Applying ecological models to communities of genetic elements: the case of Neutral Theory

Stefan Linquist[1*], Karl Cottenie[2], Tyler A. Elliott[2], Brent Saylor[2], Stefan C. Kremer[3] and T. Ryan Gregory[2]

1. Department of Philosophy, University of Guelph, Guelph, Ontario N1G 2W1 Canada

2. Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1 Canada

3. School of Computer Science, University of Guelph, Guelph, Ontario N1G 2W1 Canada

*Correspondence:

Stefan Linquist

348 MacKinnon. University of Guelph. Ontario, Canada. N1G2W1.

linquist@uoguelph.ca

Published Reference: Linquist, S. Cottenie, K. Elliott, T.A. Saylor, B. Kremer, S.C. Gregory, T.R. (2015). Applying ecological models to communities of genetic elements: the case of Neutral Theory. *Molecular Ecology*, 24(13): 3232-3242

Keywords: ecological neutral theory, genome ecology, molecular neutral theory, mutation hazard model, transposon ecology.

Running title: Neutral Models of Genomic Communities

Abstract

A promising recent development in molecular biology involves viewing the genome as a mini-ecosystem, where genetic elements are compared to organisms and the surrounding cellular and genomic structures are regarded as the local environment. Here we critically evaluate the prospects of Ecological Neutral Theory (ENT), a popular model in ecology, as it applies at the genomic level. This assessment requires an overview of the controversy surrounding neutral models in community ecology. In particular, we discuss the limitations of using ENT both as an explanation of community dynamics and as a null hypothesis. We then analyze a case study in which ENT has been applied to genomic data. Our central finding is that genetic elements do not conform to the requirements of ENT once its assumptions and limitations are made explicit. We further compare this genome-level application of ENT to two other, more familiar approaches in genomics that rely on neutral mechanisms: Kimura's Molecular Neutral Theory and Lynch's Mutational Hazard Model. Interestingly, this comparison reveals that there are two distinct concepts of neutrality associated with these models which we dub 'fitness-neutrality' and 'competitive neutrality'. This distinction helps to clarify the various roles for neutral models in genomics, for example, in explaining the evolution of genome size.

1. Introduction

There is growing enthusiasm in the fields of molecular biology and genomics for the prospect of an ecological perspective on the genome (Brookfield 2005; Le Rouzic et al. 2007; Venner et al. 2009; Linquist et al. 2013). The central idea is that genomes can be viewed as ecosystems in miniature, where particular DNA sequences are compared to organisms and the surrounding genetic and cellular structures are regarded as the local environment. Concepts and models from ecology might then be used to document and explain differences in the abundance and distribution of genetic elements among genomes. This approach has been especially promising for understanding the dynamics of transposable elements (TEs) – sequences of DNA that are capable of movement and replication within genomes (Brookfield 2005, Linquist et al. 2013). However, it is becoming clear that this approach requires careful attention to how one applies ecological concepts and models at the genomic level. For instance, it is important to be explicit about the focal entity under investigation: which types of genetic element are being identified as the organism-like entity, and what are the relevant "environmental" factors with which they interact? It is also helpful to distinguish ecological from evolutionary influences at the genomic level, something that has not been done with sufficient clarity in previous studies (Linquist et al. 2013). Detecting purely *ecological* influences at the genomic level involves testing whether local environmental factors (e.g., chromatin structure, GC content) co-vary with genetic elements of a given type, independent of the evolutionary relatedness of the elements in question. Elsewhere, we report that ecological factors do in fact influence the composition of transposable elements within the *Bos taurus* genome (Saylor et al. 2013). It is yet to be determined whether such intra-genomic ecological factors are operative within the genomes of other species.

Of particular interest in the development of ecological approaches to the genome is whether

neutral ecological models might be employed at the genomic level. Recent years have seen growing interest in neutral ecological models as they are applied to communities of whole organisms (Hubbell 1997, 2001; Rosindell et al. 2010, 2011, 2012). In that context, a focal point of controversy has involved the question of when (if ever) neutral processes can be inferred from neutral patterns: i.e. does conformity with the predictions of a neutral model imply that a community is governed primarily by neutral processes? Within ecology, there has been a growing sense of caution surrounding this issue (Cottenie 2005, Dornelas et al. 2006, McGill et al. 2006, Griffiths 2010, Clark 2012). In what follows, we review some of the arguments calling for a tempered use of neutral models in ecology and consider their implications for an ecological approach to the genome.

Notwithstanding potential uses of ecology at the genome level, it is important to note that neutral models are controversial within ecology proper. As we argue, applications of neutral models encounter problems when they: (1) neglect known ecological differences among the kinds of units being investigated (e.g., species or TE family membership), (2) ignore matters of scale, or (3) rely on questionable assumptions about parsimony in order to infer process from pattern. In what follows, we illustrate how these problems emerge in a previous attempt to apply ecological neutral models to the genome (Serra et al. 2013). We also explore how these considerations may be useful for understanding some of the controversy surrounding another recent model of genome evolution (Lynch and Conery 2003; Lynch 2007, Lynch et al. 2011), even though it is not explicitly neutralist.

2. Neutral Theory in Community Ecology

The Unified Neutral Theory of Biodiversity was developed by ecologist Stephen Hubbell (1997,

2001) as a challenge to conventional wisdom in community ecology. Prior to this, community ecologists typically endeavoured to explain the relative abundance and diversity of species within a community in terms of their suitability to available niches (Lack 1947; MacArthur 1961; Mayr and Diamond 2001). In stark contrast, Hubbell's neutral theory – also called Ecological Neutral Theory (ENT) – proposes that species-specific differences are irrelevant to population growth rates and their effect on abundance and diversity. Instead, he argued that a handful of "neutral" processes (including birth, death, dispersal, and speciation) are sufficient to explain abundance and diversity within most communities.

At least ten different neutral ecological models have since been developed to explore these ideas (reviewed by McGill et al. 2006). All of them share two core assumptions. First, they assume that all individuals existing within the same trophic level can be assigned equal likelihood of reproduction and death, regardless of species membership – or, what is also sometimes described as "equi-probable success". Second, these models imagine a metacommunity structure wherein multiple local communities are connected by dispersal. In addition to these two core assumptions are two auxiliary assumptions, specific to some but not all neutral models. Often, model-communities are assumed to have fixed carrying capacity. This generates a "zero-sum" dynamic in which individuals compete for a limiting resource. Hubbell (2001) originally took the zero-sum assumption to be an essential feature of neutral models, but subsequent models generate similar dynamics while allowing, for example, for constant positive population growth (e.g.Volkov et al. 2003). A second auxiliary assumption is that species have limited dispersal distances. This assumption can create spatial autocorrelation among species, a ubiquitous aspect of biological communities (Fortin and Dale 2005).

Historically, assessments of neutral ecological models have made extensive use of rank-

abundance curves (Hubbell 2001; Rosindell et al. 2010). These graphs plot the log abundance of each species in a community against the rank abundance of all species within that community. The curvature of the resulting graph is then taken to provide a basis for comparing a model's predictions with empirical data. McGill et al. (2006) noted that neutral models generate a hyperbolic (hollow curved) distribution known as a zero-sum multimodal distribution (ZSM). Most known communities approximate the ZSM. This apparent predictive success initially led Hubbell and colleagues to infer that neutral processes predominate in nature and that species-specific effects are therefore comparatively weak (Hubbell 1997, 2001; Rosindell et al. 2010).

However, this inference from pattern to process has generated considerable critical discussion. One set of problems stems from the use of the ZSM as a test of neutral models. Most notably, there is a wide range of non-neutral conditions under which a community might exhibit a hollow-shaped curve. Hence, this distribution does not reliably indicate that species-specific effects are absent. More generally, a variety of non-neutral models produce identical patterns to neutral models. Marquet et al. (2003) identified more than 30 such cases where non-neutral models generated equivalent predictions to neutral models. Importantly, these included many plausible ecological scenarios.

A different problem with the usage of ZSM to test neutral-model predictions is the low information content present in the data. Since all that is needed are species abundances, these curves are relatively easy to construct. However, there are more data-intensive tests of ENT, such as comparing the variation explained by proxies for niche versus neutral processes in a multivariate context (Cottenie 2005). Using this approach in an analysis of 158 datasets, Cottenie (2005) found that neutral processes were only evident in 8% of the data sets.

In the face of these criticisms, some proponents of ENT have suggested that neutral models can nonetheless play an important role as null hypotheses. This is a logically weaker proposal than Hubbell's original suggestion: conformity with neutral-model predictions is no longer taken to reveal that a community is shaped primarily by neutral processes. Instead, this approach assumes neutrality as the default condition, placing the burden of proof on advocates of niche-based models to demonstrate otherwise.

A fundamental requirement of hypothesis testing, however, is that H1 and H0 make distinct (non-overlapping) predictions. The philosopher John Beatty (1987) raised this issue over two decades ago in the context of molecular neutral theory: since changes in allele or trait frequency can be explained either by selectionist or by neutral models, he argued, neutral theory cannot serve as an adequate null hypothesis for testing adaptationist hypotheses. Yet some ecologists persist in defending this process-from-pattern use of ENT despite the problem of overlap. Their argument is that neutral models are more *parsimonious* than niche-based models (Rosindell et al. 2011; 2012). If two competing hypotheses have equal explanatory value, it is argued, the simpler one should prevail.

But this argument is problematic for several reasons. First, there is no single definition of parsimony: a hypothesis that seems relatively parsimonious from one perspective appears convoluted in other respects (Sober 1994). For example, neutral models seem parsimonious because they do not discriminate among species. However, they generally contain more tunable parameters than many niche-based models, making them relatively difficult to fit to data (Nee and Stone 2003). A further problem is that the parsimony argument prejudges the issue at hand. Placing the burden of proof on non-neutral hypotheses assumes that neutral processes are more influential in nature, and this is precisely what ENT aims to test.

In addition, it might be argued that the parsimony argument is entirely irrelevant to ENT because these types of models offer a poor explanation of community dynamics. Even under ideal circumstances, parsimony should be used only to decide among competing hypotheses that are otherwise *equivalent* in their explanatory success. Discarding a more explanatory hypothesis for one that is just simpler would run contrary to the basic aims of science. This raises the question of whether neutral models, considered as explanations, are on the same footing as their niche-based rivals. Some critics argue that ENT fails to describe the mechanisms driving community dynamics. According to Clark (2012), although ENT assumes that all species have the same likelihood of success, it is mathematically equivalent to a model in which species differ in success rates, but niches are encountered randomly by individuals. According to Clark these are two very different sorts of causal process and ENT fails to discriminate among them. Clark (2012) argues that ENT does not in fact model the *absence* of species differences; rather, it models the *ignorance* of species differences. When viewed in this light, perhaps ENT should not be lauded for its (alleged) parsimony so long as its explanatory status is in question. In our view, it remains a complex question as to whether neutral models are in fact explanatory and, indeed, whether they describe causal mechanisms at some level of abstraction (cf. Craver 2006). The main point is that, even if the other problems with the parsimony argument could be resolved (i.e. if a definition for "parsimony" could be agreed upon and the bias against niche-based models could be somehow compensated for), the matter would still not be settled: it remains a further question as to whether neutral models have the same explanatory power as their more mechanistically detailed rivals.

It bears mentioning that despite these obstacles, ENT remains a popular framework within the discipline of ecology. In some cases, neutral models have persisted in an even weaker

form. For example, Wennekes et al. (2012) appeal to the philosophical doctrine of instrumentalism, which holds that scientific theories, generally, are mere tools for prediction and control of nature that cannot, in principle, explain the structure of underlying processes. This is a considerable concession given Hubbell's original ambition, which was to uncover the processes driving species abundance and diversity in communities (Hubbell 1997, 2001). Other ecologists continue to use neutral models both as an approximation of observational data and as a null hypothesis, despite the well documented problems in doing so (see Clark 2012 for discussion). More constructively, McGill et al. (2006) have developed a rigorous framework for hypothesis testing in community ecology. They concur that neutral models are uninformative as null hypotheses. Instead, they propose a model-comparison framework in which realistic alternative models are ranked (as opposed to simply rejected or accepted) according to their degree of fit with data. The authors conclude that, according to this more rigorous framework, neutral models receive little empirical support.

In sum, the past decade has seen steady theoretical progress surrounding the application of neutral models in ecology. The trend is a gradual weakening of neutral models in their logical status from good approximations, to useful null hypotheses, to occasional predictive devices. Perhaps the most important lesson to draw from this debate is that neutral-model predictions by themselves reveal little about the nature of ecological processes. The plausibility of a neutralist explanation depends not just on the predictive success of a neutral model, but also on how well the community conforms with the model's assumptions. For example, if a community exhibits a ZSM distribution but it is known that species are very likely to be trophically different, then neutral-process interpretations should receive low prior probability.

3. **Ecological Neutral Theory at the Genomic Level**

In a recent paper, Serra et al. (2013) made the first explicit attempt to apply ecological neutral models at the genomic level. In that case, distinct categories of element (retrotransposons, DNA transposons, satellites, simple repeats, tRNA, miRNA, RNA, and genes) were identified as different 'genetic species'. Each chromosome was viewed as a local community with the nuclear genome at large serving as the metacommunity. Model simulations were then compared with whole genome data from seventeen species' genomes. The abundances of genetic species exhibited the standard ZSM distribution predicted by neutral models, which Serra et al. (2013) then took to suggest that an underlying neutral process could be the cause: "We are certainly aware that the fit of a neutral pattern does not necessarily imply the existence of a neutral process behind the pattern, but it does offer the simplest explanation consistent with current data" (2013 p. 4).

It is helpful to consider this genome-level application of ENT in light of the theoretical developments that have taken place within ecology. As noted, considerations of parsimony (ie. simplicity) alone do not licence the inference from neutral pattern to neutral process. Instead, one must consider whether neutral-model assumptions are biologically plausible as they pertain to the system under study. In this regard, it is useful to assess whether the core and auxiliary assumptions that are characteristic of (most) neutral ecological models are satisfied by the genomic 'species' analyzed by Serra et al. (2013).

1. *Equi-probable success among 'species'*. A particularly unrealistic assumption is that genetic elements have an equal likelihood of success regardless of their 'species' identity. This assumption is explicitly violated if one compares active transposons, whose copy rate is typically higher than average, to non-mobile elements. Even at this coarse-grained level, where mobile elements are lumped into a single category, it is a mistake to treat

them on par with non-mobile elements. Such differences continue to emerge at even finer levels of grain. Within the category of mobile elements there is an important functional distinction between DNA transposons, that employ a cut-and-paste mode of replication, and retrotransposons which employ a copy-and-paste strategy. Each mode of replication is associated with different genetic and cellular requirements that can influence an element's replication rate and mobility (Havecker et al. 2004, Feschotte and Pritham 2007, Han 2010). Even within each of these functional categories, particular families of TE are known to vary, for example, in their preference for certain genomic regions or 'genomic habitats' (Zou et al. 1995, Eickbush 2002, Neumann et al. 2011, Pardue and DeBaryshe 2011, Elliott et al. 2013, Kojima and Jurka 2013). Thus, some TEs appear to select genomic regions that are less vulnerable to removal by negative selection at the host level. Likewise among non-mobile DNA, there are differences in mutation rate among different regions. Protein-coding genes also experience differential magnitudes of positive or negative selection, especially as compared to many non-coding regions of the genome. It therefore seems methodologically egregious to lump these functional genes together in the same category with various forms of non-coding and mobile DNA, given that the mechanisms governing their replication and assortment within the genome are known to differ so substantially.

Perhaps this point is best illustrated with an example from ecology where known differences among species are also sometimes ignored. Ecologists Rachata Muneepeerakul and colleagues (2008) developed a neutral model to predict species richness patterns in fishes across the entire Mississippi delta. The model, which assumed no differences among species in suitability to particular habitats, successfully predicted

abundance curves across the region. However, as the authors were careful to note: "it is crucial to recognize that neutral pattern does not imply neutral process" (Muneepeerakul et al. 2008 p. 222). A noteworthy limitation of this study is that different fish species are known not to have equivalent dispersal rates (Griffiths 2010). Species also varied in their modes of subsistence (piscivores, planktivores, benthivores, etc.) as well as other ecologically relevant factors. It is also significant that the study was conducted at a very coarse grain of spatial resolution, such that species-specific interactions tend to get overwhelmed by environmental heterogeneity (Olden et al. 2010). Our point is that such known differences among species undermine the accuracy and suitability of neutral models. The same point applies at the genomic level, where functional differences among genetic element 'species' are perhaps even more pronounced.

2. *Metacommunity structure.* At first glance, it appears that chromosomes should serve as fairly discrete and well defined genomic communities. However, in order for multiple chromosomes to be viewed as a metacommunity, movement of elements within individual chromosomes must exceed the migration rate of elements among chromosomes. Some categories of element are likely to satisfy this requirement better than others. For instance, protein-coding genes, whose translocation depends largely on homologous recombination, are more mobile within than among chromosomes. Transposable elements are less likely to observe chromosomal boundaries, but even here some element families may be more highly localized than others. Such inconsistencies in intra- and inter-chromosomal mobility raise issues of identifying appropriate spatial scales for analysis. For an element with restricted dispersal ability, an individual chromosome might indeed qualify as the local community, but for a more nomadic type

of element, the community may encompass the entire genome. It would be misleading in such cases to use a single ecological neutral model to describe the behavior of a wide array of genetic elements since they do not belong to the same metacommunity.

3. *Zero sum dynamics and fixed carrying capacity*. In traditional ecological communities, organisms at a given trophic level often compete for limited resources. Most neutral ecological models represent competition by imposing a fixed carrying capacity on these communities, such that a new individual cannot migrate until an available spot becomes vacant. It is questionable whether genomic communities exhibit this dynamic. In some species there appear to be fitness benefits associated with reduced genome size (e.g., hummingbirds; Gregory et al. 2009). Selection acting on the whole organism that limits genome size might effectively create a low carrying capacity for genomic elements. But this limitation is unlikely to generalize across all species – a simple fact that is demonstrated by the enormous diversity in genome size across animals and plants (Gregory 2005a; Bennett and Leitch 2005). Nor is it clear which other factor(s) besides physical space on the chromosome might serve as the genomic equivalent of a limiting resource (access to enzymes needed for transposition or availability of raw nucleotides may be plausible candidates).

Moreover, some types of element (e.g. retrotransposons) actively incorporate themselves into a chromosome and thus increase genome size directly. Those newly established regions then become suitable 'habitat' for the colonization of additional TEs. This might be regarded as an interesting example of *ecosystem engineering* that occurs at the genomic level (cf. Wright & Jones 2006, Pearce 2011). Viewed from the perspective of ENT, however, the possibility that TEs are not resource limited in most eukaryotic

genomes implies that they violate yet another assumption common to most neutral models.

4. *Limited dispersal.* The extent to which genetic elements migrate among chromosomes is highly dependent on the genomic 'species' in question. Transposable elements are much more mobile than other categories of genetic elements. Given the broad distribution of at least some elements within particular genomes, it is clear that not all TEs are limited in their capacity to disperse throughout a chromosomal metacommunity. Consider, for example, the hyper-abundant *Alu* element, which is present in more than 1 million copies in the human genome and is widely (though not uniformly) distributed across the human karyotype (Bolzer et al. 2005). As noted, non-mobile elements (e.g., protein-coding genes) are more likely to satisfy this assumption than most transposons, but this merely reinforces the problems with treating elements of very different types in the same way within a single neutral model.

In sum, the attempt by Serra et al. (2013) to apply ENT at the level of the genome exposes some challenges facing neutral models more broadly. We have argued that mere conformity with ENT's broad-scale predictions (e.g. the ZSM curve) is an inadequate basis for inferring process from pattern. Nor is this inference justified on grounds of parsimony. Instead, one must pay close attention to the assumptions of ENT to determine whether they are violated by the genomic systems in question. This task can be especially challenging in the case of genome-level ecology, where the relevant entities and their environments are not well defined in comparison to those typically addressed by community ecologists. In this section, we have argued that the assumptions of ENT are often violated by the known properties of TEs. Given that transposable elements are a major component of most eukaryotic genomes (Gregory 2005b), it seems doubtful

that neutral models will serve as a useful tool for many genome-level ecologists.

4. **Molecular Neutral Theory**

Our finding in the previous section – that neutral *ecological* models may be inappropriately applied to entire genomes – raises important questions about other models in genetics that emphasize the role of neutral processes. In fact, ENT was partly inspired by the success of molecular neutral theory (Hubbell 2001). This raises the question of how these two theories differ.

As in ecology, neutral models in molecular genetics first emerged as something of a challenge to prevailing orthodoxy. A common view prior to the publication of the influential work by Kimura (1968, 1969, 1983) held that the properties and components of genomes are shaped primarily by natural selection. For example, it was often assumed that alleles would be removed or fixed by selection even when the fitness differences between them were slight. The discovery that most populations contain an unexpectedly high amount of allelic diversity ran counter to this expectation. As a result, neutral models gained popularity among molecular evolutionists because they better explained these observations (Dietrich 1994). In this respect, the reception of neutral models within molecular biology differs markedly from the reception of neutral theory in ecology: in ecology there is, to date, no equivalent body of data that neutral models are unambiguously better able to accommodate than non-neutral models.

Another key distinction concerns the ways in which neutral models are employed as theoretical tools in the two disciplines. In molecular evolution, neutral models specify a baseline rate of expected allelic change given non-selective factors, including mutation rate and the relative influence of genetic drift as determined by population size. A significant departure from this baseline – involving either greater or lesser rates of change or observed variation – can be

taken as evidence that other evolutionary factors are at play, and often this includes natural selection.

Beatty's (1987) philosophical discussion about predictive overlap is relevant at this juncture. His concern was that, for a particular trait or locus, any amount of change predicted by a selectionist model could also be accommodated by a neutral model. This may be true if one's focus is restricted to a limited number of loci. However, modern genomic methods implement more refined analyses that are designed to detect specific signatures of selection. This may involve an examination of patterns of variation at specific loci among populations, or, of comparisons of many loci within a single genome (e.g. Vitti et al. 2013; Berg and Coop 2014). Because most forms of natural selection deplete allelic variation, selectionist models often predict that a locus under selection will exhibit significantly less variation than other loci in the genome that have not been under selection. By contrast, neutral models predict much more consistent rates of change or amounts of standing variation across multiple, independent (i.e., unlinked) loci. A lack of variation across wide swaths of the genome is thus indicative of genetic drift via population bottlenecks or founder effects, whereas depauperate variation localized to particular regions is suggestive of natural selection on those (or closely linked) loci. Thus, the problem of predictive overlap has been overcome in molecular evolutionary biology through the development of more discriminating predictions associated with models of neutral processes, as well as increased availability of large-scale genomic data that allow investigators to test those predictions. The same degree of refinement is currently lacking for ENT.

Perhaps the most important difference between the two disciplines, however, is that the term 'neutral' has a different meaning in molecular biology than it does it ecology. This can generate confusion, especially when applying neutral ecological models at the genomic level. In

the molecular sense, a genetic element is considered neutral if it has no impact on the *fitness of the organism* in which it resides. Thus, a 'nearly-neutral' insertion is one that has negligible effects on organismal survival and reproduction. Notice that this concept of neutrality is implicitly hierarchical, such that the neutrality of an element is determined by its impact on the fitness of the whole organism in which it resides. Ecologists, by contrast, use 'neutral' to refer to the *absence of competitive differences among different types of entity* – this was described earlier as 'equi-probable success,' a core assumption of ENT. This non-hierarchical sense of neutrality compares entities existing at the same level of biological organization, i.e. the same trophic level. For sake of clarity, we refer to the first kind of neutrality as 'fitness-neutrality' and to the second as 'competitive-neutrality'.

Over the past decade, models based on the differential inputs of selection versus neutral mechanisms have attracted renewed attention within evolutionary biology. This interest has been inspired to a significant extent by the work of Michael Lynch and colleagues (Lynch and Conery 2003, Lynch 2007, Lynch et al. 2011). Their Mutational-Hazard Model offers a mechanism of how non-coding genetic elements (transposable elements, pseudogenes, introns, etc.) accumulate in some genomes. The model assumes that these elements have mildly deleterious effects on fitness. Because these detrimental effects are individually weak, whether or not such insertions are removed from populations by purifying selection depends on effective population size – that is, on the relative strengths of natural selection versus genetic drift. In large populations, selection is sufficiently strong to remove modestly deleterious elements despite small selection coefficients. In small populations, genetic drift is the dominant factor driving the evolutionary fates of such insertions, such that they may increase in number despite being slightly deleterious. The overall expectation of this model is that larger genomes evolve in taxa with small effective

population sizes.

Lynch (2011) maintains that the Mutational-Hazard Model is not a 'neutral' model despite the tendency for some of his critics to interpret it as such. Our distinction between fitness-neutrality and competitive-neutrality helps to clarify this debate over the precise status of Lynch's model. Insofar as the Mutation Hazard Model assumes that non-coding insertions are mildly deleterious, Lynch is correct to insist that that these elements are not fitness-neutral, strictly speaking. It might be argued that Lynch nonetheless assumes that non-coding insertions are *effectively* fitness-neutral, at least when population size is small, since drift and mutation are the only factors determining genome evolution under these circumstances. But this difference between 'strict' and 'effective' fitness-neutrality is a fairly minor issue. A more pressing question concerns the relationship between population size and competitive-neutrality among different types of non-coding elements in Lynch's model. As we shall now argue, the Mutational-Hazard Model assumes that different types of element are competitively-neutral in large populations. In this respect it shares a core assumption with genome-level ENT.

A central piece of evidence offered in support of the Mutational-Hazard Model involves whole-genome comparisons among broad taxonomic categories –spanning phyla, kingdoms, and even domains (Lynch and Conery 2003; Lynch et al. 2011). These comparisons suggest that effective population size negatively correlates not only with genome size, but also with the number of transposable element copies, number and size of introns, and the retention of defunct gene duplicates. Hence, genome size and complexity are said to be greater in lineages with smaller historical effective population sizes. Perhaps the most striking pattern used to support this model involves a comparison among various prokaryotic and eukaryotic genomes. Prokaryotes, which are assumed to have population sizes that are orders of magnitude larger than

eukaryotes, tend to have much lower abundances of non-coding DNA. According to the Mutational-Hazard Model, the best explanation for this pattern is that eukaryotic genomes are dominated by evolutionary processes other than selection.

With these comparisons in mind, it seems clear that the Mutational-Hazard Model makes no assumption about the competitive-neutrality of different types of genetic element *provided that effective populations are small*. For example, in very small eukaryotic populations it does not matter whether TEs replicate at a higher rate than other types of non-coding DNA because all such sequences evolve as though they are effectively neutral. The point, as far as Lynch is concerned, is that all types of non-coding element tend to accumulate under these conditions, despite their negative impact on fitness. However, we argue that the model's assumption of competitive-neutrality among different element types becomes relevant at large effective population sizes. As we have seen, Lynch and colleagues explain the relative paucity of non-coding DNA within prokaryotes as a case of selection overpowering drift and mutation in large populations. This is just to assume that all types of non-coding DNA are equally subject to purifying selection under these circumstances. In other words, the Mutational-Hazard Model assumes that, as far as large effective populations are concerned, different types of non-coding element are competitively neutral precisely when they are no longer effectively fitness-neutral.

This point becomes clearer when we consider some of the criticisms that have been raised against the Mutational-Hazard Model. As has been pointed out by Charlesworth and Barton (2004, 2007), even in large host populations, TEs can become established despite negative selection on the host provided that they exhibit a sufficiently high rate of transposition. In other cases, TE insertions may be significantly deleterious, either individually (e.g., if they have a propensity to disrupt gene function) or in the aggregate (e.g., as their repetitive nature promotes

illegitimate recombination) (Hedges et al. 2007, Belancio et al. 2009). In these cases, the biological properties of the TEs themselves – and not simply the relative strength of genetic drift versus purifying selection based on host population size – would be a primary determinant of their rate of accumulation. Another way to put this point is that TEs are known to violate the assumption of competitive-neutrality, even in large populations. Thus, it is questionable whether their scarcity in prokaryotic genomes is explained by the Mutational-Hazard Model.

Upon reflection, this objection to the Mutational-Hazard Model should by now seem familiar. The problem stems from a failure to take into account known differences among different types of genetic element. Recall that a core assumption of ENT is that species membership is irrelevant in determining individual success. As with Serra et al.'s (2013) use of ENT in the study of genomes, the Mutational-Hazard Model treats all non-coding DNA sequences as effectively identical in their (very minor) impact on organismal fitness, at least when population size is large. In other words, the model does not take into account -- indeed, it assumes the irrelevance of -- element "membership" in the same way that ENT assumes the irrelevance of species membership—at least as far as large populations are concerned. Thus, TE insertions, intron additions, gene duplicates, and other events are all assumed to respond to purifying selection in the same fashion.

This observation raises the question of whether any other similarities exist between the Mutational-Hazard Model and ENT. Notably, the Mutational-Hazard Model does not explicitly assume a meta-population structure, at least not at the level of the whole organism. However, viewed from the perspective of genome-level ecology, it is possible to view individual genomes as communities of genetic elements and the entire population of organisms as a meta-community of genomic elements. In this case, different patterns of gene flow among organisms can result, at

the genomic level, in meta-communities with potentially very different structures.

It has been argued elsewhere that the Mutational-Hazard Model is sensitive to factors affecting gene flow. This includes differences in the degree of admixture within genomes (recombination rate), among individual organisms/genomes (sexual versus asexual reproduction, breeding system), and/or host populations (different rates of gene flow) – all of which can affect the spread of transposable elements (Charlesworth and Barton 2004). Our argument here is that the reasons that population structure matter for the Mutational-Hazard Model are equivalent to the reasons that meta-population structure matters to genome-level ENT. In both cases, this factor significantly influences the accumulation of competitively-neutral entities in a population/meta-community.

A further similarity between these models is revealed by reflecting on what it means for genomes to have a fixed 'carrying capacity' and how this property might impact the accumulation of non-coding genetic elements. As noted previously, an auxiliary assumption of most neutral ecological models is that the local communities have limited carrying capacity. We have seen that transposable elements can generate positive feedback by creating additional sites into which more elements can then be inserted without disrupting protein-coding genes or regulatory regions. This poses a problem for employing ecological neutral models to transposable elements because it leads to a non-fixed carrying capacity. With regard to the Mutational-Hazard Model, this means that not all insertions will be equally deleterious. Indeed, in small number and in non-essential regions of the genome, these insertions may exert very little fitness effect on the organism. Instead, fitness effects may result from the larger-scale relationships between genome size, cell size, and cell division rate – which in turn may manifest as links between genome size and body size, metabolic rate, developmental rate, and other

organismal traits (Gregory 2005a). The selective pressure on these traits differs markedly among taxa: for example, there may be severe constraints on metabolic rate (and therefore cell size and genome size) in flying vertebrates but very few constraints in groups with low metabolic rates and simplified development such as neotenic salamanders (e.g., Gregory 2004, 2005a). In other words, the carrying capacity for transposable elements can be radically different among groups of organisms. The result is significant variation in the limits on genome expansion through TE insertion. This often occurs independently of population size, and would therefore represent another major factor that is overlooked in the Mutational-Hazard Model.

In sum, we have argued that the Mutational-Hazard Model shares several features with genome-level ENT. Indeed, criticisms of Lynch's model, and its use to explain differences among eukaryotic and prokaryotic genomes, bear an uncanny resemblance to the arguments raised in previous sections against Serra et al. (2013). This comparison helps to shed light on the structure of neutral models in general as well as the empirical challenges that they face.

6. Concluding Remarks

There is great promise in thinking of genomes as ecosystems in miniature. Not only does this perspective stand to shed light on pressing questions about the abundance and distribution of genetic elements among genomes, but, in addition, genomes might serve as excellent model systems for testing ecological hypotheses (Linquist et al. 2013). However, applying ecological models at the genomic level requires careful attention to the assumptions of those models. This can be especially challenging when one applies ecological models to unconventional entities – such as genetic elements.

ENT would appear to stand out as a likely candidate for application at the genomic level.

However, as we have argued, this model suffers from the problem of predictive overlap: both neutral and non-neutral models make numerous identical predictions, especially regarding the rank abundance of species. The extent of this overlap makes it difficult to infer neutral processes from neutral patterns. For the same reason, neutral models, as they stand, cannot serve as legitimate null hypotheses for niche-based models. Nor can the inference from pattern to process be underwritten by appeals to parsimony, that are inherently vague, that misplace the burden of proof, and that arguably irrelevant even under ideal circumstances. The only way to determine whether a neutral model accurately describes some genomic system, we argue, is by considering whether that model's assumptions are satisfied.

We have argued that the two core assumptions of ENT are not satisfied by genomic systems because (1) different types of genetic element are known to have different likelihoods of success, and (2) genomic systems do not obviously exhibit a metacommunity structure. We further suspect that transposable elements in particular violate two common auxiliary assumptions of neutral models: (3) by creating their own habitat through replication and insertion, TEs do not have limited carrying capacity in many genomes; and (4) due to their mobility, TEs often do not have limited dispersal throughout the genome. It should be noted that the failure of many genomic systems to conform with the assumptions of ENT does not undermine the promise of other ecological approaches to the genome. There are many non-neutralist ecological models that can be applied at the genomic level (e.g. Venner et al. 2009, Abrusan and Krambeck 2006, Saylor et al. 2013). It is also important to note that our criticisms have focused primarily on applications of neutral models *across the entire genome*. It is possible that some other, more restrictive application of neutral models might prove more successful. For example, it might be possible to apply neutral theory within just one category of non-coding

DNA (e.g. pseudogenes).

Suppose, then, that ENT is inappropriately applied at the genomic level. It seems paradoxical that Molecular Neutral Theory has enjoyed such success in evolutionary biology. What explains this discrepancy? The answer becomes clear once we distinguish two different senses in which genetic elements can be defined as 'neutral.' In Kimura's sense, a genetic element is considered neutral (or not) depending on whether it impacts the fitness of the organisms in which it occurs. This is a fairly stable definition of neutrality in the sense that there are no additional parameters requiring specification by the investigator. By contrast, a genetic element is considered neutral (or not) in the ecological sense if its membership in some taxonomic or functional category does not contribute to its likelihood of success. There are at least two open-ended parameters in this ecological definition of neutrality. Firstly, one must settle upon a relevant taxonomic or functional category in relation to which neutral elements are defined. At the level of genetic elements there is a wide range of candidates that could potentially serve as the genomic analogues for species. As we have seen, simply treating all genetic elements as belonging to the same taxonomic or functional type (as far as their ecological neutrality is concerned) undermines a core assumption of neutral ecological models. The second open-ended parameter is the physical scale over which relative 'success' is defined. A genome-level ecologist has the option of measuring success intra-chromosomally, within a genome, among closely related genomes, among distantly related genomes, etc. Each scale of comparison can result in a different answer to the question of whether genetic elements belonging to the same taxonomic or functional type have an equal likelihood of success. This is not to say that ENT models cannot in principle be applied at the genomic level. However, identifying a reliable genomic signature for distinguishing neutral from non-neutral processes – something that will be

required in order to avoid the problem of predictive overlap—is bound to be even more difficult when definitions of "neutrality" are open ended.

Acknowledgements

The authors thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for support, Dr. Maureen O'Malley and two anonymous reviewers for helpful comments, and Dr. Louis Bernatchez for the invitation to contribute a review.

References

Abrusan G, Krambeck H.J (2006). Competition may determine the diversity of transposable elements. *Theoretical Population Biology*, **70**, 364 – 375.

Adler, PB, Ellner SP, Levine JM (2010) Coexistence of perennial plants: an embarrassment of niches. *Ecology Letters*, **13**, 1019–1029.

Beatty, J (1987) Natural selection and the null hypothesis. *The Latest on the Best: Essays on Evolution and Optimality (Dupre J, ed.*) MIT Press: Cambridge, Mass.

Belancio VP, Deininger PL, Roy-Engel AM (2009) LINE dancing in the human genome: transposable elements and disease. *Genome Medicine*, **1**, 97.

Bennett MD, Leitch IJ (2005). Genome size evolution in plants. In: *The Evolution of the Genome*, edited by TR Gregory. Elsevier, San Diego, pp. 89-162.

Berg JJ, Coop G (2014). A population genetic signal of polygenic adaptation. PLOS Genetics 10(8): e1004412. doi:10.1371/journal.pgen.1004412.

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Muller S, Eils R, Cremer C, Speicher MR, Cremer T (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology*, **3**, e157.

Brookfield J FY (2005) The ecology of the genome –mobile DNA elements and their hosts. *Nature Reviews Genetics*, **6**, 128 – 136.

Caswell H (1976) Community structure: a neutral model analysis. *Ecological Monographs*, **46**, 327-254.

Charlesworth B, Barton B (2004) Genome size: does bigger mean worse? *Current Biology*, 14, R233-R235.

Charlesworth B (2007) The origin of genomes – not by natural selection? *Current Biology*, 18, R140

Chisholm RA, Pacala SW (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proc. Natl. Acad. Sci*. U.S.A., **107**, 15821–15825.

Clark J (2012) The coherence problem with the Unified Neutral Theory of Biodiversity. *Trends in Ecology and Evolution*, 27, 198-202.

Connolly SE, MacNeil, MA, Caley MJ, Knowlton N et al. (2014). Commonness and rarity in the marine biosphere. Proc. Natl. Acad. Sci. U.S.A., 111, 8524-8529.

Cottenie K (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters* 8, 1175-1182

Craver C (2006) When mechanistic models explain. *Synthese*, **153**, 355-376.

Daubin V, Moran NA (2004) Comment on " The origins of genome complexity". Science 306, 978a-978b

Dietrich M, Millstein R, (2008) The role of causal processes in neutral and nearly neutral theories. *Philosophy of Science,* **75**, 548-559.

Dornelas M, Connolly SR, Hughes TP (2006) Coral reef diversity refutes the neutral theory of biodiversity. Nature, **440**, 80–82.

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, **41**, 331-368.

Eickbush TH (2002) R2 and related site-specific non-long terminal repeat retrotransposons. In: Mobile DNA II (eds. Craig NL, Craigie R, Gellert M, Lambowitz AM), pp. 813-835. American Society of Microbiology Press, Washington, DC

Elliott TA, Stage DE, Crease TJ, Eickbush TH (2013) In and out of the rRNA genes: characterization of *Pokey* elements in the sequenced *Daphnia* genome. *Mobile DNA*, 4, 20 doi:10.1186/1759-8753-4-20.

Fortin MJ, Dale MRT (2005) Spatial analysis – A guide for ecologists. Cambridge University

Press, Cambridge. xiii + 365 pp.

Gregory TR (2004). Macroevolution, hierarchy theory, and the C-value enigma. Paleobiology 30: 179-202.

Gregory TR (2005a). Genome size evolution in animals. In: *The Evolution of the Genome*, edited by TR Gregory. Elsevier, San Diego, pp. 3-87.

Gregory TR (2005b). Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics*, **6**, 699 – 708.

Gregory TR., Andrews CB, McGuire JA, Witt CC (2009). The smallest avian genomes are found in hummingbirds. Proceedings of the Royal Society B 276: 3753-3757.

Griffiths D (2010) Pattern and process in the distribution of North American freshwater fishes. *Biological Journal of the Linnean Society*, **100**, 46-61.

Han JS (2010) Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mobile DNA* **1**, 15.

Havecker ER, Xiang Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. *Genome Biology*, **5**, 225

Hedges DJ, Deininger PL (2007) Inviting instability: transposable elements, double-strand breaks, and the maintenance of genomic integrity. *Mutation Research*, **616**, 46-59

Hubbell SP (1997) A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs*, **16**, S9-S21

Hubbell SP (2001) *A Unified Theory of Biodiversity and Biogeography*. Princeton University Press: Princeton, NJ.

Kimura M (1968). Evolutionary Rate at the Molecular Level. *Nature*, **217**, 624-6.

Kimura M(1969), "The Rate of Molecular Evolution Considered from the Standpoint of Population Genetics", *Proceedings of the National Academy of Sciences* 63: 1181-1188.

Kimura, M. (1983). *The neutral theory of molecular evolution.* Cambridge University Press, Cambridge, UK.

King JL, Jukes TH (1969) Non-Darwinian Evolution. *Science*, **164**, 788-97.

Kojima KK, Jurka J (2013) A superfamily of DNA transposons targeting multicopy small RNA genes. *PLoS One*, **8**, e68260.

Lack D (1947) *Darwin's Finches*. Cambridge, UK: Cambridge University Press.

Le Rouzic A, Dupas S, Capy P (2007). Genome ecosystem and transposable elements species.

Gene, **390**, 214 – 220.

Linquist S, Saylor B, Cottenie K, Elliott TA, Kremer SC, Gregory TR (2013). Distinguishing ecological from evolutionary approaches to transposable elements. *Biological Reviews*, **88**, 573-584.

Lynch M (2007) The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics*, **8**, 803-813.

Lynch M, Conery JS (2003) The origins of genome complexity. Science, **302**, 1401-1404.

Lynch M, Bobay L-M, Catania F, Gout J-F, Rho M (2011) The repatterning of eukaryotic genomes by random genetic drift. Annual Review of Genomics and Human Genetics 12, 347-366.

MacArthur RH (1961) *Geographical Ecology: Patterns in the Distribution of Species*. Princeton University press: Princeton, NJ.

Marquet PA, Keymer JE, Cofre H (2003) Breaking the stick in space: of niche models, metacommunities and patterns in the relative abundance of species. Macroecology: concepts and consequences Blackwell Publishing: Malden Mass. Pp 68-84.

McGill BJ, Maurer BA, Weiser MD (2006) Empirical evaluation of neutral theory. *Ecology*, **87**, 1411-1423.

Muneepeerakul R, Bertuzzo E, Lynch HJ, Fagan WF, Rinaldo A, Rodrigues-Itrurbe I (2008) Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. *Nature*, **453**, 220-223.

Nee S, Stone G (2003) The end of the beginning for neutral theory. *Trends in Ecology and Evolution*, **18**, 433-434.

Neumann P, Navrátilová A, Koblížková A, et al. (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* 2, 4.

Olden JD, Kennard MJ, Leprieur F, Tedesco PA, Winemiller KO, Garcia-Berthou E (2010) Conservation biogeography of freshwater fishes: recent progress and future challenges. *Diversity and Distributions*, **16**, 496-513.

Pardue ML, DeBaryshe PG (2011) Retrotransposons that maintain chromosome ends. *Proceedings of the National Acadademy of Sciences. U S A*, **108**, 20317-24.

Pearce T (2011) Ecosystem engineering, experiment, and evolution. *Biology & Philosophy*, **26**, 793-812.

Rosindell J, Cornell SJ, Hubbell SP, Etienne RS (2010) Protracted speciation revitalizes the neutral theory of biodiversity. *Ecology Letters*, **13**, 716-727.

Rosindell J, Hubbell SP, Etienne RS (2011) The unified neutral theory of biodiversity and

biogeography at age ten. *Trends in Ecology and Evolution*, **26**, 340-348.

Rosindell J, Hubbell SP, He F, Harmon LJ, Etienne RS (2012) The case for ecological neutral theory. *Trends in Ecology and Evolution*, **27**, 203-208.

Saylor B, Elliott TA, Linquist S, Kremer SC, Gregory TR, Cottenie K (2013). A novel application of ecological analyses to transposable element distributions in the *Bos taurus* genome. *Genome,* **56**, 521-533.

Serra F, Becher V, Dopazo H (2013) Neutral theory predicts the relative abundance and diversity of genetic elements in a broad array of eukaryotic genomes. *PLoS One,* **8**, E63915.

Sober E (1994) Let's razor Occham's razor. *From a Biological Point of View*. Cambridge: Cambridge University Press, pp. 136–57.

Turnbull LA, *et al.* (2005) Niches, rather than neutrality, structure a grassland pioneer guild. *Proceedings of the Royal Society B: Biological Sciences,* **272**, 1357–1364.

Venner S, Feschotte C, Biemont C (2009). Dynamics of transposable elements: towards a community ecology of the genome. *Trends in Genetics*, **25**, 317 – 323.

Vitte JJ, Grossman SR, Sabeti PC (2013) Detecting natural selection in genomic data. *Annual Reviews of Genetics*, **47**, 97-120.

Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature,* **424**, 1035–1037.

Wennekes PL, Rosindel J, Etienne RS (2012) The neutral-niche debate: a philosophical perspective. *Acta Biotheoretica*, **60**, 257-271.

Wright JP, Jones CG (2006) The concept of organisms as ecosystem engineers ten years on: progress, limitations, and challenges. *Bioscience*, **56**, 203-209.

Zou S, Ke N, Kim JM, Voytas DF (1996) The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes & Development* 10, 634-645.