Compensatory mutations are repeatable and clustered within proteins

Brad H. Davis^{1,*}, Art F. Y. Poon^{1,2} and Michael C. Whitlock¹

¹*Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4*

²*Department of Pathology and UCSD Antiviral Research Center, University of California, San Diego, CA 92103, USA*

Compensatory mutations improve fitness in genotypes that contain deleterious mutations but have no beneficial effects otherwise. As such, compensatory mutations represent a very specific form of epistasis. We show that intragenic compensatory mutations occur non-randomly over gene sequence. Compensatory mutations are more likely to appear at some sites than others. Moreover, the sites of compensatory mutations are more likely than expected by chance to be near the site of the original deleterious mutation. Furthermore, compensatory mutations tend to occur more commonly in certain regions of the protein even when controlling for clustering around the site of the deleterious mutation. These results suggest that compensatory evolution at the protein level is partially predictable and may be convergent.

Keywords: compensatory mutation; deleterious mutations; experimental evolution; epistasis; primary structure

1. INTRODUCTION

Compensatory mutations are the result of a particular form of epistasis, in which the new mutation has a beneficial effect on fitness when a deleterious mutation is present but is otherwise neutral or deleterious. Compensatory mutations are an important yet poorly understood aspect of biological evolution with profound implications. For instance, antibiotic resistance in bacteria, pesticide resistance in agricultural pest species and failure of antiretroviral therapy in HIV-1-infected patients have all been linked to the occurrence of compensatory mutations (Schrag & Perrot 1996; Maisiner-Patin & Andersson 2004). Resistance mutations are often associated with substantial fitness costs in non-selective environments. Compensatory mutations can at least partially offset these costs, allowing populations to retain their resistance in the absence of the selective agent. Compensatory mutations may also play an important role in conservation genetics, because compensation allows small populations to recover from fixation of deleterious mutations by genetic drift (Whitlock 2000; Poon & Otto 2000; Whitlock *et al.* 2003). Compensatory mutations have been implicated in the formation of Dobzhansky–Muller incompatibilities (Kondrashov *et al.* 2002), which makes them of general interest to evolutionary biology.

Despite the obvious importance of compensatory mutations, we understand relatively little about their basic biology, although there has recently been a surge of interest in compensatory mutations. For example, Kascser & Burns (1973, 1981) developed metabolic control theory by thinking about metabolic pathways from an evolutionary point of view. Using this metabolic control theory, Kascser & Burns (1981) showed that most enzymes in linear pathways could have their performances changed with little impact on

fitness. Hartl & Taubes (1996) showed that, under these circumstances, a great ability exists to compensate for mildly deleterious mutations. Whitlock *et al.* (1995) showed that compensatory epistasis is likely to be a general consequence of any form of stabilizing selection.

There has also been considerable experimental evidence supporting the existence of compensatory mutations. Burch & Chao (1999) investigated fitness recovery in a strain of the $\phi 6$ virus fixed with a deleterious mutation causing a 90 per cent reduction in its fitness. They observed fitness recovery over many population sizes, including a population with an effective size of 60 and nearly perfect fitness recovery in the populations with large effective sizes. In the smaller populations, fitness recovery occurred in a stepwise fashion, indicating that the fitness recovery was not due to a single back mutation, but by new compensatory mutations at other sites. Rokyta *et al.* (2002) found that following the deletion of the ligase gene in the bacteriophage T7, fitness dropped dramatically, but most of this fitness loss was recovered by compensatory changes to other genes. Moore *et al.* (2000) found that low-fitness mutant genotypes recovered fitness more rapidly than high-fitness mutant genotypes. The eukaryote *Caenorhabditis elegans* has also demonstrated rapid fitness recovery from the accumulation of deleterious mutations (Estes & Lynch 2003). However, it is unclear whether the resulting fitness improvements observed in the latter two experiments were due to back mutation or compensatory mutations at other sites. Compensatory mutations are not the only way to overcome a deleterious mutation. Crill *et al.* (2000) performed an experiment with the bacteriophage $\phi X174$ in which the target gene was mutated, and they found no evidence for compensatory changes, only back mutations.

There is substantial evidence in favour of relatively high compensatory mutation rates, but the properties of compensatory mutations are not well understood.

* Author and address for correspondence: Biotechnology Institute, University of Minnesota, 1479 Gortner Avenue, St Paul, MN 55108, USA (davi1847@umn.edu).

Figure 1. Location of compensatory and deleterious mutations along the length of their genes. Dots denote the location of deleterious mutants and lines denote the location of compensatory mutations. The height of the bar above each compensatory mutation site is proportional to the number of lines that showed this mutation.

Poon *et al.* (2005) investigated the distribution of the number of compensatory mutations and the proportion of compensatory mutations that were intragenic rather than intergenic, across a broad taxonomic range covering the viral, prokaryotic and eukaryotic kingdoms. Poon *et al.* (2005) found that compensatory mutations were abundant overall, with a mean of 11.8 per deleterious mutation and substantial variation in fitness effect that was best described by an L-shaped gamma distribution function. Furthermore, the majority of compensatory mutations were intragenic, with a significantly lower fraction in viruses (69%) than in prokaryotes (92%) or eukaryotes (90%). Therefore, understanding intragenic relationships both among compensatory mutations and between compensatory mutations and their associated deleterious mutations is important to improving our understanding of compensatory mutations in general. Furthermore, studies on three viral proteins have found that compensatory mutations tend to be more effective when found closer to the site of the deleterious mutation in terms of the protein's primary structure (Poon & Chao 2006), but this pattern has not been examined on a broader scale.

While analysing the data in the previous study (Poon *et al.* 2005), we observed what appeared to be non-random associations between the location of compensatory mutations and their associated deleterious mutations in terms of their positions within the primary sequence of the protein (figure 1). In this paper, we investigate the relationship between the position of deleterious mutations and their compensatory mutations. We asked three related questions:

- (i) Are all amino acid residues within a protein's primary structure equally likely to produce compensatory mutations?
- (ii) Do compensatory mutations tend to occur around the site of their associated deleterious mutations?

- (iii) Accounting for the location of the deleterious mutation, do compensatory mutations show evidence of clustering around particular amino acid residues within a protein's primary structure?

We addressed each of these questions using the entire dataset of Poon *et al.* (2005), for all taxa combined and for each of three taxonomic groups for which there is enough data: eukaryotes, prokaryotes, and viruses. Unfortunately, we lacked sufficient data to examine these trends at greater taxonomic resolution.

2. MUTATIONAL DATA

We used the dataset collected by Poon *et al.* (2005), which comprised compensatory mutations from 67 published articles. Among 77 different deleterious mutations for which compensatory mutations were recovered, a total of 602 compensatory mutations were identified. The data were sampled from across a broad taxonomic spectrum including four viral, five prokaryotic and nine eukaryotic species. Most of these represented experimental model systems (e.g. *C. elegans*, *Escherichia coli*). For this study, for a mutation to be considered compensatory, it must have occurred in a different codon than the deleterious mutation. All compensatory mutations considered in this study were intragenic point mutations that occur within the protein-coding region.

(a) Question 1: are some amino acid residues more likely to mutate with compensatory effects than others?

To evaluate the biological significance of the location of compensatory mutations in the primary structure, we first determined whether such mutations occurred at similar codon positions more often than expected by chance. For this purpose, we employed an index of dispersion $\rho = \sigma/\mu$, where σ is the variance across the sequence in the number of mutations per amino acid residue and μ is the mean number of compensatory mutations per amino acid residue. The index of dispersion, ρ_i , was calculated for each deleterious mutation, i.e. $\bar{\rho}$ is the average across all deleterious mutations. We randomly placed the observed number of mutations into each locus, reflecting the null hypothesis that compensatory mutations are randomly distributed across all codon positions. A ratio greater than that observed in the randomization indicates that some amino acid residues are more likely to produce compensatory mutations than is expected by chance, whereas an index greater than the randomized value would indicate that mutations are more evenly distributed across all codons in the gene.

The index of dispersion averaged across all the taxa, $\bar{\rho} = 2.65$, was much larger and statistically significantly different from that observed in the randomization $\rho = 1.05$ ($p < 10^{-6}$). The index was significantly greater than expected by chance for each of the three kingdoms considered separately (eukaryotes: $\bar{\rho} = 2.65$, $p < 10^{-6}$; prokaryotes: $\bar{\rho} = 2.84$, $p < 10^{-6}$; viruses: $\bar{\rho} = 2.06$, $p < 10^{-6}$). These data demonstrate that multiple compensatory mutations occur at the same amino acid residue much more often than is expected by chance, across the three kingdoms surveyed.

The foregoing analysis shows that in response to a single deleterious mutation, some sites are more likely to evolve compensatory alleles. We can also ask whether there are any sites that are likely to compensate for more than one deleterious mutation. In our dataset, there are 11 proteins that have been studied with more than one deleterious mutation. Of these 11, five showed at least one site where a compensatory mutation evolved independently in response to distinct deleterious mutations. (The remaining six that did not show this pattern were among the loci which had the fewest compensatory mutations, therefore limiting the scope for multiple mutations.) We tested whether more proteins than expected by chance showed convergent evolution at compensatory sites in response to distinct deleterious mutations. To perform this test, we used the hypergeometric distribution to calculate the expected number of proteins in the dataset that would show no compensatory mutations in common for different deleterious mutations, under the null hypothesis that compensatory mutations are distributed equally through the protein sequence. The hypergeometric distribution describes the probability of getting a given number of sites that appear for one deleterious mutation when sampled without replacement from the possible sites that compensate for another deleterious mutation. We excluded any amino acid that was within 5 per cent of the total sequence length of both the deleterious mutations, because, as we show in the following section, this region contains an excess of compensatory mutations. From this analysis, we expect that on average 1.5 of the 11 proteins ought to show a compensatory mutation at the same site for more than one deleterious mutation just by chance. The observed value, 5 out of 11, is significantly more than expected by chance (binomial test, $p=0.01$).

(b) Question 2: are compensatory mutations near their associated deleterious mutations?

Given that some sites are more likely to produce compensatory mutations than others, we ask whether proximity to the deleterious mutation might explain some of this pattern. We quantified the degree of clustering of compensatory mutations around their associated deleterious mutations using the following scheme. We used d_i to represent the sequence location of the i th deleterious mutation and $c_{j,i}$ to represent the location of the j th compensatory mutation identified for that deleterious mutation. Thus, the absolute distance in the primary structure between the deleterious and compensatory mutations is given by $|c_{j,i} - d_i|$. The mean distance standardized by gene length L_i between the i th deleterious mutation and its N_i compensatory mutations is given by

$$\delta_i = \sum_{j=1}^{N_i} \frac{|c_{j,i} - d_i|}{(N_i L_i)}.$$

We use $\bar{\delta}$ to refer to the mean of δ_i . We compared $\bar{\delta}$ to its expectation under a random placement model by a simulation method. Positions of all compensatory mutations were drawn randomly from all sites in the gene (except the deleterious mutation site). For each simulation, $\bar{\delta}$ was recalculated and stored. After repeating this process 10^7 times, the observed test statistic was compared against the simulated null distribution.

Figure 2. The frequency distribution of the locations of compensatory mutations relative to deleterious mutations, expressed as a proportion of the gene length. Distance values less than zero indicate compensatory mutations that are upstream of the deleterious mutations, and distances greater than zero imply that the compensatory mutation is downstream of the deleterious mutation. The black line shows the expected distribution assuming random placement of the compensatory mutations. (The expected distribution declines away from the deleterious mutation, because deleterious mutations are not typically at the centre in the sequence of the gene.) The data show an excess of mutations near the deleterious mutation.

We found significant evidence of compensatory mutations clustering with respect to the position of their associated deleterious mutations (figure 2). Compensatory mutations were located at a mean standardized distance of $\bar{\delta}=0.228$, averaged over all deleterious mutations. By contrast, the null expectation of $\bar{\delta}$ was 0.321, and the ratio of observed versus expected was 0.710 ($p<10^{-6}$, for the test comparing this ratio with the null expectation of 1). For eukaryotes, $\bar{\delta}=0.202$, compared with a null expectation of $\bar{\delta}=0.29$ ($p=0.0023$), and the ratio of observed versus expected was 0.70. For prokaryotes, $\bar{\delta}=0.266$, compared with a null expectation of $\bar{\delta}=0.331$, and a ratio of observed to expected was 0.68 ($p=0.00041$). For viruses, $\bar{\delta}=0.194$, compared with a null expectation of $\bar{\delta}=0.311$ ($p<10^{-6}$), and the observed to expected ratio was 0.622. Thus, in all taxonomic groups considered, compensatory mutations tended to occur closer to the original deleterious mutation than expected by chance.

We also considered whether compensatory and deleterious mutations are closer together in the protein's tertiary structure than would be expected by chance. This was accomplished using published three-dimensional crystal structures that exist for 10 of the proteins used above. We measured the Euclidean distance in angstroms between the α -carbon of the deleterious and compensatory mutation sites, as reported in the three-dimensional structure files obtained from Research Collaboratory for Structural Bioinformatics at www.rcsb.org (Berman et al. 2000). We calculated the average distance by dividing the mean distance between the compensatory and its associated deleterious mutations by the average distance between the deleterious mutation and all the other amino acid residues in the protein. To test statistically for deviations between the observed relative distances and that expected by chance, the positions of the

compensatory mutations were then randomly relocated in the gene, and the average distance between the compensatory and deleterious mutations was recorded for the randomized data. This randomization process was recreated 100 000 times to generate a null distribution for the test statistic. For cases where there were more than one replicate evolutionary line with a given deleterious mutation, the data across lines were collated to give a combined p -value using the Z -transform test (Whitlock 2005). Out of the 22 deleterious mutations that had structural data available, seven showed strongly significant evidence that the compensatory mutations were closer to the deleterious mutation than expected by chance, and six of these remain significant following adjustment for false discovery rate (Benjamini & Hochberg 1995). Of these six, compensatory mutations were on average only 51 per cent of the expected distance in angstroms as expected by chance. In no case were compensatory mutations significantly farther from the deleterious mutation than expected by chance.

(c) Question 3: are compensatory mutations clumped within the gene?

The mean standardized nearest-neighbour distance in primary sequence between the N_i compensatory mutations having length L_i amino acid residues (associated with the i th deleterious mutation) was calculated as

$$\bar{\delta}_i^{nm} = \frac{\sum_j \min_j |\delta_{j,i} - \delta_{j \neq i,i}|}{N_i L_i},$$

where $\delta_{i,j}$ represents the position of compensatory mutation j for deleterious mutation i . We calculated the grand mean over all deleterious mutations to obtain our test statistic, which we denote as $\bar{\delta}^{nm}$. Given that we know compensatory mutations are likely to clump because of their increased probability of being near the deleterious mutation (see previous section), we statistically removed the effect of the location of the deleterious mutation through a two-step process. First, we excluded all compensatory mutations that lie within 5 per cent of the length of the gene from the deleterious site, because, as shown in figure 2, there is a large excess of compensatory mutations near the site of the deleterious mutation (most of the excess seemed to occur within 1–2 per cent of the distance of the compensatory mutation, but to be conservative we eliminated a larger range). This removed 25.8, 12.03, 22.1 and 40% of the compensatory mutations that appear in the immediate neighbourhood of the site of the deleterious mutation for the full dataset, the eukaryote dataset, the prokaryote dataset and the virus dataset, respectively. After removing the mutations in the immediate neighbourhood of the deleterious mutation, the probability of a compensatory mutation as a function of distance from the deleterious mutation is an approximately linear function of the proportional distance. We then divide the genes into bins representing 5 per cent of the total length of the gene, and we performed a linear regression of the absolute distance of the remaining compensatory mutations upon their probability of occurrence within each bin. This regression model was used to generate a random probability of compensatory mutation placement in the gene, accounting for the location of the deleterious mutation. By simulating this random model, we determined the null distribution for $\bar{\delta}^{nm}$.

The average standardized distance for the whole dataset is $\bar{\delta}^{nm} = 0.078$, which is statistically significantly different from the random expectation $\bar{\delta}_{\text{random}}^{nm} = 0.128$ ($p < 10^{-6}$), indicating that compensatory mutations cluster significantly around each other on average. Breaking the dataset down by kingdom into eukaryote, prokaryote and virus yields for eukaryotes $\bar{\delta}_{\text{euk}}^{nm} = 0.042$ (compared with $\bar{\delta}_{\text{euk,random}}^{nm} = 0.113$; $p < 10^{-6}$), prokaryotes $\bar{\delta}_{\text{prok}}^{nm} = 0.087$ ($\bar{\delta}_{\text{prok,random}}^{nm} = 0.130$; $p = 0.0026$) and viruses $\bar{\delta}_{\text{virus}}^{nm} = 0.059$ ($\bar{\delta}_{\text{virus,random}}^{nm} = 0.106$; $p = 0.0016$). This result shows that compensatory mutations cluster more closely to each other than would be expected by chance regardless of the taxonomic group considered.

3. DISCUSSION

Our analysis of the sequence of compensatory mutations has revealed several novel patterns (figure 1). Some sites are much more likely than others to successfully compensate for specific deleterious mutations; these sites are closer than expected to the deleterious site; and these compensatory sites are close to each other. These sequence spatial patterns have some important implications for evolutionary biology.

Multiple compensatory mutations occur at the same amino acid residue more often than is expected by chance. Given our understanding of the importance of protein structure on protein function, the functional relationships among amino acid residues within proteins are not surprising. Structural studies of proteins demonstrate that some amino acid residues are more important than others in affecting a protein's function. Consequently, we might expect that some amino acid positions are more likely to produce compensatory mutations than others. Indeed, we have seen that some amino acids sites are far more likely to evolve compensation than others, with variability among sites two to three times that expected by chance. Evolution has a high probability of convergence at the molecular level. The response to fixation of deleterious alleles is partially predictable.

Our results have some potentially important implications for building phylogenetic trees using molecular data. Most phylogenetic reconstruction methods assume independent evolution of each mutation in the tree. However, if compensatory evolution is common, then multiple, nearby mutations may give only highly correlated information.

Biochemical insights also predict that some nearby parts of proteins are likely to be involved in the same functions, for example in binding sites. As a result, we can predict that deleterious mutations are more likely to be compensated by nearby amino acids sites. For the data we have collated, we have shown that compensatory mutations occur approximately two-thirds as far away from the site of their associated deleterious mutations compared with the distance expected by chance. Moreover, compensatory mutations tend to occur closer to the site of the deleterious mutation in tertiary structure.

This clustering also has important implications for evolutionary biology. Recombination has been shown to be an important force affecting the frequencies of alleles interacting epistatically with each (Phillips & Johnson 1998), and compensatory mutations are an example of mutations interacting epistatically. Here, we have shown

that compensatory mutations tend to occur extremely close to the site of the deleterious mutation; on average, they occur within 22 per cent of the length of the gene around the site of the deleterious mutation, and Poon *et al.* (2005) have shown that compensation is much more likely to be intragenic than expected by chance. This clustering of compensatory mutations around a particular site within a gene means that recombination is unlikely to break apart, or recombine together, deleterious mutations with their compensatory counterpart.

Finally, biochemistry has informed us that different regions of a protein will probably perform specific functions (e.g. active sites, hydrophobic core, etc.) necessary for overall protein performance. For example, mutations that affect a protein's active site are unlikely to be compensated for by a secondary mutation that occurs within the hydrophobic core. This leads to the prediction that compensatory mutations ought to occur closer to each other in the primary and tertiary structures than would be expected by random chance. We have shown that compensatory mutations do tend to occur closer to their associated deleterious mutations than is expected by chance. Such clustering of compensatory mutations is expected because of the importance of local interactions affecting the overall shape of the protein (Chikenji *et al.* 2006), and these results reinforce the conclusions of phylogenetic analysis that show frequent coevolution of nearby amino acid residues (Pollock 2002; Wang & Pollock 2005; Castoe *et al.* 2008). We have confirmed this prediction in these data, finding that the nearest-neighbour distance between compensatory mutations is approximately 40 per cent lower than would be expected by chance. These strong patterns show that the path that evolution takes is influenced by the basic constraints of biochemistry in predictable yet important ways. Future work could examine the specific biochemical properties of compensatory mutations, to ask whether successful compensatory mutations are predictable from their biochemistry (see Poon & Chao 2006).

This paper has been much improved by comments from Sally Otto, Leithen M'Gonigle, Dolph Schluter, David Houle, Alex Kondrashov and an anonymous reviewer. This work was supported by a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (NSERC PGS-B) and a University of British Columbia Paetzold Fellowship to B.H.D. and an NSERC Discovery Grant to M.C.W.

REFERENCES

- Benjamini, Y. & Hochberg, Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. 2000 The protein data bank. *Nucleic Acids Res.* **28**, 235–242. (doi:10.1093/nar/28.1.235)
- Burch, C. B. & Chao, L. 1999 Evolution by small steps and rugged landscapes in the RNA virus $\phi 6$. *Genetics* **151**, 921–927.
- Castoe, T. A., Jiang, Z. J., Gu, W., Wang, Z. O. & Pollock, D. D. 2008 Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE* **3**, e2201. (doi:10.1371/journal.pone.0002201)
- Chikenji, G., Fujitsuka, Y. & Takada, S. 2006 Shaping up the protein folding funnel by local interactions: lesson from a structure prediction study. *Proc. Natl Acad. Sci. USA* **103**, 3141–3146. (doi:10.1073/pnas.0508195103)
- Crill, W. D., Wichman, H. A. & Bull, J. J. 2000 Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* **154**, 27–37.
- Estes, S. & Lynch, M. 2003 Rapid fitness recovery in mutationally degraded lines of *Caenorhabditis elegans*. *Evolution* **57**, 1022–1030. (doi:10.1554/0014-3820(2003)057[1022:RFRIMD]2.0.CO;2)
- Hartl, D. L. & Taubes, C. H. 1996 Compensatory nearly neutral mutations: selection without adaptation. *J. Theor. Biol.* **182**, 303–309. (doi:10.1006/jtbi.1996.0168)
- Kascier, H. & Burns, J. A. 1973 The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104.
- Kascier, H. & Burns, J. A. 1981 The molecular basis of dominance. *Genetics* **97**, 639–666.
- Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. 2002 Dobzhansky–Muller incompatibilities in protein evolution. *Proc. Natl Acad. Sci. USA* **99**, 14 878–14 883. (doi:10.1073/pnas.232565499)
- Maisiner-Patin, S. & Andersson, D. I. 2004 Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res. Microbiol.* **155**, 360–369. (doi:10.1016/j.resmic.2004.01.019)
- Moore, F. B. G., Rozen, D. E. & Lenski, R. E. 2000 Pervasive compensatory adaptation in *Escherichia coli*. *Proc. R. Soc. B* **267**, 515–522. (doi:10.1098/rspb.2000.1030)
- Phillips, P. C. & Johnston, N. A. 1998 The population genetics of synthetic lethals. *Genetics* **150**, 449–458.
- Pollock, D. D. 2002 Genomic biodiversity, phylogenetics, and coevolution in proteins. *Appl. Bioinformatics* **1**, 81–92.
- Poon, A. & Otto, S. P. 2000 Compensating for our load of mutations: freezing the mutational meltdown. *Evolution* **54**, 1467–1479.
- Poon, A. F. Y. & Chao, L. 2006 Functional limits on the fitness effect size of compensatory mutations in the bacteriophage ϕ X174. *Evolution* **60**, 2032–2043.
- Poon, A., Davis, B. H. & Chao, L. 2005 The coupon collector and the suppressor mutation: estimating the number of compensatory alleles by maximum likelihood. *Genetics* **170**, 1323–1332. (doi:10.1534/genetics.104.037259)
- Rokyta, D., Badgett, M. R., Molineux, I. J. & Bull, J. J. 2002 Experimental genomic evolution: extensive compensation for loss of DNA activity in a virus. *Mol. Biol. Evol.* **19**, 230–238.
- Schrag, S. J. & Perrot, V. 1996 Reducing antibiotic resistance. *Nature* **381**, 120–121. (doi:10.1038/381120b0)
- Wang, Z. O. & Pollock, D. D. 2005 Context dependence and coevolution among amino acid residues in proteins. *Methods Enzymol.* **395**, 779–790. (doi:10.1016/S0076-6879(05)95040-4)
- Whitlock, M. C. 2000 Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution* **54**, 1855–1861. (doi:10.1554/0014-3820(2000)054[1855:FONAAT]2.0.CO;2)
- Whitlock, M. C. 2005 Combining probability from independent tests: the weighted Z method is superior to Fisher's approach. *J. Evol. Biol.* **18**, 1368–1373. (doi:10.1111/j.1420-9101.2005.00917.x)
- Whitlock, M. C., Phillips, P. C., Moore, F. B. G. & Tonsor, S. 1995 Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**, 601–629. (doi:10.1146/annurev.es.26.110195.003125)
- Whitlock, M. C., Griswold, C. K. & Peters, A. D. 2003 Compensating for the meltdown: the critical effective size of a population with deleterious and compensatory mutations. *Annales Zoologici Fennici* **40**, 169–183.