus involved in maize domestication (Clark *et al.*, 2004); for loci surrounding the *waxy* locus involved in domestication of japonica rice (Olsen *et al.*, 2006) and for loci flanking the lactase locus in human populations that adopted dairying and evolved adult lactose persistence (Burger *et al.*, 2007). Sweeps originating from new favourable mutations are referred to as hard sweeps, whereas those due to pre-existing polymorphisms are referred to as soft sweeps and typically have lesser impacts (Hermisson and Pennings, 2005).

Natural selection against new deleterious mutations also reduces genetic diversity at linked loci (Charlesworth *et al.*, 1993; Charlesworth, 1996). The proportionate reduction in nucleotide diversity (π/π_0) at a neutral locus owing to background selection depends upon the deleterious mutation rates (ud), the selection coefficients against

heterozygotes for deleterious alleles (s) and the recombination rates (c), and the effects are summed across all linked loci that mutate to deleterious alleles, as follows (Nordborg *et al.*, 1996):

$$\pi/\pi_0 \sim \exp - \sum_i (ud_i/s_i) / \{1 + c_i(1 - s_i)/s_i\}^2 \quad (4)$$

This is a weak form of selection that reduces N_e . Over many generations it reduces genetic diversity at linked loci, especially in regions of low recombination.

In what follows, I typically do not distinguish the impacts of selective sweeps from those of background selection, as both reduce genetic diversity compared with neutral expectations and both usually operate simultaneously. In addition, selection at one locus impedes selection response at linked loci, especially in regions of low recombination (Hill and Robertson, 1966), and interacts with other forms of selection, especially background selection (Charlesworth *et al.*, 2009).

Balancing selection

Balancing selection in a region will usually lead to nearby neutral loci retaining more genetic diversity than expected with neutrality (see Latter, 1998; Charlesworth, 2006). Associative balancing selection can be due to linked loci showing balancing selection or to the short-term effects of deleterious alleles at several loci in linkage disequilibrium (Latter, 1998) (that contrast with the long-term effects involved in background selection). For example, loci near self-incompatibility loci (that are subject to frequency-dependent selection) in *Arabidopsis lyrata*, but not involved in self-incompatibility themselves, have elevated levels of genetic diversity compared with other loci in the genome (Kamau *et al.*, 2007; Ruggiero *et al.*, 2008). Similarly, the major histocompatibility complex in vertebrates and the complementary sex-determining locus in Hymenoptera are both subject to balancing selection, and flanking sequences that are not themselves subject to selection also show elevated levels of genetic diversity (O'Uigin *et al.*, 2000; Hasselmann and Beye, 2006).

CORRELATIONS OF GENETIC DIVERSITY WITH POPULATION SIZE AND FITNESS

Relationships of genetic diversity and effective population size

Equations (1) and (2) predict that Hardy-Weinberg expected genetic diversity across populations will be positively correlated with effective population sizes, provided mutation rates are similar. This prediction has been verified experimentally in pedigreed *Drosophila* populations for allozymes (correlation $r_{H_e - \log N_e} = 0.59$; Montgomery *et al.*, 2000) and microsatellites ($r_{H_e - \log N_e} = 0.91$; Montgomery *et al.*, 2010). Further, Palstra and Ruzzante (2008) reported a correlation of 0.73 between H_e for microsatellites and $\log N_e$ across 26 closed populations of diverse species.

Effective population sizes and N_e/N ratios

For most species N_e is unknown, so inferences are often based upon census sizes (N) and N_e/N ratios from other species. Based on a meta-analysis, ratios in unmanaged wild populations with all relevant variables included averaged 0.11 (Frankham, 1995). No consistent significant differences were detected in ratios across a broad range of major taxa, indicating that positive correlations between genetic diversity and census population sizes are expected. Palstra and Ruzzante (2008) reviewed temporal estimates of N_e/N and reported a median value of 0.14, based on 64 estimates from diverse animal and plant taxa. However, species with high fecundity have significantly reduced N_e/N ratios on the order of 10^{-3} to 10^{-6} based on data from

Table 1 Correlations between genetic diversity and population size, or its surrogates

Taxa (surrogate)	Correlation	Reference
<i>Nuclear molecular markers</i>		
Across species		
Animals	0.7	1
Animal species	0.12	2
Vertebrates	0.35	2
Invertebrates	-0.08	2
All species	0.81	3
Plants	0.32	4
Fish (abundance)	~0.7	5
Populations within species		
Diverse taxa	Mean 0.46	3
Animals (island area)	Mean 0.36	3
Loci subject to balancing selection		
MHC across vertebrate species	0.66 ^a	6
Populations within species		
MHC native rats (log island area)	0.61	7
MHC pocket gophers (allozyme heterozygosity)	0.69	6
SI alleles	0.93	8
<i>mtDNA</i>		
Vertebrate species	0.45	3
1683 animal species (allozyme heterozygosity)	-0.14 ^{NS,b}	9
Eutherian mammal orders (allozyme heterozygosity)	0.86 ^b (silent) 0.84 ^b (total)	10
Mammal species (allozyme heterozygosity)	0.43	11
Mammal species (log body mass)	-0.50 ^a	12
Bird species (log body mass)	-0.30 ^a	12
Fish species (abundance)	0.32-0.43	5
Human populations	0.98	13

^aSpearman's ρ .

^bKendall's τ .

References: 1, Soulé (1976); 2, Nevo *et al.* (1984); 3, Frankham (1996); 4, Leimu *et al.* (2006); 5, McCusker and Bentzen (2010); 6, Zegers (2000); 7, Seddon and Baverstock (1998); 8, Young *et al.* (2000); 9, Bazin *et al.* (2006); 10, Mulligan *et al.* (2006); 11, Nabholz *et al.* (2008b); 12, Berlin *et al.* (2007); 13, Atkinson *et al.* (2008).

fish, oysters, shrimp and seaweed (see Coyer *et al.*, 2008; Palstra and Ruzzante, 2008). Presumably variances in family sizes increase with average fecundity, as N_e/N ratios decline as variances in family sizes increase (Wright, 1969).

Relationship of genetic diversity to census population size

Equation (2) predicts that genetic diversity will be positively correlated with census population sizes, provided current N reflects long-term N_e , and that mutation rates are similar across populations. Empirical estimates of these correlations are overwhelmingly positive (Table 1). Soulé (1976) reported a correlation of 0.7 between genetic diversity for allozymes and logarithm of N across animal species. Subsequent studies that had sufficient statistical power (particularly meta-analyses) have confirmed Soulé's conclusions across a broad range of major taxa and population sizes (up to 10^{20}).

Threatened species have, by definition, small or declining population sizes, so are expected to have reduced genetic diversity. Threatened species across a broad taxonomic range had on average 35% less genetic diversity than taxonomically related non-threatened species (Spielman *et al.*, 2004), whereas the reduction was 25% for birds (Evans and Sheldon, 2008) and 30% for tetrapods (Flight, 2010).

A significant correlation between genetic diversity and log N within species was also reported by Frankham (1996), based on a meta-analysis (mean $r=0.46$). Correlations between genetic diversity and population size are not always significant, but this seems to be due primarily to statistical power. In my meta-analysis, significant correlations were found in only seven studies, but 22 of the 23 reported correlations were positive (highly significant sign test). Several studies have also reported correlations between surrogates of population size and genetic diversity, such as range size in plants, island area (both positive), rates of chromosomal evolution and body size (both negative) (see Frankham, 1996; Table 1). Further, average genetic diversity of island populations (presumed to have smaller N) is lower than that for mainland populations (Frankham, 1997).

Populations subject to short size bottlenecks typically show reduced genetic diversity for molecular markers, compared with non-bottlenecked populations (see England *et al.*, 2003; Garner *et al.*, 2005; Frankham *et al.*, 2010, Chapter 8). In *Drosophila*, the effect of a single pair bottleneck on microsatellite diversity was close to the theoretical reduction of 25% in genetic diversity (England *et al.*, 2003) and the expected reduction in allelic diversity (Frankham *et al.*, 2010, p. 172).

Positive correlations between genetic diversity and population size are also expected for loci that are subject directly to weak balancing selection (Robertson, 1962). These have been observed for loci subject to balancing selection (Table 1), including the major histocompatibility complex in vertebrates, self-incompatibility loci in plants and inversions in *Drosophila* (Montgomery *et al.*, 2000).

Overall, there is overwhelming evidence for positive correlations between genetic diversity and population size for nuclear markers. The correlations are less than 1, but this is expected owing to sampling variation and variation in N_e/N ratios.

However, the relationship between mtDNA genetic variation and population size is controversial (Table 1). Frankham (1996) reported a correlation of 0.45, based on data for 18 vertebrate populations from 12 species. Conversely, Bazin *et al.* (2006) reported a non-significantly negative correlation (Kendall $\tau=-0.14$) between mtDNA sequence diversity and nuclear allozyme genetic diversity across eight major animal groups encompassing 1683 species, while finding a significant positive correlation between nuclear DNA diversity and allozyme diversity ($\tau=0.87$) across the same groups. Subsequent analyses using the same data set revealed significant positive correlations between mtDNA and allozyme genetic diversity across eutherian mammalian orders (Mulligan *et al.*, 2006) and within mammals (Nabholz *et al.*, 2008b). Several other studies have also reported positive correlations between mtDNA diversity and population size or its surrogates (Table 1), but these have all involved data across a narrower taxonomic range than used by Bazin *et al.* (2006).

Selection and variation in mutation rates have been suggested as the reason for the equivocal relationship between mtDNA diversity and population size (Bazin *et al.*, 2006; Eyre-Walker, 2006). First, there are multiple lines of evidence that selective sweeps and/or background selection affect mtDNA (see Rand, 2001; Ballard and Rand, 2005; Bazin *et al.*, 2006; Kivisild *et al.*, 2006; Meiklejohn *et al.*, 2007; Wares, 2009).

Second, mtDNA mutation rates differ across animal taxa (Nabholz *et al.*, 2008a, 2009, but see Charlesworth, 2010) and Bazin *et al.* (2006) did not correct for such differences. mtDNA silent site substitution rates and presumably mutation rates fall into high (flatworms, molluscs, annelid worms, bryozoans, arthropods, nematodes, echinoderms, tunicates and vertebrates) and low (angiosperms, fungi, sponges, corals, sea fans and Medusozoa) rates, with the former rates being about 10 times nuclear substitution rates and the latter about an order of magnitude slower (Hellberg, 2006). Further, there

have been several independent transitions from slow to fast rates, in both plants and animals, possibly owing to inactivation of mtDNA proof-reading or repair enzymes. Within mammals, mtDNA shows mutational hotspots and site-specific mutation rates that vary rapidly over time (Galtier *et al.*, 2006), whereas plants show rates that vary among species, over time and between loci (see Sloan *et al.*, 2009). Evolutionary changes in mutation rates are well known in asexual bacteria (see Denamur and Matic, 2006). High mutation rates are favoured during adaptation to new environments, and lowered rates once the costs of deleterious mutations exceed the benefits from new advantageous mutations, as predicted by theory (Leigh, 1970). Mutation rates are also expected to evolve in a related manner in mtDNA, cpDNA and other non-recombining chromosomes and species, provided that loci controlling mutation rates lie within the genome being affected. Conversely, in recombining genomes, mutation rates typically evolve towards minima (see Sniegowski *et al.*, 2000).

Correlations between fitness and genetic diversity

As loss of genetic diversity in random mating populations is directly related to the population average inbreeding coefficient (equation (1)), and inbreeding has deleterious impacts on reproductive fitness that are approximately linearly related to F (see Lynch and Walsh, 1998), a positive correlation is expected between population average genetic diversity and population average fitness for small populations in similar environments. This prediction has been supported in meta-analyses of data from animal and plant species ($r=0.45$ (Reed and Frankham, 2003); $r\sim 0.3$ (Leimu *et al.*, 2006); $r=0.40, 0.49$ (Markert *et al.*, 2010)).

By contrast, little relationship is expected between individual multi-locus genetic diversity for near-neutral loci and reproductive fitness within populations, unless there is heterozygote advantage, inbreeding or population structure. A positive correlation is expected if the heterozygote fitness for marker loci exceeds the weighted mean fitness of the homozygote genotypes (Deng and Fu, 1998). Meta-analyses have revealed only very weak relationships (Britten, 1996; David, 1998; Coltman and Slate, 2003; Chapman *et al.*, 2009). For example, Chapman *et al.* (2009) found a correlation of only 0.036 and effects that did not fit causal relationships between individual markers and fitness, based on a meta-analysis of 628 estimates. Further, Szulkin *et al.* (2010) concluded the available data are qualitatively and quantitatively consistent with the inbreeding hypothesis, based upon theoretical analyses and a review of the empirical evidence.

DEVIATIONS FROM NEUTRAL EXPECTATIONS FOR LOSS OF GENETIC DIVERSITY OVER GENERATIONS

Quantitative analyses reveal that there are frequently significant deviations from neutral predictions for changes in genetic diversity over generations in finite populations. Rigorous tests for deviation from neutrality that do not rely on extraneous assumptions can be performed by regressing H_t/H_0 on the pedigree inbreeding coefficient of populations: Neutrality yields a slope of -1 (equation (1)), whereas directional selection results in faster than neutral declines in genetic diversity, and balancing selection usually results in slower than neutral declines.

Slower than neutral declines in genetic diversity

Declines of average allozyme genetic diversity with pedigree inbreeding coefficients (F_p) are often slower than predicted by neutral theory (see Rumball *et al.*, 1994; Gilligan *et al.*, 2005). For example, the regression coefficient of H_t/H_0 on F_p for 40 captive *Drosophila* populations maintained for 50 generations with diverse effective

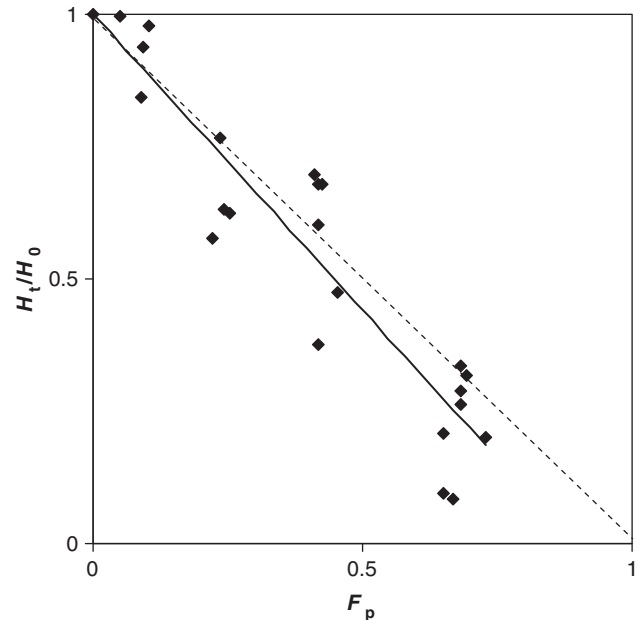


Figure 1 Relationship between proportion of microsatellite genetic diversity retained (H_t/H_0) and pedigree inbreeding coefficient (F_p) for 23 populations maintained with effective population sizes of between 25 and 500 for 48 generations (from Montgomery *et al.*, 2010). The solid line is the fitted regression to the data and the dotted line is the neutral expectation.

population sizes (all derived from the same wild population) was -0.79 ± 0.10 , significantly slower than the neutral expectation of -1 (Gilligan *et al.*, 2005). These could be due to either balancing selection on the loci themselves (see Kreitman and Hudson, 1991) or short-term associative balancing selection (see Latter, 1998; Charlesworth, 2006). If associative balancing selection is involved, the direction of deviations from neutrality for different markers will be the same. As microsatellites showed a faster than neutral decline (Figure 1), the associated balancing selection hypothesis was rejected. This implicates balancing selection on the allozymes themselves. As temporal or spatial variations in selection are improbable in our constant laboratory environment, heterozygote advantage or frequency-dependent selection are the most probable explanations for the allozyme results. The above conclusions relate to average genetic diversity and concealed some allozyme loci showing neutral behaviour and others subject to selective sweeps (Montgomery *et al.*, 2010). Molecular studies have detected selection on several allozyme loci with intermediate frequency polymorphisms in wild populations (see Eanes, 1999; Hey, 1999). However, our results reflect only the last 50 generations in a different, captive environment. Analyses indicate that balancing selection probably only affects a small proportion of loci in human genomes (Asthana *et al.*, 2005; Bubb *et al.*, 2006; Andrés *et al.*, 2009) and mice allozyme loci (Storz and Nachman, 2003).

Selective sweeps in populations adapting to new environments

Genetic diversity for non-coding microsatellites declined at a 12% faster than neutral rate in 23 *Drosophila* populations adapting to captivity over 48 generations (Figure 1; Montgomery *et al.*, 2010). Further, allele frequency changes were 33% greater than the neutral expectation, and variation among replicate populations was 25% greater than predicted by neutrality. Direct selection on the microsatellite loci themselves would have caused drift variances among

replicates to be less than neutral predictions. Our populations experienced conditions that were highly conducive to selective sweeps. Adaptations to captivity were highly deleterious when populations were returned to simulated wild conditions (Woodworth *et al.*, 2002), indicating that adaptation was due largely to initially rare alleles that were deleterious in the wild, but beneficial in captivity. Linkage disequilibrium around alleles in mutation–selection balance is expected because of the turnover of deleterious alleles. All eight microsatellite loci spread throughout the *Drosophila* genome showed at least some signals of selective sweeps, implying that the effects were genome-wide. Given the short duration and the large deviations from neutral predictions, it is improbable that background selection caused the deviation from neutrality (Charlesworth, 1996). Selective sweeps are expected in all genetically variable populations subject to environmental change (Montgomery *et al.*, 2010). Widespread selective sweeps across genomes may be missed in many current analyses, such as those based on outlier analyses (Hahn, 2007), as regions affected by sweeps may be subsumed into the common ‘control’ group, and only sequences with extreme behaviour identified as selected. Rigorous controls, such as those provided by pedigrees, provide powerful means for detecting widespread selective sweeps.

LOW GENETIC DIVERSITY IN REGIONS WITH LOW RECOMBINATION RATES

Nucleotide diversity across chromosomes is positively correlated with recombination rate in *Drosophila* (Begun and Aquadro, 1992; Shapiro *et al.*, 2007; Kulathinal *et al.*, 2008), humans (Hellmann *et al.*, 2005), white-throated sparrows (Huynh *et al.*, 2010), tomatoes (Stephan and Langley, 1998) and maize (Tenailon *et al.*, 2001), but not in *A. lyrata* (Wright *et al.*, 2006). In several *Drosophila* species, nucleotide diversity differs by ~10-fold between regions of high and low recombination (Stephan *et al.*, 1992; Aquadro *et al.*, 1994; Begun *et al.*, 2007). Selective sweeps and/or background selection could account for these results.

The hypothesized relationship between genetic diversity and recombination leads to the prediction that genetic diversity should be lower than expected from neutral theory for loci on chromosomes with very low recombination (W and Y chromosomes, mtDNA and cpDNA in eukaryotes, and small fourth chromosomes in several *Drosophila* species), provided the chromosomes retain some functional loci. Further, translocation of such a chromosome to a region of normal recombination should lead to a higher equilibrium level of genetic diversity.

By contrast, any deviations from neutrality owing to recombination rate differences between X and autosomes should be modest and of variable direction, as the effective rates of recombination on X chromosomes versus autosomes should be 2/3:1 in species with recombination (most species) and 4/3:1 in species such as *Drosophila* where there is no recombination in males (assuming that recombination rates in sex chromosomes and autosomes are otherwise similar on average).

Many studies have reported effects of linked selection on genetic diversity, but I am unaware of any compilation of the magnitude of these effects for regions of low recombination. Below I review levels of nucleotide diversity in chromosomes with low rates of recombination and compare them with data for chromosomes with more normal recombination rates. To minimize variation due to differences among populations in N_e and other extraneous variables, I compared ratios of nucleotide diversity in chromosomes with low rates of recombination (W, Y, dot fourth, cpDNA) on a within-species basis with ‘normally’ recombining autosomal, X or Z chromosomal loci, or mtDNA and

cpDNA with nuclear DNA. Further, ratios for relatively normally recombining chromosomes (X and Z) with autosomal loci were computed. All ratios were adjusted for differences in copy number of the chromosomes and for mutation rate differences (based on genetic divergences from another species at silent sites, or using the method of Ellegren (2007) that is based upon differences in male and female mutation rates derived from numbers of germ cell divisions in females versus males), so that the neutral expectation is 1. Ellegren (2009) reported X:A and Z:A ratios across species, but did not correct them for mutation rate differences.

I avoided using data where introgression from other species or sub-species was suspected and from recently bottlenecked populations, as these may distort comparisons (see Pool and Nielsen, 2007). Uninformative comparisons where mtDNA and nuclear diversity were both 0 are not included (Zhou *et al.*, 2010). As ratios are often not normally distributed, I present both means and medians, and tests for deviation from neutral predictions using non-parametric tests. Statistical tests were performed both for all estimates and for species means. Tests were one-tailed for chromosomes with low recombination rates and two-tailed for those with ‘normal’ rates.

W and Y chromosomes

Adjusted ratios of genetic diversity for non-recombining regions on W and Y chromosomes compared with that on autosomes or Z or X chromosomes were much lower than predicted by neutrality (Tables 2 and 3). For example, the ratio of nucleotide diversity for non-coding regions on the W chromosome to that on the autosomes in domestic chickens is 0.0108, and 0.075 after adjusting for the four-fold higher autosomal N_e and the 1.75-fold higher mutation rate in autosomes than W chromosomes (Berlin and Ellegren, 2004).

All 12 W chromosome ratios were much less than 1 (sign test $P < 0.0002$), the mean ratio being 0.011 ± 0.007 and the median 0.00 (Tables 2 and 3). Similarly, 22 of 23 estimates of adjusted genetic diversity for non-recombining regions on Y chromosomes were less than 1 (sign test $P < 0.0001$), with a mean of 0.226 ± 0.058 and a median of 0.110 (Tables 2 and 3).

Very low adjusted ratios of W chromosome to autosomal variation in birds are only realistically attributable to selection, as any effects of male polygamy are expected to bias the ratio upwards (Berlin and Ellegren, 2004). Conversely, adjusted ratios for Y chromosomes may be biased downwards owing to lower effective population sizes in males than females (see Frankham *et al.*, 2010). Handley *et al.* (2006) showed that the Y chromosome in the greater white-toothed shrew still showed a deficit in variation relative to the X, after accounting for mutation rate, copy number and demography, leaving directional selection as the probable explanation. Further, Gerrard and Filatov (2005) reported positive selective for two of three Y chromosomal loci across 6–12 mammalian species.

How does selection have such large impacts when W and Y chromosomes typically have few functional loci and low polymorphism? While *Drosophila melanogaster* has only 10–20 protein-coding loci on the Y chromosome, Lemos *et al.* (2008) reported that it harbours substantial polymorphic regulatory variation that affects hundreds of X linked and autosomal loci with important functions.

Drosophila chromosome-4

The small dot chromosome-4 found in several species of *Drosophila* has only about 1/100th the recombination rate of the rest of the nuclear genome. All 12 estimates of the adjusted ratio of its nucleotide diversity, compared with other recombining autosomes or X chromosomes in six species of *Drosophila*, were less than 1 (sign test $P = 0.002$),

Table 2 Nucleotide diversity for Y, W, dot fourth chromosomes (compared with diversity for X or Z chromosomal or autosomal loci) and mtDNA and cpDNA (compared with nuclear loci) as a proportion of that expected from neutral theory (corrected for differences in copy number and mutation rates)

Species	Proportion of neutral expectation ^a	References
<i>W chromosome</i>		
Barn swallow	0	1
Black grouse	0	1
Blue tit	0	1
Chiffchaff	0	2
Chicken	0.075	3
Collared flycatcher	0	1
Dusky warbler	0	2
Kestrel	0	1
Pied flycatcher	0	1
White-throated sparrow	0.056	3, 4
Willow warbler	0, 0	1, 2
<i>Y chromosomes</i>		
Bonobo	1.08	5, 6
Cattle	0	7
Chimpanzee ^b	0.02	5, 6
Field vole	0.43	7
Greater white-toothed shrew	0.49	8
Horse	0	9
Human	0.44, 0.62, 0.29, 0.47	5, 6, 10, 11, 12
Lynx	0	7
Mouse	0.53	6, 11
Reindeer	0	7
Sheep	0.11	13
Wolf	0.16	7
<i>Drosophila melanogaster</i>	0	14
<i>Drosophila miranda</i>	0.11	15, 16
<i>Drosophila santomea</i>	0.05	17
<i>Drosophila simulans</i>	0.08	14, 16
<i>Drosophila teissieri</i>	0	17
<i>Drosophila yakuba</i>	0.09	17
<i>Silene dioica</i>	0.08	18
<i>Silene latifolia</i>	0.15	19
<i>Chromosome-4 in Drosophila</i>		
Dot (very low recombination)		
<i>D. americana</i>	0.06	20
<i>D. mauritiana</i>	0.05	21
<i>D. melanogaster</i>	0, 0.04, 0.41, 0.13	21, 22, 23, 24, 25
<i>D. sechellia</i>	0	21
<i>D. simulans</i>	0.03, 0.02, 0.08, 0.07	21, 22, 23, 25, 26, 27
<i>D. yakuba</i>	0.10	15, 23, 25
Fused to another autosome (~normal recombination)		
<i>D. insularis</i>	1.34	28
<i>D. willistoni</i>	0.92, 1.28	28
<i>mtDNA</i>		
Greater white-toothed shrew	0.48	8
Mice (two sub-species)	0.91, 0.88	29

Table 2 (Continued)

Species	Proportion of neutral expectation ^a	References
Beetle (<i>Adalia bipunctata</i>) (with three strains of <i>Ricksettsia</i>)	10.12	30
Butterfly (<i>Acraea encedana</i>) (with <i>Wolbachia</i>)	0	31
<i>Drosophila innubila</i> (with <i>Wolbachia</i>)	0.17	32
<i>D. melanogaster</i>	0.18	14, 33
<i>D. recens</i> (with <i>Wolbachia</i>)	0.03	34, 35
<i>D. santomea</i>	0.18	17, 35
<i>D. simulans</i>	0.04	17, 33
<i>D. subquinaria</i> (no <i>Wolbachia</i>)	0.63	34, 35
<i>D. teissieri</i>	0.05	17, 35
<i>D. yakuba</i>	0.09	17, 35
<i>Arabidopsis lyrata</i>	0.81	36
<i>Caenorhabditis remani</i>	1.19	35, 37
<i>Chlamydomonas reinhardtii</i>	0.53	38
<i>cpDNA</i>		
<i>Arabidopsis lyrata</i>	0.80	36
<i>Arabidopsis thaliana</i>	0.75 ^c	36

^aWhere the referenced publication did not provide the corrected ratio, I have calculated it from the original data.

^bOnly *Pan troglodytes verus* data were used as there are diverged sub-species that may complicate matters and this sub-species had much higher sample sizes than the others.

^cCorrected for selfing.

References: 1, Montell *et al.* (2001); 2, Bensch *et al.* (2006); 3, Berlin and Ellegren (2004); 4, Huynh *et al.* (2010); 5, Stone *et al.* (2002); 6, Ellegren (2007); 7, Hellborg and Ellegren (2004); 8, Handley *et al.* (2006); 9, Lindgren *et al.* (2004); 10, Sachidanandam *et al.* (2001); 11, Nachman (1998); 12, Shen *et al.* (2000); 13, Meadows *et al.* (2004); 14, Zurovcova and Eanes (1999); 15, Bachtrog and Charlesworth (2000); 16, Bauer and Aquadro (1997); 17, Bachtrog *et al.* (2006); 18, Filatov *et al.* (2001); 19, Qiu *et al.* (2010); 20, Betancourt *et al.* (2009); 21, Hilton *et al.* (1994); 22, Jensen *et al.* (2002); 23, Moriyama and Powell (1996); 24, Wang *et al.* (2002); 25, Arguello *et al.* (2010); 26, Wang *et al.* (2004); 27, Begun *et al.* (2007); 28, Powell *et al.* (2011); 29, Baines and Harr (2007); 30, Jiggins and Tinsley (2005); 31, Jiggins (2003); 32, Dyer and Jaenike (2004); 33, Haag-Liautard *et al.* (2008); 34, Shoemaker *et al.* (2004); 35, Kondrashov and Kondrashov (2010); 36, Wright *et al.* (2008); 37, Graustein *et al.* (2002); 38, Smith and Lee (2008). The neutral expectation is 1.

with a mean of 0.083 ± 0.032 and a median of 0.055 (Tables 2 and 3). A causal relationship between very low recombination and low genetic diversity has been established, as translocation of the dot chromosome onto another autosome increased recombination rates and resulted in normal levels of genetic diversity for dot chromosomal loci (Tables 2 and 3; Powell *et al.*, 2011). The difference between the adjusted ratios for free and fused fourth chromosomes was highly significant (Mann–Whitney test $P=0.006$).

mtDNA and cpDNA diversity

In all of the cases in Table 2 mtDNA and cpDNA are maternally inherited or transmitted by only one mating type (*Chlamydomonas*). The criteria applied to detect deviations from neutrality for nucleotide diversity in mtDNA and cpDNA versus nuclear DNA were very stringent, as they assume a four-fold difference in ploidy in species with two sexes and double in hermaphrodites (Wright *et al.*, 2008), that only one mtDNA and cpDNA haplotype is transmitted per female gamete, that there are no paternal contributions, that none of the mtDNA sequences are integrated into the nucleus, that there has been no introgression between taxa and that male and female effective population sizes are equal. Consequently, the adjusted ratios of mtDNA and cpDNA nucleotide diversity compared with nuclear loci will typically be biased upwards by higher effective population sizes in females than males (Frankham *et al.*, 2010), widespread

Table 3 Mean and median nucleotide diversity ratios (adjusted for copy number and mutation rate differences) as compared with neutral expectation of 1 for comparisons of chromosomes for regions of low and 'normal' recombination

Chromosomal comparison	All estimates Genetic diversity		Species means Genetic diversity	
	Mean	Median	Mean	Median
<i>Low recombination</i>				
W:A ^a or X	0.011 ± 0.007	0.00	0.012 ± 0.008	0.00
Y:A or X	0.226 ± 0.058	0.11	0.192 ± 0.062	0.09
<i>Drosophila</i> dot 4th: other A	0.082 ± 0.032	0.06	0.068 ± 0.020	0.06
mtDNA: nuclear	1.018 ^b ± 0.614	0.33 ^b	1.026 ^b ± 0.656	0.18 ^b
(minus outlier)	0.394 ± 0.087	0.26	0.377 ± 0.102	0.18
cpDNA: nuclear	0.775 ± 0.025	0.78	0.775 ± 0.025	0.78
<i>'Normal' recombination</i>				
X:A	1.091 ± 0.050	1.02	1.009 ± 0.072	1.00
Z ^c :A	0.363 ± 0.100	0.40	0.363 ± 0.100	0.40
<i>Drosophila</i> dot 4th fused to another A:other A	1.180 ± 0.131	1.28	1.220 ± 0.120	1.22

^aAutosome.

^bOne outlier (see text).

^cZ recombination ~40% that of autosomes.

Means and medians are presented for all estimates and for the means of species.

occurrence of low levels of paternal leakage and recombination (see Ballard and Rand, 2005; White *et al.*, 2008), nuclear integration of mtDNA, introgression between taxa and multiple mtDNA genomes being transmitted across generations (~30 in *Drosophila melanogaster*, Haag-Liautard *et al.* (2008); ~100 in humans and Chinook salmon and ~200 in mice, White *et al.* (2008)).

Most estimates of the adjusted ratio of mtDNA diversity to nuclear diversity were less than 1 (17 out of 19, sign test $P=0.007$), but one species had a large excess of mtDNA diversity compared with nuclear levels (Table 2). Further, Hellberg (2006) reported 0 within-population mtDNA nucleotide variation in *Balanophylla elegans* coral, but high levels of allozyme variation at nuclear loci. Substantial effects on mtDNA diversity in arthropods are caused by infections with maternally transmitted endosymbionts, such as *Wolbachia* bacteria (Hurst and Jiggins, 2005). Typically invertebrates infected with one strain of endosymbiont have reduced levels of mtDNA variation, whereas those infected with multiple strains may have elevated levels of mtDNA variation. For example, *Drosophila recens* infected with the maternally transmitted *Wolbachia* shows much lower mtDNA variation than predicted by neutrality, whereas the closely related uninfected *Drosophila quinaria* has more normal levels of mtDNA. Conversely, *Adalia bipunctata* is infected with three endosymbiont strains and has highly elevated mtDNA levels, overall, but very low mtDNA levels in beetles infected with a particular endosymbiont strain (Jiggins and Tinsley, 2005).

The two adjusted ratios for cpDNA were lower than the neutral expectation (Table 2). Further, Banks and Birky (1985) reported very low cpDNA variation in *Lupinus texensis*, a species with very high levels of nuclear-encoded allozyme variation ($H_e=0.41$). cpDNA has been shown to display selective sweeps (Muir and Filatov, 2007). However, there are insufficient estimates to be sure about general conclusions.

There are strong opportunities for selection among organelle genomes within cells, among cells within individuals and among individual females within populations, especially given that mtDNA codes for critical functions in energy metabolism and cpDNA codes

for critical proteins involved in photosynthesis (Rand, 2001; White *et al.*, 2008). There is substantial evidence of selection on mtDNA (see Ballard and Whitlock, 2004; White *et al.*, 2008) and on the cpDNA-encoded *rbcl* locus (which codes for the large subunit of the Rubisco, an enzyme with a critical role in photosynthesis) of most analysed land plants (Kapralov and Filatov, 2007).

Additional comparative estimates of mtDNA and cpDNA versus nuclear variation are required across a broader range of taxa.

X and Z chromosomes

Adjusted ratios of X:autosomal genetic diversity varied on either side of the neutral expectation, with a mean of 1.091 ± 0.050 and a median of 1.02 (Tables 3 and 4). Seventeen estimates were less than 1 and 20 greater (sign test $P=0.74$). Part of the variation in these ratios is associated with differences in recombination and distance from coding loci (and effects of selection). Hammer *et al.* (2010) found that the adjusted ratio increased from 0.89 to 1.48 as the recombination distance from functional loci increased. Further, Vicoso and Charlesworth (2009) reported an adjusted ratio for nucleotide diversity in *Drosophila melanogaster* for non-coding regions of X chromosomes: autosomes of 1.57, but the adjusted ratio was 0.97 for regions with similar effective recombination rates on X and autosomes.

Deviations from neutral expectation for X chromosomal genetic diversity may reflect different effective population sizes in the two sexes, selection on linked loci, or modest differences in effective rates of recombination on X and autosomes.

Adjusted ratios of Z chromosome nucleotide diversity compared with the autosomes in birds were all substantially less than 1 (Tables 3 and 4), but with only four estimates they do not differ significantly from 1 (mean 0.363 ± 0.100 and median 0.40). However the Z:A ratios were significantly lower than the X:A ones (Mann-Whitney test $P=0.007$). The Z chromosome in chickens has a 60% lower recombination rate than the autosomes (Sundström *et al.*, 2004). The lower ratios for Z than X chromosomes are probably attributable to male polygamy and stronger selection effects for the Z chromosomes owing to its lower recombination rate.

Table 4 Nucleotide diversity for X and Z chromosome (compared with diversity for autosomal loci) as a proportion of that expected from neutral theory (corrected for differences in copy number and mutation rates)

Chromosomal comparison Species	Proportion of neutral expectation	References
<i>X:autosomes</i>		
Chimpanzee (<i>Pan troglodytes verus</i>)	0.82	1, 2
Humans	0.88	3
Humans	0.94	1, 2
Humans	0.81, 0.85, 1.01 1.45, 1.34, 1.20 1.29, 1.59, 1.56	4
Humans	0.87, 0.95, 1.11, 1.37	5
Humans	1.02, 0.85, 0.82	6
Humans	0.83	2, 7
Humans	1.37	2, 8
<i>Mus musculus</i>	0.91	9
<i>M. musculus</i> (two sub-species)	0.91, 0.76	10
<i>Drosophila melanogaster</i>	0.83	9
<i>D. melanogaster</i>	1.03	11
<i>D. melanogaster</i>	1.16, 1.96	12
<i>D. melanogaster</i>	1.36, 1.27	13
<i>D. melanogaster</i> (with comparative recombination)	1.57	14, 15
	0.97	15
<i>Drosophila simulans</i>	0.64	9
<i>D. simulans</i>	0.61	11
<i>D. simulans</i>	1.02, 1.33	12
<i>Silene latifolia</i>	1.10	16
<i>Z:autosomes</i>		
Chickens	0.35	17
Collared flycatcher	0.45	18
Pied flycatcher	0.56	18
White-throated sparrow	0.09	17, 19

References: 1, Stone *et al.* (2002); 2, Ellegren (2007); 3, Sachidanandam *et al.* (2001); 4, Bustamante and Ramachandran (2009); 5, Hammer *et al.* (2010); 6, Keinan *et al.* (2009); 7, Stephens *et al.* (2001); 8, Yu *et al.* (2002); 9, Hedrick and Parker (1997); 10, Baines and Harr (2007); 11, Zurovcova and Eanes (1999); 12, Andolfatto (2001); 13, Singh *et al.* (2007); 14, Vicoso and Charlesworth (2009); 15, Hutter and Stephan (2009); 16, Qiu *et al.* (2010); 17, Sundström *et al.* (2004); 18, Borge *et al.* (2005); 19, Huynh *et al.* (2010). The neutral expectation is 1.

DISCUSSION

Correlations between population size and genetic diversity

There is compelling evidence for positive associations between genetic diversity and population size both across species and within species. These correlations extend from non-coding sequences and allozymes to loci subject to balancing selection. The only equivocal case is mtDNA in animals and here most data sets report significant correlations, especially for analyses restricted to specific groups of organisms. Overall, these results support neutral theory, or background selection, rather than the proposal of Gillespie (2000, 2001) that there is little relationship between genetic diversity and population size owing to genetic draft (selective sweeps). Further, the results constrain the range of acceptable models of selective sweeps.

Variation in mutation rates and different selection scenarios, rather than lack of drift effects, may explain why correlations between mtDNA diversity and populations sizes are non-significant in broadly

based surveys, but significantly positive for more narrowly based ones. First, wide variation in mtDNA mutation rates have been observed (see above). Second, mtDNA and the W chromosome are associated in birds and selection on either affects genetic diversity for the other, but there is no association between mtDNA and the Y chromosomes in mammals. Third, maternally transmitted *Wolbachia* endosymbionts lead to large selective changes and reduced genetic diversity for mtDNA in many invertebrate species infected with a single strain, whereas species infected with multiple strains may have elevated mtDNA diversity (Hurst and Jiggins, 2005).

Widespread deviations from neutral expectations for loci in regions with low recombination and high linkage disequilibrium

There is overwhelming evidence of reduced genetic diversity compared with neutral expectations in circumstances with high linkage disequilibrium (Tables 3 and 5). The ratios differed significantly between the combined low-recombination and the combined 'normal' recombination data set (Mann-Whitney test $P < 0.0001$). Of the 65 estimates for low-recombination situations, as compared with 'normally' recombining controls, 62 were less than the neutral expectation (sign test $P < 0.0001$). By contrast, deviations from neutral expectation for genetic diversity on X chromosomes, Z chromosomes and fused dot chromosomes (that all have much higher recombination rates) varied on either side of neutral expectations, 22 being above 1 and 22 below (sign test $P = 1.00$). Even here there was evidence that low recombination was associated with low genetic diversity, as described above.

The conclusions above are not due to the overrepresentation of estimates from humans, laboratory species and model species, as the magnitude of effects are similar for all estimates and for species means, and the conclusions are the same (Table 3). Some of the species have structured populations where sex-specific differences in migration rates could yield biases in ratios. However, this does not apply to the fourth chromosome data or to the correlations between genetic diversity and recombination rates within chromosomes. Its effects must be modest overall given the consistent signal of low versus normal recombination rates on genetic diversity.

Related, but less extreme, effects are found in populations adapting to new environments owing to alleles that were previously deleterious and subject to mutation-selection balance (but are now favoured) and that are expected to show substantial initial linkage disequilibrium with flanking loci owing to the recent mutational origin of most mutant alleles.

While the effects of very low recombination rates on genetic diversity vary across species owing to species-specific life-history attributes, the common factor causing deviations from neutrality is selection. Reduced genetic diversity in regions with low recombination is expected with either selective sweeps (Maynard Smith and Haigh, 1974; Gillespie, 2000) or background selection (Charlesworth *et al.*, 1993), and both are probably implicated.

All non-recombining regions share many features with asexual bacteria in chemostats (with no recombination and no gene transfer among genotypes) that undergo periodic selection and evolutionary changes in mutation rates. These analogies in evolutionary behaviour deserve more attention, especially as mtDNA and cpDNA derive from captured bacteria. All are expected to show periodic selective sweeps, linkage disequilibrium, genetic diversity that is lower than simple neutral predictions, background selection, Hill-Robertson effects and evolutionary changes in mutation rates. Quantitative deviations from neutrality will differ among them according to effective population sizes; advantageous, deleterious and neutral mutation rates; effectiveness of selection (largely determined by N_e) and recombination rates

Table 5 Summary of levels of genetic diversity as compared with neutral predictions for comparisons of chromosomes in relation to levels of recombination and other causes for deviations

Chromosomal comparison	Genetic diversity	Recombination	Cause of deviation
W:A ^a or Z	<<	Very low	Selection
Y:A or X	<<	Very low	Selection, male N_e < female N_e
<i>Drosophila</i> dot 4th:other A	<<	Very low	Selection
<i>Drosophila</i> dot 4th fused to another A:other A	~	Normal	—
mtDNA:nuclear	<	Very low	Selection, variable mutation rates, male N_e < female N_e
X:A	~ (but variable)	~ 2/3 × A (recombination in both sexes) ~ 4/3 × A (no male recombination)	—
Z:A	<	0.4 × autosomes	Selection, male N_e < female N_e
Correlation between genetic diversity and recombination	Positive		Selection
Microbes adapting to new environments	<	Very low	Periodic selection, varying mutation rates, BGS
<i>Eukaryotes adapting to new environments</i>			
Microsatellites	<	Normal	SS selection
Allozymes	>	Normal	Balancing selection
Quantitative genetic variation	>	Normal	Balancing selection
Loci linked to MHC, SI and other loci experiencing balancing selection	>	Normal	Associative balancing selection

Abbreviations: BGS, background selection; SS, selective sweeps.
^aAutosomal.

(Maruyama and Birky, 1991). The assumption of neutral behaviour is not credible for any of these chromosomes: The default hypothesis should be that they are being affected by selection (see also Ballard and Whitlock, 2004; Ballard and Rand, 2005; Hurst and Jiggins, 2005; Wares, 2009).

At first it may seem surprising that W chromosomal diversity appears, if anything, to deviate more from neutrality than Y chromosomes, whereas male polygamy would cause a difference in the opposite direction. However, the W chromosome and mtDNA in birds (and Lepidoptera) are both maternally inherited and selection on one affects the other and vice versa (Berlin *et al.*, 2007). Conversely, such combined effects are not expected for paternally transmitted Y chromosomes. A similar combined impact of selection is expected in plants with maternal inherited cpDNA and mtDNA (Mohanty *et al.*, 2003).

mtDNA often yields phylogenies that are correct in spite of background selection and periodic selective sweeps (see above), but there are many reported conflicts between phylogenies derived from mtDNA versus nuclear loci (see Ballard and Whitlock, 2004; Hurst and Jiggins, 2005; Kapralov and Filatov, 2007). Levels of mtDNA within populations and divergences among populations will represent the net effects of mutation, drift, selection, gene flow and any recombination. Despite reduced levels of mtDNA compared with neutral predictions, most animal species show mtDNA polymorphisms (Bazin *et al.*, 2006) due largely to its high mutation rate. Phylogenies will often be adversely affected by selective sweeps on mtDNA and cpDNA in the initial phase of divergence from a polymorphic common ancestor (Hickerson *et al.*, 2006). Selective sweeps of new mutations later in the process are unlikely to affect the phylogenetic structure, but may distort branch lengths. Background selection will probably have only a modest effect on branch lengths and little effect on their relative values. The use of relaxed models that account for differences in substitution rates among lineages is expected to reduce problems due to selection and mutation rate differences, and they seem to provide better results (Whelan *et al.*, 2001). However, phylogenies based on mtDNA can, at best, be

considered as generating a hypothesis about relationship. They need to be independently corroborated with analyses of DNA sequences for multiple nuclear loci or irreversible transposon insertions.

DNA barcoding, the molecular identification system that is being used to discover new species and to estimate the approximate number of animal species on Earth, is based on sequencing a section of the mtDNA cytochrome oxidase-I locus (see Hebert *et al.*, 2003). Its efficacy is affected by selective sweeps, especially those associated with *Wolbachia* (see Hurst and Jiggins, 2005). Barcoding is less than 70% successful in identifying Dipteran species, a group susceptible to this bacterium (Meier *et al.*, 2006; Whitworth *et al.*, 2007).

Some inferences based on genetic data will suffer only mild distortions when markers show behaviour that deviates from neutrality, but for others distortions may be large. Great care is required in using markers on non-recombining chromosomes (Y and W chromosomes, mtDNA and cpDNA) for making inferences about populations. Inferences may also be distorted in populations that have recently moved to new environments.

What can be done about deviations from neutrality for markers?

- (1) Routinely test non-coding 'neutral' markers for signals of selection (see Frankham *et al.*, 2010).
- (2) Test conclusions from non-recombining chromosomal loci against results for nuclear autosomal loci (from regions with 'normal' recombination).
- (3) Test the robustness of conclusions to deviations from neutrality of non-coding loci that are being used as controls (for example, by using simulations).
- (4) To reflect genome-wide measures derived from genetic data (N_e , migration rates, population structures, etc.), non-coding loci need to be sampled from across the genome, and to encompass loci from regions of the genome with high, medium and low recombination rates in a representative manner. This is difficult for species that have not been genetically mapped and sequenced.

CONCLUSIONS

- (1) There is extensive empirical evidence for positive correlations between genetic diversity and population size, as predicted by neutral theory.
- (2) It is not credible to assume that loci on W and Y chromosomes, mtDNA and cpDNA, and other non-recombining chromosomes will follow neutral predictions, as they lose genetic diversity at a faster than neutral rate owing to selective sweeps and background selection.
- (3) Inherently neutral loci will typically show higher genetic diversity and slower loss of genetic diversity in regions flanking loci subject to balancing selection, especially those surrounding complementary sex-determining loci (in Hymenoptera), major histocompatibility complex and self-incompatibility loci.
- (4) Even inherently neutral loci may not behave neutrally, especially those in regions of low recombination or in populations adapting to new environments where faster than neutral declines may be evident for loci across the genome.
- (5) In populations adapting to new environments, use of non-coding sequences and synonymous sites as neutral controls to detect selected loci as outliers may miss many selected loci, owing to the occurrence of many selective sweeps that affect control loci.
- (6) Sampling of loci to estimate effective population sizes, dispersal rates, population structure, etc., is important. It should avoid regions known to be subject to balancing or directional selection, or to have very low recombination rates.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGEMENTS

I thank Shannon Corrigan, Mark Eldridge, Michael Gillings, Matthew Hahn, Roosa Leimu, Adam Stow and two anonymous referees for comments on the manuscript, and Don Colgan for answering queries about the use of mtDNA in taxonomy.

Andolfatto P (2001). Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol Biol Evol* **18**: 279–290.

Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR *et al.* (2009). Targets of balancing selection in the human genome. *Mol Biol Evol* **26**: 2755–2764.

Aquadro CF, Begun DJ, Kindahl EC (1994). Selection, recombination and DNA polymorphism in *Drosophila*. In: Golding B (ed). *Non-neutral Evolution: Theories and Molecular Data*. Chapman & Hall: New York. pp 46–56.

Arguello JR, Zhang Y, Kado T, Fan C, Zhao R, Innan H *et al.* (2010). Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. *Mol Biol Evol* **27**: 848–861.

Asthana S, Schmidt S, Sunyaev S (2005). A limited role for balancing selection. *Trends Genet* **21**: 30–32.

Atkinson QD, Gray RD, Drummond AJ (2008). mtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol Biol Evol* **25**: 468–474.

Atwood KC, Schneider LK, Ryan FJ (1951). Periodic selection in *Escherichia coli*. *Proc Natl Acad Sci USA* **37**: 146–155.

Bachtrog D, Charlesworth B (2000). Reduced levels of microsatellite variability on the neo-Y chromosome of *Drosophila miranda*. *Curr Biol* **10**: 1025–1031.

Bachtrog D, Thornton K, Clark A, Andolfatto P (2006). Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution* **60**: 292–302.

Baines JF, Harr B (2007). Reduced X-linked diversity in derived populations of house mice. *Genetics* **175**: 1911–1921.

Ballard JWO, Rand DM (2005). The population biology of mitochondrial DNA and its phylogenetic implications. *Annu Rev Ecol Syst* **36**: 621–642.

Ballard JWO, Whitlock MC (2004). The incomplete natural history of mitochondria. *Mol Ecol* **13**: 729–744.

Banks JA, Birky CWJ (1985). Chloroplast DNA diversity is low in a wild plant, *Lupinus texensis*. *Proc Natl Acad Sci USA* **82**: 6950–6954.

Bauer VL, Aquadro CF (1997). Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. *Mol Biol Evol* **14**: 1252–1257.

Bazin E, Glémin S, Galtier N (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570–572.

Begun DJ, Aquadro CF (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW *et al.* (2007). Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* **5**: e310.

Bensch S, Irwin DE, Irwin JH, Kvist L, Akesson S (2006). Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus* warblers. *Mol Ecol* **15**: 161–171.

Berlin S, Ellegren H (2004). Chicken W: a genetically uniform chromosome in a highly variable genome. *Proc Natl Acad Sci USA* **101**: 15967–15969.

Berlin S, Tomaras D, Charlesworth B (2007). Low mitochondrial variability in birds may indicate Hill–Robertson effects on the W chromosome. *Heredity* **99**: 389–396.

Berry AJ, Ajioka JW, Kreitman M (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**: 1111–1117.

Betancourt AJ, Welch JJ, Charlesworth B (2009). Reduced effectiveness of selection caused by a lack of recombination. *Curr Biol* **19**: 655–660.

Borge T, Webster MT, Andersson G, Saetre G-P (2005). Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. *Genetics* **171**: 1861–1873.

Britten HB (1996). Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* **50**: 2158–2164.

Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M *et al.* (2006). Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* **173**: 2165–2177.

Burger J, Kirchner M, Bramanti B, Haak W, Thomas MG (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci USA* **104**: 3736–3741.

Bustamante CD, Ramachandran S (2009). Evaluating signatures of sex-specific processes in the human genome. *Nat Genet* **41**: 8–10.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009). A quantitative review of heterozygosity–fitness correlations in animal populations. *Mol Ecol* **18**: 2746–2765.

Charlesworth B (1996). Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet Res* **68**: 131–149.

Charlesworth B (2009). Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* **10**: 195–205.

Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I (2009). Genetic recombination and molecular evolution. *Cold Spring Harb Symp Quant Biol* **74**: 177–186.

Charlesworth B, Morgan MT, Charlesworth D (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.

Charlesworth D (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* **2**: e64.

Charlesworth D (2010). Don't forget the ancestral polymorphisms. *Heredity* **105**: 509–510.

Clark RM, Linton E, Messing J, Doebley JF (2004). Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci USA* **101**: 700–707.

Cohan FM (2005). Periodic selection and ecological diversity in bacteria. In: Nurmsky D (ed). *Selective Sweeps*. Kluwer: New York. pp 78–93.

Coltman DW, Slate J (2003). Microsatellite measures of inbreeding: a meta-analysis. *Evolution* **57**: 971–983.

Coyer J, Hoarau G, Kjersti S, Olsen J (2008). Being abundant is not enough: a decrease in effective population size over 8 generations in a Norwegian population of the seaweed, *Fucus serratus*. *Biol Lett* **4**: 755–757.

Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper and Rowe: New York.

David P (1998). Heterozygosity–fitness correlations: new perspectives on old problems. *Heredity* **80**: 531–537.

Denamur E, Matic I (2006). Evolution of mutation rates in bacteria. *Mol Microbiol* **60**: 820–827.

Deng H-W, Fu Y-X (1998). Conditions for positive and negative correlations between fitness and heterozygosity in equilibrium populations. *Genetics* **148**: 1333–1340.

Dyer KA, Jaenike J (2004). Evolutionarily stable infection by a male-killing endosymbiont in *Drosophila innubila*: molecular evidence from the host and parasite genomes. *Genetics* **168**: 1443–1455.

Eanes WF (1999). Analysis of selection on enzyme polymorphisms. *Annu Rev Ecol Syst* **30**: 301–326.

Ellegren H (2007). Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc R Soc Lond B Biol Sci* **274**: 1–10.

Ellegren H (2009). The different levels of genetic diversity in sex chromosomes and autosomes. *Trends Genet* **25**: 278–284.

England PR, Osler GHR, Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2003). Effect of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conserv Genet* **4**: 595–604.

Evans SR, Sheldon BC (2008). Interspecific patterns of genetic diversity in birds: correlations with extinction risk. *Conserv Biol* **22**: 1016–1025.

Eyre-Walker A (2006). Size does not matter for mitochondrial DNA. *Science* **312**: 537–538.

Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*, 4th edn. Longman: Harlow, England.

- Filatov DA, Laporte V, Vitte C, Charlesworth D (2001). DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Mol Biol Evol* **18**: 1442–1454.
- Flight PA (2010). Phylogenetic comparative methods strengthen evidence for reduced genetic diversity among endangered tetrapods. *Conserv Biol* **24**: 1307–1315.
- Frankham R (1995). Effective population size/adult population size ratios in wildlife: a review. *Genet Res* **66**: 95–107.
- Frankham R (1996). Relationship of genetic variation to population size in wildlife. *Conserv Biol* **10**: 1500–1508.
- Frankham R (1997). Do island populations have lower genetic variation than mainland populations? *Heredity* **78**: 311–327.
- Frankham R, Ballou JD, Briscoe DA (2010). *Introduction to Conservation Genetics*, 2nd edn. Cambridge University Press: Cambridge, UK.
- Galtier N, Enard D, Radonyi Y, Bazin E, Belkhir K (2006). Mutation hotspots in mammalian mitochondrial DNA. *Genome Res* **16**: 215–222.
- Garner A, Rachlow JL, Hicks JF (2005). Patterns of genetic diversity and its loss in mammalian populations. *Conserv Biol* **19**: 1215–1221.
- Gerrard DT, Filatov DA (2005). Positive and negative selection on mammalian Y chromosomes. *Mol Biol Evol* **22**: 1423–1432.
- Gillespie JH (2000). Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**: 909–919.
- Gillespie JH (2001). Is the population size of a species relevant to its evolution? *Evolution* **55**: 2161–2169.
- Gilligan DM, Briscoe DA, Frankham R (2005). Comparative losses of quantitative and molecular genetic variation in finite populations of *Drosophila melanogaster*. *Genet Res* **85**: 47–55.
- Graustein A, Gaspar JM, Walters JR, Palopoli MF (2002). Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics* **161**: 99–107.
- Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD (2008). Direct estimation of mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol* **6**: e204.
- Hahn MW (2007). Towards a selection theory of molecular evolution. *Evolution* **62**: 255–265.
- Hammer MF, Woerner AE, Mendez FL, Watkins JC, Cox MP, Wall JD (2010). The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat Genet* **42**: 830–831.
- Handley LJJ, Berset-Brändli L, Perrin N (2006). Disentangling reasons for low Y chromosome variation in the greater white-toothed shrew (*Crocidura russula*). *Genetics* **173**: 935–942.
- Hasselmann M, Beyre M (2006). Pronounced differences of recombination activity at the sex determination locus of the honeybee, a locus under strong balancing selection. *Genetics* **174**: 1469–1480.
- Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003). Biological identification through DNA barcodes. *Proc R Soc Lond B Biol Sci* **270**: 313–321.
- Hedrick PW, Parker JD (1997). Evolutionary genetics of haplo-diploid organisms and X-linked chromosomes. *Annu Rev Ecol Syst* **28**: 55–83.
- Hellberg ME (2006). No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol Biol* **6**: 24.
- Hellborg L, Ellegren H (2004). Low level of nucleotide diversity in mammalian Y chromosomes. *Mol Biol Evol* **21**: 158–163.
- Hellmann I, Prüfer K, Ji H, Zody MC, Pääbo S, Ptak SE (2005). Why do human diversity levels vary at a megabase scale? *Genome Res* **15**: 1222–1231.
- Hermisson J, Pennings PS (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**: 2335–2352.
- Hey J (1999). The neutralist, the fly and the selectionists. *Trends Ecol Evol* **14**: 35–38.
- Hickerson MJ, Meyer CP, Moritz C (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst Biol* **55**: 729–739.
- Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection. *Genet Res* **8**: 269–294.
- Hilton H, Kliiman RM, Hey J (1994). Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. *Evolution* **48**: 1900–1913.
- Hurst GDD, Jiggins FM (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc Lond B Biol Sci* **272**: 1525–1534.
- Hutter S, Stephan W (2009). Reply to Beatriz Vicoso and Brian Charlesworth. *Genetics* **181**: 1703.
- Huynh LY, Maney DL, Thomas JW (2010). Contrasting population genetic patterns within the white-throated sparrow genome (*Zonotrichia albicollis*). *BMC Genet* **11**: 96.
- Jensen MA, Charlesworth B, Kreitman M (2002). Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. *Genetics* **160**: 493–507.
- Jiggins FM (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps. Hybrid introgression and parasite population dynamics. *Genetics* **164**: 5–12.
- Jiggins FM, Tinsley MC (2005). An ancient mitochondrial polymorphism in *Adalia bipunctata* linked to a sex-ratio-distorting bacterium. *Genetics* **171**: 1115–1124.
- Kamau E, Charlesworth B, Charlesworth D (2007). Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. *Genetics* **176**: 2357–2369.
- Kapralov MV, Filatov DA (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol Biol* **7**: 73.
- Keinan A, Mullikin JC, Patterson N, Reich D (2009). Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet* **41**: 66–70.
- Kimura M (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press: Cambridge.
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K *et al.* (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**: 373–387.
- Kondrashov FA, Kondrashov AS (2010). Measurements of spontaneous rates of mutations in the recent past and the near future. *Philos Trans R Soc Lond B Biol Sci* **365**: 1169–1176.
- Kreitman M, Hudson RR (1991). Inferring the evolutionary histories of the Adh and Adh-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**: 565–582.
- Kulathinal RJ, Sarah M, Bennett SM, Fitzpatrick CL, Noor MAF (2008). Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. *Proc Natl Acad Sci USA* **105**: 10051–10056.
- Latter BDH (1998). Mutant alleles of small effect are primarily responsible for the loss of fitness with slow inbreeding in *Drosophila melanogaster*. *Genetics* **148**: 1143–1158.
- Leigh EGJ (1970). Natural selection and mutability. *Am Nat* **104**: 301–305.
- Leimu R, Mutikainen P, Koricheva J, Fischer M (2006). How general are positive relationships between plant population size, fitness and genetic variation? *J Ecol* **94**: 942–952.
- Lemos B, Araripe LO, Hartl DL (2008). Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319**: 91–93.
- Lindgren G, Backström N, Swinburne J, Hellborg L, Einarsson A, Sandberg K *et al.* (2004). Limited number of patrilineal lines in horse domestication. *Nat Genet* **36**: 335–336.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland, MA.
- Markert JA, Champlin DM, Gutjahr-Gobell R, Grear JS, Kuhn A, McGreevy TJJ *et al.* (2010). Population genetic diversity and fitness in multiple environments. *BMC Evol Biol* **10**: 205.
- Maruyama T, Birky CWJ (1991). Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. *Genetics* **127**: 449–451.
- Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35.
- McCusker MR, Bentzen P (2010). Positive relationships between genetic diversity and abundance in fishes. *Mol Ecol* **19**: 4852–4862.
- Meadows JRS, Hawken RJ, Kijas JW (2004). Nucleotide diversity on the ovine Y chromosome. *Anim Genet* **35**: 379–385.
- Meier R, Shiyang K, Vaidya G, Ng PKL (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst Biol* **55**: 715–728.
- Meiklejohn CD, Montooth KL, Rand DM (2007). Positive and negative selection on the mitochondrial genome. *Trends Genet* **23**: 259–263.
- Mohanty A, Martín JP, González LM, Aguinalde I (2003). Association between chloroplast DNA and mitochondrial DNA haplotypes in *Prunus spinosa* L. (Rosaceae) populations across Europe. *Ann Bot* **92**: 749–755.
- Montell H, Fridolfsson A-K, Ellegren H (2001). Contrasting levels of nucleotide diversity on the avian Z and W sex chromosomes. *Mol Biol Evol* **18**: 2010–2016.
- Montgomery ME, Woodworth LM, England PR, Briscoe DA, Frankham R (2010). Widespread selective sweeps affecting microsatellites in *Drosophila* populations adapting to captivity: implications for captive breeding programs. *Biol Conserv* **143**: 1842–1849.
- Montgomery ME, Woodworth LM, Nurthen RK, Gilligan DM, Briscoe DA, Frankham R (2000). Relationships between population size and genetic diversity: comparisons of experimental results with theoretical predictions. *Conserv Genet* **1**: 33–43.
- Moriyama EN, Powell JR (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Mol Biol Evol* **13**: 261–277.
- Muir G, Filatov D (2007). A selective sweep in the chloroplast DNA of dioecious *Silene* (section *Elisanthe*). *Genetics* **177**: 1239–1247.
- Mulligan CJ, Kitchen A, Miyamoto MM (2006). Comment on “Population size does not influence mitochondrial genetic diversity in animals”. *Science* **314**: 1390a.
- Nabholz B, Glémin S, Galtier N (2008a). Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol* **25**: 120–130.
- Nabholz B, Glémin S, Galtier N (2009). The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol Biol* **9**: 54.
- Nabholz B, Mauffrey J-F, Bazin E, Galtier N, Glémin S (2008b). Determination of mitochondrial genetic diversity in mammals. *Genetics* **178**: 351–361.
- Nachman MW (1998). Y chromosome variation of mice and men. *Mol Biol Evol* **15**: 1744–1750.
- Nevo E, Biales A, Ben-Shlomo R (1984). The evolutionary significance of genetic diversity: ecological, demographic and life history correlates. In: Mani GS (ed). *Evolutionary Dynamics of Genetic Diversity*. Springer-Verlag: Berlin, pp 13–213.
- Nordborg M, Charlesworth B, Charlesworth D (1996). The effect of recombination on background selection. *Genet Res* **67**: 159–174.
- Nurminsky D (ed). (2005). *Selective Sweep*. Kluwer: New York.
- O’Hugin C, Satta Y, Hausmann A, Dawkins RL, Klein J (2000). The implications of intergenic polymorphism for major histocompatibility complex evolution. *Genetics* **156**: 867–877.
- Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD (2006). Selection under domestication: Evidence for a sweep in the rice *Waxy* genomic region. *Genetics* **173**: 975–983.
- Palstra FP, Ruzzante DE (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* **17**: 3428–3447.
- Pool JE, Nielsen R (2007). Population size changes reshape genomic patterns of diversity. *Evolution* **61**: 3001–3006.

- Powell JR, Dion K, Papaceit M, Aguadé M, Vicario S, Garrick RC (2011). Non-recombining genes in a recombination environment: the *Drosophila* 'dot' chromosome. *Mol Biol Evol* **28**: 825–833.
- Qiu S, Bergero R, Forrest A, Kaiser VB, Charlesworth D (2010). Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes. *Proc R Soc Lond B Biol Sci* **277**: 3282–3290.
- Rand DM (2001). The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst* **32**: 415–448.
- Reed DH, Frankham R (2003). Correlation between fitness and genetic diversity. *Conserv Biol* **17**: 230–237.
- Robertson A (1962). Selection for heterozygotes in small populations. *Genetics* **47**: 1291–1300.
- Ruggiero MV, Jacquemin B, Castric V, Vekemans X (2008). Hitch-hiking to a locus under balancing selection: high sequence diversity and low population subdivision at the S-locus genomic region in *Arabidopsis halleri*. *Genet Res* **90**: 37–46.
- Rumball W, Franklin IR, Frankham R, Sheldon BL (1994). Decline in heterozygosity under full-sib and double first-cousin inbreeding in *Drosophila melanogaster*. *Genetics* **136**: 1039–1049.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G *et al.* (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928–933.
- Seddon JM, Baverstock PR (1998). Variation on islands: major histocompatibility complex (*Mhc*) polymorphism in populations of the Australian bush rat. *Mol Ecol* **8**: 2071–2079.
- Shapiro JA, Huang W, Zhang C, Hubisz MJ, Lu J, Turissini DA *et al.* (2007). Adaptive genic evolution in the *Drosophila* genomes. *Proc Natl Acad Sci USA* **104**: 2271–2276.
- Shen P, Wang F, Underhill PA, Franco C, Yang W-H, Roxas A *et al.* (2000). Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci USA* **97**: 7354–7359.
- Shoemaker DD, Dyer KA, Ahrens M, McAbee K, Jaenike J (2004). Decreased diversity but increased substitution rate in host mtDNA as a consequence of *Wolbachia* endosymbiont infection. *Genetics* **168**: 2049–2058.
- Singh ND, Macpherson JM, Jensen JD, Petrov DA (2007). Similar levels of X-linked and autosomal nucleotide variation in African and non-African populations of *Drosophila melanogaster*. *BMC Evol Biol* **7**: 202.
- Sloan DB, Oxelman B, Rautenberg A, Taylor DR (2009). Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe *Sileneae*. *BMC Evol Biol* **9**: 260.
- Smith DR, Lee RW (2008). Nucleotide diversity in the mitochondrial and nuclear compartments of *Chlamydomonas reinhardtii*: investigating the origins of genome architecture. *BMC Evol Biol* **8**: 156.
- Sniegowski PD, Gerrish PJ, Johnson T, Shaver A (2000). The evolution of mutation rates: separating causes from consequences. *BioEssays* **22**: 1057–1066.
- Soulé ME (1976). Allozyme variation, its determinants in space and time. In: Ayala FJ (ed). *Molecular Evolution*. Sinauer: Sunderland, MA. pp 60–77.
- Spielman D, Brook BW, Frankham R (2004). Most species are not driven to extinction before genetic factors impact them. *Proc Natl Acad Sci USA* **101**: 15261–15264.
- Stephan W, Langley CH (1998). DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**: 1585–1593.
- Stephan W, Wiehe THE, Lenz MW (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor Pop Biol* **41**: 237–254.
- Stephens CJ, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE *et al.* (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**: 489–493.
- Stone AC, Griffiths RC, Zegura SL, Hammer MF (2002). High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc Natl Acad Sci USA* **99**: 43–48.
- Storz JF, Nachman MW (2003). Natural selection on protein polymorphisms in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution* **57**: 2628–2635.
- Sundström H, Webster MT, Ellegren H (2004). Reduced variation on the chicken Z chromosome. *Genetics* **167**: 377–385.
- Szulkin M, Bierne N, David P (2010). Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* **64**: 1202–1217.
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* **98**: 9161–9166.
- Vicoso B, Charlesworth B (2009). Recombination rates may affect the ratio of X to autosomal noncoding polymorphism in African populations of *Drosophila melanogaster*. *Genetics* **181**: 1699–1701.
- Wang W, Thornton K, Berry A, Long M (2002). Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* **295**: 134–137.
- Wang W, Thornton K, Emerson JJ, Long M (2004). Nucleotide variation and recombination along the fourth chromosome in *Drosophila simulans*. *Genetics* **166**: 1783–1794.
- Wares JP (2009). Natural distribution of mitochondrial sequence diversity support new null hypothesis. *Evolution* **64**: 1136–1142.
- Whelan S, Liò P, Goldman N (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet* **17**: 262–272.
- White DJ, Wolff JN, Pierson M, Gemmill NJ (2008). Revealing the hidden complexities of mtDNA inheritance. *Mol Ecol* **17**: 4925–4942.
- Whitworth TL, Dawson RD, Magalon H, Baudry E (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protophila* (Diptera: Calliphoridae). *Proc R Soc Lond B Biol Sci* **274**: 1731–1739.
- Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2002). Rapid genetic deterioration in captivity: causes and conservation implications. *Conserv Genet* **3**: 277–288.
- Wright S (1969). *Evolution and the Genetics of Populations. 2. The Theory of Gene Frequencies*. University of Chicago Press: Chicago.
- Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Loosely M, Gaut BS *et al.* (2006). Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* **174**: 1421–1430.
- Wright SI, Nano N, Foxe JP, Nisa Dar V-U (2008). Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*. *Genet Res* **90**: 119–128.
- Young AG, Brown AHD, Murray BG, Thrall PH, Miller CH (2000). Genetic erosion, restricted mating and reduced viability in fragmented populations of the endangered grassland herb *Rutidosis leptorrhynchoides*. In: Young AG, Clarke GM (eds). *Genetics, Demography and Viability of Fragmented Populations*. Cambridge University Press: Cambridge, UK. pp 335–359.
- Yu N, Chen F-C, Ota S, Jorde LB, Pamilo P, Patthy L *et al.* (2002). Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**: 269–274.
- Zegers G (2000). Genetic variability and resistance to infectious disease with particular emphasis on the major histocompatibility complex in the valley pocket gopher. PhD thesis, University of California, Santa Cruz.
- Zhou R, Qiu S, Zhang M, M G, Chen S, Shi S (2010). *Sonneratia ovata*—a genetically depauperate mangrove species. *Biochem Syst Ecol* **38**: 697–701.
- Zurovcova M, Eanes WF (1999). Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*. *Genetics* **153**: 1709–1715.