PERSPECTIVE

The Importance of the Neutral Theory in 1968 and 50 Years On

Jeffrey D. Jensen[1], Bret A. Payseur[2], Wolfgang Stephan[3], Charles F. Aquadro[4], Michael Lynch[5], Deborah Charlesworth[6], and Brian Charlesworth[6]

[1]*School of Life Sciences, Arizona State University, Tempe, USA*

[2]*Laboratory of Genetics, University of Wisconsin-Madison, Madison, USA*

[3]*Leibniz-Institute for Evolution and Biodiversity Science, Berlin, Germany*

[4]*Department of Molecular Biology & Genetics, Cornell University, Ithaca, USA*

[5]*Center for Mechanisms of Evolution, Arizona State University, Tempe, USA*

[6]*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK*

corresponding author: jeffrey.d.jensen@asu.edu

ABSTRACT

A recent article reassessing the Neutral Theory of Molecular Evolution claimed that it is less important than is widely believed (Kern & Hahn 2018). Although this is probably a minority view, it is important to evaluate such a claim carefully in the context of current knowledge, as inaccuracies can sometimes morph into an accepted narrative for those not familiar with the underlying science. We here critically examine and ultimately reject Kern & Hahn's arguments and assessment.

The Neutral Theory of Molecular Evolution asserts that most *de novo* mutations are either sufficiently deleterious in their effects on fitness that they have little chance of becoming fixed in the population, or are under such weak selection that they may become fixed as a result of genetic drift (Kimura 1968, 1983; King & Jukes 1969; Ohta 1973). While drifting to fixation or loss, such mutations contribute to DNA sequence variation within populations. Furthermore, the evolutionary rate of neutral nucleotide substitutions between species should equal the mutation rate, and the proportion of the genome represented by selectively constrained sites (where mutations have low probabilities of fixation by drift) will co-vary with the effective population size of the species. Finally, under the Neutral Theory, advantageous mutations are hypothesized to be uncommon enough, compared to the constant input of neutral and deleterious variants, that they are rarely present in samples of segregating variants, especially because of their rapid spread to fixation.

These ideas greatly changed the thinking of evolutionary biologists. Genetic drift was taken much more seriously than previously, stimulating a large body of fruitful empirical research into molecular evolution and variation, as well as fundamental advances in the stochastic theory of evolution, summarized by Kimura's influential book (Kimura 1983). It is now difficult to appreciate how radical a departure this view of evolution represented: in the 1950s and 1960s, almost all evolutionary changes were attributed to directional natural selection, and most polymorphisms with alleles at intermediate frequencies were thought to be caused by balancing selection. R.A. Fisher, for example, famously rejected any significant evolutionary role for genetic drift (Fisher 1930), though it is notable that S. Wright simultaneously developed a deep appreciation for the importance of these stochastic effects that was later justified when molecular variants began to be studied (Wright 1931).

It is against this backdrop that Kern & Hahn (2018) discuss a purported controversy in population genetics concerning the predictive power and applicability of the Neutral Theory, beginning with the suggestion that "the ubiquity of adaptive variation both within and between species means that a more comprehensive theory of molecular evolution must be sought." Although those who supported the Neutral Theory, including Kimura, did not claim that all sequence changes are neutral — indeed, Kimura himself developed some of the most fundamental theoretical formulations of selection and its interactions with genetic drift — Kern & Hahn (2018) conclude that modern data have demolished the original evidence supporting the Neutral Theory. Below, we briefly re-state their claims and then examine their validity.

Kern & Hahn's arguments rely heavily on population-genomic studies suggesting that as many as 50% of amino-acid replacement substitutions in *Drosophila* are adaptive (see, for example, the review by Sella *et al*. 2009), apparently contradicting Kimura's (1968) and King & Jukes' (1969) claim that most substitutions result from genetic drift. Apart from the inherent uncertainty in these estimates (discussed by Fay 2011), using this observation to make the general claim that the Neutral Theory is insufficient to explain genome-wide patterns of variation and evolution is misleading, as these inferred frequencies of adaptive substitutions mostly relate to only the small fraction of the genome that codes for proteins (*e.g.,* < 2% of the human genome; see Lander *et al.* 2001). Kern & Hahn further overstate the pervasiveness of adaptive substitutions by highlighting studies in humans and plants that focus on the limited subset of genes that evolve rapidly. In 'Molecular Population Genetics', Hahn (2018; pg.21) states that "even if the vast majority of substitutions across a genome are neutral, what most researchers care about are substitutions that could have an effect on fitness — either those in coding regions or those in functional noncoding regions. And for these types of substitutions the neutral theory continues to be rejected." This circularity of ignoring the "vast majority" of neutral or nearly neutral substitutions, and then rejecting a significant role for neutrality, hardly justifies the need

for the "selection theory of molecular evolution" advocated by Hahn (2008).

Even if one focuses solely on coding regions, Kern and Hahn's assertions remain poorly supported. As noted by Cutter & Payseur (2013), more than 95% of amino-acid replacement sites in the multi-species comparisons in the genus *Drosophila* discussed above remain conserved. In addition, replacement sites that do evolve mostly do so at a fraction of the rate of selectively unconstrained sites (Halligan *et al*. 2013). Furthermore, with a high level of selective constraint, the few amino acids that become fixed through the action of positive selection will do so alongside others that are fixed under genetic drift, making it difficult to detect any deviation from a constant substitution rate. Thus, while there is good evidence for positive selection acting on coding and some classes of noncoding sequences (*e.g.,* Andolfatto 2005; Haddrill *et al.* 2008), there is ample evidence for recurrent, and overwhelmingly more frequent, purifying selection in these parts of the genome.

Concerning within-species variability, Kern & Hahn (2018) emphasize the well-established positive correlation between recombination rates and levels of variation. However, they make the unsupported claim that "these results imply that almost no loci are free from the effects of selection, in any organism." The correlation, first documented by Begun & Aquadro (1992) in *Drosophila melanogaster* and subsequently observed in many species (Cutter & Payseur 2013), indeed suggests that selection reduces neutral variability at linked sites. Purifying selection plays a central role in explaining this pattern: neutral variants linked to deleterious mutations are routinely eliminated through background selection (Charlesworth *et al.* 1993; Charlesworth 2012), without affecting neutral fixation probabilities. In explicit comparisons between models of widespread purifying selection on weakly deleterious alleles versus recurrent positive selection on beneficial alleles, Lohmueller *et al*. (2011) found a much better fit of the former to the observed pattern in humans, as did Comeron (2014) for *Drosophila* (reviewed in Charlesworth & Charlesworth 2018). In addition, observations from organisms with large genomes, including humans and mice, show that levels of polymorphism are low in the neighbourhood of coding or conserved non-coding sequences and increase approximately monotonically away from such sequences (see Cutter & Payseur 2013, as well as the recent results of Johri *et al*. 2017 and Lynch *et al*. 2017). While selective sweeps may well contribute to this pattern, and are also required to explain other observations, such as lower rates of protein sequence evolution in genes with high silent-site diversity (Campos *et al*. 2017), these findings imply that any selective sweeps involved must have rather local effects. It follows that many parts of genomes are little affected by them. Given these results, as well as the commonness of purifying relative to positive selection noted above, background selection must be an important contributor to the patterns of sequence diversity in genomic regions with different recombination rates. Furthermore, the effect of background selection in reducing diversity only serves to increase the number of genomic mutations that will behave as effectively neutral, as it increases the role of chance in evolution.

Despite this substantial evidence to the contrary, Kern & Hahn (2018) arbitrarily reject background selection as the primary driver of the effects of selection at linked sites. Describing background selection as an alternative to the Neutral Theory, Hahn's (2018; pg.24) book quickly summarizes it as being unable to adequately explain observed patterns, citing Stephan (2010) for this result. Curiously, Stephan's paper is a review of the historical development of these models, and proposes background selection as a globally sufficient explanation, followed by a discussion of recent progress towards co-estimating selective sweeps alongside it (*e.g.,* Kim & Stephan 2000). Hahn (2018) then introduces a hitchhiking model involving selective sweeps as another alternative to the Neutral Theory, simply describing it as being "consistent with the high rates of adaptive substitution observed in many species." Kern & Hahn (2018) similarly suggest that the supply of beneficial mutations is not "a major limiting factor over evolutionary time", citing studies that invoke pervasive positive selection to explain genome-wide patterns of variation (including Garud *et al.* 2015 and Schrider & Kern 2017). While it may be correct that the supply of beneficial mutations does not limit the rate of adaptive evolution, it does not follow that adaptive mutations can explain all patterns of variability across the genome. Moreover, the role of adaptive mutations must be evaluated with caution, as studies where their action has been inferred often fail to exclude or take proper account of the effects of the (unknown) non-equilibrium demographic histories of the populations in question. A recent study has demonstrated that the statistical methodologies underlying these claims falsely detect positive selection under a wide range of neutral scenarios (Harris R, Sackman A, & Jensen JD, unpublished data).

Given the above considerations, we here propose a simple interpretation of the existing evidence in terms of a modern version of the Neutral Theory, whose individual components should not be controversial:

1) A large fraction of the genome of organisms studied to date is subject to mutations that are effectively neutral with respect to fitness, and hence evolve under genetic drift.

2) The great majority of newly arising mutations that do affect fitness (*i.e.*, non-neutral mutations) are deleterious, and the major form of natural selection is purifying selection, removing these deleterious mutations from populations.

3) Natural populations are rarely at demographic equilibrium, and commonly have undergone recent historical changes. The combined effects of population-size change, structure, and migration all further shape patterns of variability in genomes. These demographic histories cannot be assumed to affect genomic patterns of variation uniformly, and indeed may produce different effects in different genomic regions, mimicking expectations under selection (*e.g.*, Wall *et al*. 2002; Thornton & Jensen 2007).

4) A combination of genetic drift (as modulated by the demographic history of the population) with both direct and linked purifying selection shapes patterns of genomic variation. Thus, a model taking joint account of all of these effects is essential for genomic analysis (Comeron 2017), and progress is being made towards this goal (*e.g.,* Zeng & Charlesworth 2010).

5) Beneficial mutations occasionally arise and may reach fixation or high frequencies, and convincing, localized hitchhiking effects related to such events have been described in a variety of organisms. In some cases, these genotypic changes have been meaningfully connected with both phenotype and fitness. However, the effects of these comparatively rare, localized hitchhiking events may be best-characterized and quantified as effects additional to the genome-wide, common processes described above. In the absence of this appropriate null model accounting for these common processes, inappropriate adaptive story-telling will be likely to proliferate.

All of our above points are fully consistent with the ground-breaking work of Kimura. Furthermore, developments made in light of empirical observations subsequent to his initial publication are straightforward extensions of the Neutral Theory, demonstrating its continued importance, not a demolition of it. Over the past five decades such extended insights have enhanced our understanding of the interplay of population size with drift-selection dynamics (Ohta 1973), and described the hitchhiking effects of selection induced by the comparatively rare class of beneficial mutations (Maynard Smith & Haigh 1974), as well as those induced by the common class of deleterious mutations (Charlesworth *et al*. 1993). Moreover, our use of the term "ground-breaking" to describe the Neutral Theory is not meant to imply a scientific advance that was fully formed at the outset. Other major scientific advances have similarly been adjusted and modified over time in light of later observations and thought. For example, Darwin's findings and reasoning supporting the operation of natural selection were not abandoned owing to his lack of a satisfactory theory of heredity — indeed, the incorporation of that subsequent knowledge only strengthened the underlying concepts.

In sum, the transition to molecular biology has increased the importance of population genetics, and this influx of data has lent support to many pre-genomic theoretical developments. Though the edifice may not yet be complete, those calling for the demolition and wholesale replacement of fundamental principles ought first to recognize the underlying support for, and insights arising from, these foundational ideas. The Neutral Theory changed how people thought about evolution at the molecular level, and great credit is owed to the scientists who worked it out in detail and understood what it could tell us once genes (and genomes) could be sequenced.

CITATIONS

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature* 437: 1149-52.

Begun, D.J. and C.F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlated with recombination rates in *D. melanogaster*. *Nature* 356: 519-20.

Campos, J.L., L. Zhao, and B. Charlesworth. 2017. Estimating the parameters of background selection and selective sweeps in Drosophila in the presence of gene conversion. *Proc Natl Acad Sci USA* 114: E4762-71.

Charlesworth, B., M.T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289-1303.

Charlesworth, B. 2012. The effects of deleterious mutations on evolution at linked sites. Genetics *190*: 5-22.

Charlesworth, B., and D. Charlesworth. 2018. Neutral variation in the context of selection. *Mol Biol Evol* 35: 1359-61.

Comeron, J.M. 2014. Background selection as baseline for nucleotide variation across the Drosophila genome. *PLoS Genet* 10(6):e1004434.

Comeron, J.M. 2017. Background selection as null hypothesis in population genomics: insights and challenges from Drosophila studies. *Philos Trans R Soc Lond B Biol Sci* 372(1736).

Cutter, A.D. and B.A. Payseur. 2013. Genomics signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* 14(4): 262-74.

Fay, J.C. 2011. Weighing the evidence for adaptation at the molecular level. *Trends Genet* 27(9): 343-9.

Garud, N., P. Messer, E. Buzbas, and D. Petrov. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet* 11(2):e1005004.

Haddrill, P.R., D. Bachtrog, and P. Andolfatto. 2008. Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol Biol Evol* 25: 1825-34.

Hahn, M.W. 2008. Toward a selection theory of molecular evolution. *Evolution* 62: 255-65.

Hahn, M.W. 2018. *Molecular Population Genetics*. Sinauer Associates.

Halligan, D., A. Kousathanas, R. Ness, B. Harr, L. Eory, T. Keane, D. Adams, and P. Keightley. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet* 9(12):e1003995.

Harris, R.B., A. Sackman, and J.D. Jensen. 2018. On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. In revision.

Fisher, R.A. 1930. *The genetical theory of natural selection.* Clarendon Press, Oxford.

Johri, P., S. Krenek, G.K. Marinov, T.G. Doak, T.U. Berendonk, and M. Lynch. 2017. Population genomics of paramecium species. *Mol Biol Evol* 34: 1194-1216.

Kern, A.D. and M.W. Hahn. 2018. The neutral theory in light of natural selection. *Mol Biol Evol* 35: 1366-71.

Kim, Y., and W. Stephan. 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* 155: 1415-27.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624-6.

Kimura, M., 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

King, J.L., and T.H. Jukes. 1969. Non-Darwinian evolution. *Science* 164: 788-98.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M. Zody, J. Baldwin, *et al*. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lohmueller, K.E., A. Albrechtsen, Y. Li, S. Kim, T. Korneliussen, N. Vinckenbosch, G. Tian, E. Huerta-Sanchez, A. Feder, N. Grarup, T. Jorgensen, T. Jiang, D. Witte, A. Sandbaek, I. Hellmann, T. Lauritzen, T. Hansen, O. Pedersen, J. Wang, and R. Nielsen. 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet* 7(10):e1002326.

Lynch, M., R. Gutenkunst, M. Ackerman, K. Spitze, Z. Ye, T. Maruki, and Z. Jia. 2017. Population genomics of *Daphnia pulex*. *Genetics* 206: 315-32.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* 23: 23-35.

Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96-8.

Schrider, D.R. and A.D. Kern. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol Biol Evol* 34: 1863-77.

Sella, G., D. Petrov, M. Przeworski, and P. Andolfatto. 2009. Pervasive natural selection in the Drosophila genome? *PLoS Genet* 5(6):e1000495.

Stephan, W. 2010. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos Trans R Soc Lond B Biol Sci* 365(1544): 1245-53.

Thornton, K.R. and J.D. Jensen. 2007. Controlling the false-positive rate in multi-locus genome scans for selection. *Genetics* 175: 737-50.

Wall, J.D., P. Andolfatto, and M. Przeworski. 2002. Testing models of selection and demography in *Drosophila simulans*. *Genetics* 162: 203-16.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.

Zeng, K. and B. Charlesworth. 2010. The effects of demography and linkage on the estimation of selection and mutation parameters. *Genetics* 186: 1411-24.