Individuating population lineages: a new genealogical criterion

Beckett Sterner[1] ![ORCID]

Abstract Contemporary biology has inherited two key assumptions from the Modern Synthesis about the nature of population lineages: sexual reproduction is the exemplar for how individuals in population lineages inherit traits from their parents, and random mating is the exemplar for reproductive interaction. While these assumptions have been extremely fruitful for a number of fields, such as population genetics and phylogenetics, they are increasingly unviable for studying the full diversity and evolution of life. I introduce the "mixture" account of population lineages that escapes these assumptions by (1) dissolving the Modern Synthesis's sharp line separating reproduction and development and (2) characterizing reproductive integration in population lineages by the ephemerality of isolated subgroups rather than random mating. The mixture account provides a single criterion for reproductive integration that accommodates both sexual and asexual reproduction, unifying their treatment under Kevin de Queiroz's generalized lineage concept of species. The account also provides a new basis for empirically assessing the effect of random mating as an idealization on the empirical adequacy of population genetic models.

Keywords Biological species · Meta-population · Hybridization · Darwinian population · Biological individuality · Neutral model

✉ Beckett Sterner
beckett.sterner@asu.edu

1   Arizona State University, Tempe, AZ, USA

Introduction

Contemporary biology has inherited two key assumptions from the Modern Synthesis about the nature of population lineages: sexual reproduction is the exemplar for how individuals inherit traits from their parents, and random mating is the exemplar for reproductive interaction. Both assumptions take a particular process that contributes to reproductive integration in some populations as a model for how all population lineages arise and persist. Sexual reproduction generates reproductive integration through the fusion of haploid gametes into a diploid zygote and chromosomal recombination during gamete production. Biologists typically judge the contributions of other reproductive processes to integration, such as bacterial transformation, by their similarity to sex. In an analogous way, random mating ensures that any groups uniformly share hereditary material over time. When biologists try to locate populations on a continuum between random mating and total fragmentation, for example using Wright's F_{st} measure, they treat random mating as the paradigm for 'good' population lineages. While these assumptions have been extremely fruitful for a number of fields, such as population genetics and phylogenetics, they are increasingly unviable for studying the full diversity and evolution of life (Jablonka and Lamb 2005; Doolittle and Zhaxybayeva 2010; Landguth et al. 2015).

In this article, I introduce a new definition of population lineages that escapes these assumptions and expands our ability to define biological species and study the effect of geography on evolvability. Biologists have typically defined population lineages as the units of evolution that emerge when individual chains of parents and offspring are integrated into a cohesive whole that persists over time (Simpson 1961; Hull 1980; de Queiroz 1999). Generally, population lineages have two characteristic features: they are reproductively isolated from other groups, and they are internally integrated through reproductive interactions between individuals. Hence clades do not count as population lineages because they exhibit persistent splits between subgroups, and clones—the descendants of a single individual through asexual reproduction—do not generally form population lineages because the offspring fail to "mix" reproductively over time.

The primary theoretical challenge is to make informal notions of 'isolation' and 'integration' more precise. In order to organize and articulate the possible approaches we might pursue, we can classify existing accounts in terms of how they prioritize structural or functional perspectives on reproductive integration, and whether they characterize the dynamics of populations in terms of patterns or processes. Structural perspectives individuate population lineages in terms of causal interactions among individuals and their environment, while functional perspectives characterize population lineages in terms of their capacities as unitary wholes, such as the capacity to evolve under natural selection (Hull 1980; Godfrey-Smith 2009; Matthewson 2015).

Challenging the paradigmatic status of sex and random mating forces us to reconsider the structural possibilities of individual life cycles and the ways in which they should interact in order to constitute a population lineage. After reviewing existing perspectives on population lineages, I summarize a recent modification of

James Griesemer's work on reproduction and development that introduces the concept of a "demarcator" as a suitably general foundation for individuating biological units (Sterner 2017). I then develop a formal model for reproductive interactions and use it to articulate a new "mixture" account of population lineages based on a probabilistic view of what it means to have "ephemeral" branching within a population over time (de Queiroz 1999).

Perspectives on population lineages

Historically, the concept of a population lineage has not been a major topic in its own right in biology or philosophy; instead, it usually comes up in discussions of other issues, such as the species debate (de Queiroz 1999), the conditions for natural selection to occur (Hull 1980; Godfrey-Smith 2009), and the nature of biological populations more generally (Millstein 2015). As a result, the relevant literature is scattered and lacks a shared starting point. This section provides a common framework that relates the key assumptions and aims of different approaches, although it is not necessary for the main argument of the paper.

Any definition of a population lineage needs to address four essential elements: (1) which individuals form the population over time, (2) what causal relations are possible between individuals, (3) what causal relations are relevant between individuals and their environments, and (4) what criteria determine if a group of individuals forms an integrated whole. Within the category of causal relations between individuals, definitions typically distinguish between reproductive or inheritance relations and ecological interactions such as competition for resources (c.f. Millstein 2015; Matthewson 2015). Any definition will privilege some of these elements, e.g. the reproductive process, and remain neutral on others, such as the capacity for natural selection.

We can classify the different approaches using the common biological distinctions between structure and function and between pattern and process. Structure and function have a long history as contrasting perspectives on biology (Griesemer 2005; Laubichler and Maienschein 2007; Boucher 2015; Brigandt 2017), but for our purposes here I will treat them as providing complementary rather than antithetical views of living systems. In our context, functional approaches individuate population lineages according to the capacities of the group as a whole, while structural approaches individuate population lineages based on causal interactions among individuals and their environments. In other words, functional approaches specify what population lineages are able to do as units and structural approaches specify how the parts of population lineages are causally interrelated. These two perspectives are complementary insofar as knowledge within one perspective provides a basis to formulate novel explanations or predictions addressing questions prioritized by the other perspective (Sterner 2015).

Pattern and process are similarly complementary ways of characterizing a natural phenomenon. The distinction comes from evolutionary biology and is more recent than structure and function (Avers 1989). Patterns are empirical relationships among observable properties of the phenomenon, and they are individuated independently

from the actual causes that produced the empirical relationship. Process, by contrast, refers to a causal mechanism or series of events that generate a phenomenon. Biologists distinguish pattern from process in order to reason about when an empirical relationship is a reliable indicator that some process is responsible for the phenomenon.

Table 1 applies these two distinctions to organize existing work relevant to the problem of defining population lineages. Both approaches in the structure-pattern category rely on the topology of interactions between individuals over time. The first, Kornet's internodal species concept, provides a definition of biological species relevant to understanding population lineages (Kornet 1993; Kornet et al. 1995; Kornet and McAllister 2005). An internodon is an equivalence group: a set of individuals located between the same two branching points in the population genealogy that are connected to each other by chains of ancestor–descendant relations. Internodons are externally isolated by definition, because any "outside" individual mating with a member inside the internodon would automatically count as a member. Internodons are also internally integrated in the limited sense that all members are connected by chains of parent–offspring relationships, although in some cases the offspring may fall outside the internodon that contains the parents. Kornet's account falls into the structure-pattern category because it uses the network of parent–offspring relationships to identify equivalence groups as the key structural pattern characterizing internodons.

The internodon species concept has several limitations, however. Even a single mating between two otherwise unconnected internodons merges the two into a single unit. Such a low threshold for hybridization is especially problematic if being a parent is a matter of degree, such as with lateral gene transfer or behavioral learning. Moreover, determining which populations count as internodons depends on comprehensive knowledge of their genealogical history, with little guarantee that these groups will remain stable in the future.

The second approach defines populations as a group of conspecific individuals that interact much more often with each other than with outsiders in ways that affect survival and reproduction (Millstein 2009, 2010, 2015). Millstein's causal interactionist population concept individuates groups by looking for patterns of modularity extending over time in the network of interactions among individuals. Emphasizing interactions that affect survival and reproduction increases the likelihood that populations are important units of evolution for the purposes of evolutionary biology.

While modularity is more flexible than Kornet's internodons, Millstein's approach inverts the logical relationship between population lineages and species.

Table 1 Perspectives on the nature of population lineages

	Structural	Functional
Pattern	Topology of interaction networks	None
Process	Gene flow by sexual reproduction and random mating	Capacity to evolve by natural selection

On de Queiroz's view, the concept of a population lineage is defined logically prior to any species concept, but Millstein presupposes that we already know which individuals are conspecifics. Assuming conspecificity matters because populations otherwise might include individuals from widely diverged taxa that are in strong ecological competition with eachother (Matthewson 2015). Another issue is that Millstein's interaction networks only allow one type of edge, so very different kinds of interactions affecting survival and reproduction are lumped together to form the aggregate network. Any variation in the networks formed by different kinds of interactions is therefore lost.

Sexual reproduction and random mating have been central for structure-process approaches. Many key models and measures of population structure in phylogenetics and population genetics assume random mating with sexual reproduction (Liu et al. 2009; Wakeley 2009). However, research in the past few decades has undermined the adequacy of sexual reproduction as a general model for how individuals form offspring that hybridize distinct lineages of descent (Kendig 2014). Several key properties that are true of sex but fail to generalize include: (1) a descendant's ancestors are fixed at the start of its life cycle; (2) every descendant in a population has the same number of ancestors; (3) the integration of ancestor–descendant chains involves genetic recombination. The universality of these assumptions are undermined by a number of the same phenomena, which I won't attempt to review in detail, but include epigenetic inheritance, niche construction, horizontal gene transfer, chimeric fusion, and symbiotic relationships in holobionts (Odling-Smee et al. 2003; Jablonka and Lamb 2005; Doolittle 2010; Chiu and Gilbert 2015).

Additionally, random mating models idealize away the spatial distribution of individuals, eliminating a major factor affecting reproductive integration in real populations. The precise spatial relations between individuals in a population can have a qualitative impact on how speciation proceeds (Mallet et al. 2009; Church and Taylor 2002), and it is not always possible to simplify the geography of a group into a set of populations that each internally exhibit random mating and are linked by migration. More broadly, biologists' growing recognition of the importance of geography is reflected in the rapid growth of two new subfields, landscape genetics and phylogeography (Manel et al. 2003; Knowles 2009). While random mating is one of the most powerful idealizations in population genetics, it does not offer a universally adequate basis for theorizing about population lineages.

Functional approaches to population lineages are dominated by the function-process category. The process perspective prioritizes a population's capacity to undergo natural selection (Hull 1980; Godfrey-Smith 2009; Clarke 2013; Matthewson 2015). A "Darwinian population," to use Godfrey-Smith's term, is a population with properties that make it highly evolvable under natural selection. Recent work in this vein relies on Lewontin's three criteria for natural selection to occur (Lewontin 1970). While useful as a way to theorize about units of selection, functional approaches to populations have severe disadvantages as tools for phylogenetics or population genetics. Each account relies on abstract capacities of whole populations, such as the capacity to undergo selection, which are difficult to estimate empirically in the best of cases and are simply unavailable in many

contexts, especially across long time periods. In addition, Lewontin's criteria depend on precisely those concepts, e.g. reproduction and inheritance, whose paradigm exemplars are being stretched by phenomena such as epigenetic inheritance, lateral gene transfer, and holobionts.

One benefit of the structure-pattern approach I articulate is that it opens new possibilities for the empirical investigation of the causes of high evolvability, which may challenge the criteria used by existing functional accounts. In general, I take a pluralist stance to the different approaches I have articulated here, each of which provides distinctive epistemic value to advancing biological research. Nonetheless, sometimes new discoveries and theoretical results arise that require us to criticize and renew the foundations of an approach.

Rethinking individuals in population lineages

Existing structural views are constrained by relying on sexual reproduction and random mating as paradigms. In this context, James Griesemer's reproducer account of individuality has several relevant virtues: it introduces a generalized view of reproduction that is not limited to genetic replication, and it explicitly incorporates the ability of individuals to hybridize during their life cycles (Griesemer 2000, Griesemer 2014a, b). Although the topic of biological individuality may seem like a digression, structuralist views of population lineages are determined by the interactions they allow among individuals, so a new perspective on the space of possible interactions provides a fruitful starting point for new theory.[1] In this section I briefly motivate and review a modification to Griesemer's approach based on the concept of a "demarcator" (Sterner 2017).

Griesemer aims to develop a theory of life cycles that breaks down the strict separation of reproduction and development imposed by the Modern Synthesis. Traditionally, biologists have assumed that reproduction and development form two distinct phases of an individual's life cycle: reproduction is the fertilization of haploid gametes into a diploid zygote, and development consists of the zygote's growth into a sexually mature individual. The Modern Synthesis also follows a Weismannian view of inheritance: the cells of a developing embryo segregate into a germ line and a soma (the rest of the body), and only cells in the germ line may contribute genetic material to the next generation (Griesemer 2005). Development can therefore be ignored when studying inheritance and the genealogies of individuals.

In contrast, Griesemer conceives of reproduction as "a process with two aspects: progeneration and development. Progeneration is the multiplication of entities with material overlap of old (parent) and new (offspring) entities. Material overlap means that some of the parts of the offspring were once parts of the parents" (Griesemer 2000, p 242). After cell division, the genomes inherited by two daughter cells

---

[1] The relevance of demarcators for existing views of biological individuality is developed in more detail in (Sterner 2017). Developing the full implications of the demarcator formalism for other conceptions of individuality and lineages (Bouchard 2008; Bapteste and Dupré 2013; O'Malley 2014; Doolittle and Booth 2016) remains a topic for future work.

structurally resemble each other because their DNA was copied from a parental template. This resemblance is not the full story, though, because each daughter cell acquires its DNA from the parent through the transmission of one old and one new DNA strand. The parent–offspring relationship therefore involves material overlap and not just resemblance because off-spring are made from "organization-preserving physical parts" (Griesemer 2014b, p 26). Material overlap offers a more general way of thinking about the reliable inheritance of developmental capacities compared to biologists' typical focus on DNA. In other words, reproduction on Griesemer's view encompasses but does not require the inheritance of genes.

He also argues that not all developmental resources are passed on through material overlap. "Successful developers are not only born organized but are also often born into environments that 'scaffold' them in ways that use order in the environment to organize aspects of the developing system" (Griesemer 2014b, p 26). A scaffold facilitates a process that would otherwise be more difficult or costly and is often temporary from the perspective of the overall life cycle. Scaffolding therefore serves as a complementary mechanism for inheritance beyond what is passed on by a germ line.

Both material overlap and scaffolding provide pathways for hybridization, and hybridization events are crucial to identifying important stages in life cycles. Griesemer understands hybridization here as "biological systems incorporating parts of different provenance," i.e. from different lineages or genealogies, rather than the narrower genetic sense (Griesemer 2014a, p 191). Material overlap and scaffolding therefore provide for complementary kinds of material interaction that jointly delineate structural stages in life cycles. For an extended example of how Griesemer uses these concepts to analyze the life cycle of the HIV virus, see (Griesemer 2014a, b).

Griesemer's account provides two key insights for a generalized structuralist approach: first, identify causally significant entities that establish boundaries which in turn mediate the production and transfer of material; second, track how material moves across these boundaries. These insights follow in the spirit of Kornet's approach: "The relationship from parent to offspring involves the transfer of genetic material. This transfer of matter takes place, directly or indirectly, between any two individuals in a network. This pattern of transfers of matter is what makes a genealogical network a material system. A genealogical network is cohesive in virtue of being such a system" (Kornet 1993, p 428). However, we don't have to make special commitments to genes as the units of inheritance or sex as the process by which reproductive lineages hybridize.

Instead of distinguishing between hybrids and non-hybrids as Griesemer does, I will use the term "demarcator" for the boundary-establishing entities that underwrite the possibility of hybridization (Sterner 2017). The primary conceptual role of the demarcator is to characterize an individual by an objective distinction between what it includes as parts from what it excludes as non-parts. A demarcator may be a material object that generates a spatial boundary, such as semi-permeable surfaces like membranes or skin. Alternatively, it may be a causal system that discriminates among classes of objects, such as an immune system or mechanism

for mate recognition. Material overlap then requires that a material part of one system becomes a material part of another system at a later time. This could happen, for example, by the first system dividing in half or by the part moving across the two boundaries. Scaffolding involves material that originates or ends up excluded by at least one of the systems.

The significance of demarcators for understanding a life cycle follows from the interactions they facilitate or prohibit among parts of the overall system. Cell membranes, for example, are porous to small molecules such as water or ions but block large proteins and polysaccharides unless the cell installs specific channels for import and export. Similarly, membranes influence how likely molecules inside the cell (or embedded in the membrane itself) are to interact by constraining their diffusion (Clarke 2013; Godfrey-Smith 2016).

More precisely, a demarcator is important for understanding a life cycle to the extent that it serves as a focal point for causal control of events in the life cycle. A demarcator is a focal point for control to the extent that the variation possible within a life cycle can be explained by: (a) the demarcator's effects on what becomes a part of the object in question; (b) its effects on what causal interactions are possible among parts, among non-parts, and between the two; and (c) changes in its properties that alter its effects on the life cycle.

For simplicity, I will not address the distinction between material overlap and scaffolding in the remainder of the paper or attempt to give formal definitions of reproduction, development, and life cycle. As a result, the semantics of the demarcator formalism are not exhaustive and rely on expert judgment to capture the evolutionary dynamics of interest.

Defining population lineages: closure and mixture

This section introduces the demarcator formalism and characterizes population lineages in terms of two key properties, closure and mixture. In the formalism, demarcators establish quasi-spatial relationships between material entities by dividing the world into two sets: things that are included by the demarcator, and things that are excluded. We can track entities over time by watching how they move among demarcators and thereby generate a collective network of material transfer and production relations across generations. If this network exhibits closure and mixture over time, then we can say that the series forms a population lineage.

Note that I use "population" here in the mathematical sense as an arbitrary set of individuals, recognizing that in practice which sets we consider is often guided by features such as geography and ecology (Millstein 2015). "Population lineage" as I use it may encompass sets of individuals located in different spatial regions or habitats so long as these sets exhibit the appropriate patterns of reproductive interaction. Biologists, however, would call the individuals in each region a separate population and the whole collection of individuals a meta-population.

We can now start laying out key concepts in the formalism. A *demarcator type* is a kind of entity that (1) partitions the set of all other entities in the world into two subsets: included and excluded; and (2) imposes constraints on the types of entities

it may include or exclude. An individual demarcator, such as the membrane of some particular cell, is a token instance of some demarcator type. The differences between demarcator types lie in the constraints they impose on what can or must be included or excluded by tokens of that type. One could also extend these constraints to include rules about what kinds of entities can be transferred "into" and "out of" the demarcator type. Similarly, tokens of one demarcator type could be forbidden from including entities that are also included by tokens of another type. In what follows, I will use demarcator to refer to a token instance of some type unless otherwise stated.

Each demarcator can be represented by two sets that partition all the entities in the world. The included set contains all the entities that are parts of the demarcator, while the excluded set contains all the entities that are non-parts. Figure 1 illustrates a simple example that models bacteria engaging in the transfer and production of plasmids (small, circular DNA strands capable of independent replication). Note that there is nothing fancy going on here: we are just carving up a big set (the world) at each time step into complementary subsets (included and excluded) a number of times equal to the number of demarcator tokens in the world.

We can define the *location state* of an entity as a vector that lists the relations (included or excluded) that the entity has to each demarcator in the world. The model is *logically consistent* if the location state of every entity satisfies the constraints imposed by the types of the demarcators.
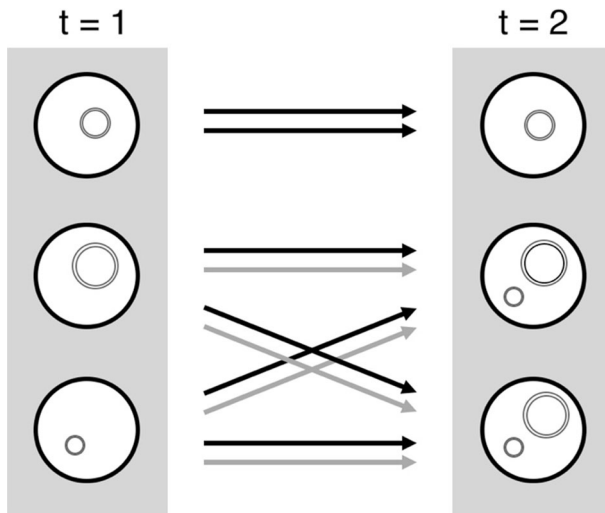

Fig. 1 A population series with two time steps. The figure is drawn to suggest cells demarcated by membranes (*black circles*) which include plasmids, drawn as *double circles* to reflect the two complementary DNA strands that make up each plasmid. *Black arrows* represent transfer events and *gray arrows* represent production events. The *top pair of arrows* therefore show the transfer of both strands in the medium-sized plasmid. In contrast, the large and small plasmids reproduce between between t = 1 and t = 2. Since plasmids reproduce through template-based DNA replication, their offspring each contain one old and one new strand. As a result, the offspring plasmids are shown as linked to their parent by both transfer and production relations

Now let's consider several kinds of relations among entities. To do so, we need to add a time index to the state of the world. In order to talk about an entity being transferred from one location to another, we also need to be able to say that entities existing at two different times are the same individual. So let's define *sameness* in this sense as a primitive relation between two entities present in two time steps of the model (not necessarily consecutive). A *transfer event* is a relation between two entities that exist at consecutive times and are the same individual but have different location states. Furthermore, we can specify a *production event* as a primitive relation between two entities existing at two consecutive times such that the later entity is not the same individual as any entity existing at a previous time. Production events relate new entities to the entities responsible for creating them in the previous time step.[2] Figure 1 illustrates these kinds of relations.

We are now in a position to define closure and mixture for population lineages. Let W_t denote the world of entities and their location states at time t. A model in the formalism consists of a sequence of worlds (W_1 … W_T) and a network of transfer and production relations connecting the entities. Figure 1, for example, represents a model with two time steps. In order to define closure, we need to look at the production and transfer relations between a subpopulation of entities and the rest of the world. Let P_t be a subpopulation of the entities in W_t, and $P_{1..T}$ be a temporal series of subpopulations. For notational simplicity, I will suppress the time index in $P_{1..T}$ and just refer to it as P going forward.

First, let's define *closure* relative to an entity type and then work up to a definition of total closure over a set of entity types. What we need to check, in effect, is whether any demarcators in the population series P have acquired entities of some type E from "outside" P. In other words, closure means that demarcators in P contain only entities of type E that came from other demarcators in P. We can start by collecting all the entities of type E that are included by any demarcators in P, ignoring the first time point. For each entity in this collection, we then identify all other entities of type E in the previous time step that are related to it by transfer or production relations. Is every entity related in this way included by some demarcator in P? If no, then closure fails. If yes, then we move to consider the next entity in our initial collection. If all the included entities of type E pass this test, then P is closed for entity type E. P is *totally closed* for a set of entity types if and only if it is closed over each type in the set.

Notice that closure can be partial in two distinct ways. First, P may be closed for only a subset of entity types we consider. Second, P may prove to be more or less "permeable" to intrusion by excluded tokens for any single entity type. For example, two nearly isolated breeding populations could produce occasional sterile hybrids, which violates closure if we count the hybrids as members of either population over time. However, the fraction of material with external origins at any point earlier in time would remain close to zero. Strictly speaking, closure fails to

---

[2] Restricting our attention to relations between entities of the same type serves as a stand-in here for a fully developed theory of inheritance. As defined, production events only track causal responsibility for the existence of an entity, not causal responsibility for its specific properties. We would need a more refined subtype of production event to express the idea that DNA sequences are copied during replication and hence exhibit a stable mapping between the properties of the original strand and its product.

hold, but quantifying the fraction of externally originating material is a way to expand the concept to allow partial closure as a matter of degree.

Closure is not sufficient to define a population lineage because the population may still contain what we would want to say are multiple, independent lineages. As de Queiroz rightly points out, clades should not count as population lineages since they have effectively split into distinct, persisting units of evolution. However, he also notes that population lineages can exhibit internal branching so long as these branches "would have to be judged as ephemeral" (de Queiroz 1999, p 53). Similarly, accounting for ephemeral branching was Kornet's motivation for expanding the internodal species concept to allow species to be composites of internodons (Kornet and McAllister 2005). The idea behind the mixture criterion is to express what it means for branches to be "ephemeral" in the right sort of way.

Typically, biologists think of a branch as something that is mutually closed relative to the rest of the population series: nothing goes in, nothing comes out. This criterion for isolation is too strong, however, because it overlooks the possibility of a series of subpopulations that is closed to outside intrusions but emits material into the rest of the population. See Fig. 2 for an example. The subpopulation series would then be connected with other individuals while being closed to any external input. In the language of graph theory, it is a "source" of material for the rest of the network without also being a "sink." Mixture should fail in this scenario because
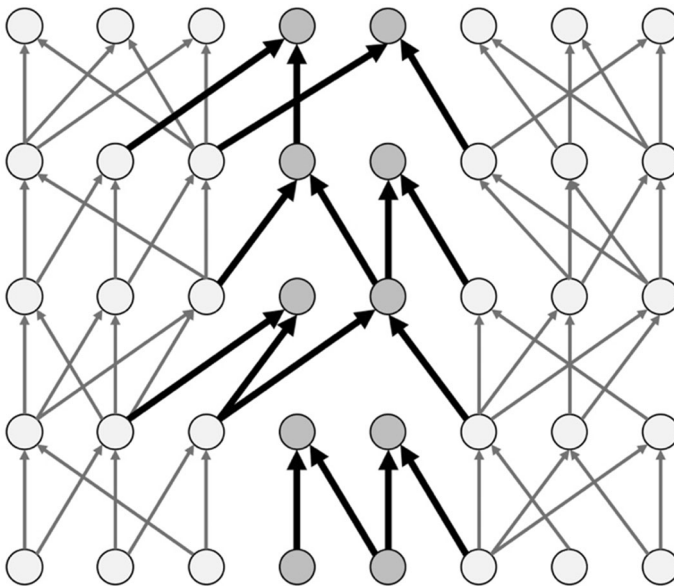

Fig. 2 Population series may contain subgroups that contribute material to the population as a whole without receiving material in turn. The two subgroups (*light grey* individuals clustered on the *left* and *right*) are connected in the sense that there are some individuals (*dark grey*) in the overall population which receive material from both subgroups (*black arrows*, difference between transfer and production events not shown). However, there is never transfer or production of material from one subgroup into the other

the closed subpopulation series does not receive any material from the rest of the population series.

Let's call this case an *insulated subgroup*. To identify insulated subgroups, we start by selecting a strict subset S of demarcators in P, at some time t, and a set of entity types E to track. Next, we collect any demarcators in next time step that only receive material of those types from the members of S, including both transfer and production relations. If the collection at time t + 1 is empty or equal to the whole population, the search halts. In this case, no insulated subgroup exists with S as its initial members. Otherwise, we repeat the same process for the next time step (t + 2) using the subset of demarcators found in t + 1. We continue until a time step returns an empty result or the whole population. The insulated subgroup is then the series of demarcators we identified at each time step. This iterative search process returns a maximal and closed subseries of P given the initial choice of subset S and entity types E.

Intuitively, if P exhibits mixture, then there should be very few insulated subgroups with durations longer than some relevant timescale T. There are multiple ways to make this criterion more precise, depending on whether we want deterministic or probabilistic conditions and whether the population series P has finite or infinite length. See Table 2 for the possible criteria we could use. The finite, probabilistic criterion is the most interesting and relevant for biological practice, but it is also the hardest to know with certainty.

There are also multiple ways to select a relevant timescale, each of which assumes different background knowledge. We can set a macroevolutionary timescale for a population series based on an estimate of the speciation rate in its clade using comparative phylogenetics (Barraclough and Nee 2001; Fontaneto and Barraclough 2015). If a population's existence is tied strongly to a particular habitat or niche, such as fish confined to a single network of caves, we could also set a meso-level timescale according to the expected longevity of the habitat or niche.

At a micro-scale, we can compare an actual population series to a neutral model of the population, i.e. where positive and negative selection are absent. Using the model, we could calculate the expected distribution of insulated subgroup durations, either analytically or numerically, and compute a threshold time that marks when the frequency of insulated subgroups lasting at least that long became sufficiently rare. For example, we could set the threshold to be the 95th percentile of insulated subgroup durations generated by a random-mating model, so that only 5% of insulated subgroups are expected to last longer. Random-mating models thus re-

Table 2 Criteria for the existence of mixture in the history of a population series P

	Finite	Infinite
Deterministic	P has no insulated subgroups of length T or longer	P contains only finite length insulated subgroups
Probabilistic	The probability of insulated subgroups T or longer is approximately zero	The probability of insulated subgroups T or longer converges to zero as T increases to infinity

enter the picture here as analytically tractable idealizations that we can use to set a timescale for mixture. However, random-mating models receive no special status as definitional for population lineages, and models that include explicit geographical factors may be more desirable when feasible. Note also that the timescale-relative nature of population lineages reflects the actual complexity of evolution, where speciation is a matter of degree and even populations isolated by millions of years may still hybridize under the appropriate circumstances (Seehausen et al. 2014; Grant and Grant 2014).

Putting everything together, a population series P is a *population lineage* for entity types E if P is closed and mixes over entities in E. Closure means that the population series never receives new material transferred from or produced by the "outside." Mixture means that "ephemeral branches" never persist indefinitely or only rarely for a long time.

Note that closure and mixture are straightforwardly satisfied in a scenario where each individual receives one chromosome from each of two randomly selected parents via sexual reproduction. No chromosomes migrate into the population, which satisfies closure. Panmixis ensures that the probability of a closed subpopulation sequence of duration T declines to zero as T increases. An actual population close to panmixis would typically satisfy mixture relative to a threshold calculated from a random mating model.

Types, causes, and identities of population lineages

How does the mixture account relate to existing views of population lineages? Compared to the paradigm processes of sex and random mating, it articulates a more general view of the structural possibilities for population lineages. As a result, the account accommodates biologically important cases that were previously excluded, marginalized, or problematic. This section examines four types of cases: populations with mosaic inheritance, limited gene flow, complex patterns of lateral transfer, and significant hybridization. I introduce each type of case through a common expectation about population lineages that is not necessarily true under the account given here.

Expectation 1: symbiotic individuals cannot form population lineages without vertical transmission

As we saw, functionalist approaches define population lineages as units of evolution. Paradigm population lineages are thus highly evolvable, but how do we tell which ones are highly evolvable? On this point Godfrey-Smith, for example, reaches to supplement his functionalist theory with knowledge about material features of individual reproduction and development (Sterner 2015). In particular, he uses three key properties of individual reproduction and development to identify paradigm population lineages: the existence of a genetic bottleneck, a germ-soma distinction, and functional integration among parts (Godfrey-Smith 2009).

Supplementing functionalism with these more particular properties runs into a problem if there are highly evolvable collections of living things that end up excluded or marginal. Ereshefsky and Pedroso (2012), for example, have argued that biofilms achieve a higher degree of individuality than one would expect on Godfrey-Smith's account because they use lateral gene transfer to produce genetic similarity instead of a bottleneck stage in the life cycle.[3] Kim Sterelny has argued a related point regarding the mutualistic symbiosis between bullhorn acacia trees (*Acacia cornigera*) and the acacia ant, *Pseudomyrmex ferruginea* (Sterelny 2011). The two species exhibit functional integration, since the trees provide food and housing for the ants while the ants defend the tree from other insects and plants, but they reproduce independently: juvenile trees do not generally inherit an ant colony from their parents, nor, vice versa, do new ant colonies establish themselves in an offspring of their former home. Godfrey-Smith argues that this symbiosis is not a Darwinian population, despite its high evolvability, and is better understood as two separate, co-evolving Darwinian populations (Godfrey-Smith 2011).

In contrast, the mixture account identifies population lineages without making assumptions about their evolvability. It is also neutral with respect to empirical claims about the importance of particular reproductive mechanisms, such as bottlenecks or germ lines, for evolvability (Clarke 2013). As a result, the demarcator formalism provides us with a platform for investigating the effects of these processes on the evolvability of population lineages through mathematical modeling or simulation. For instance, the demarcator formalism enables us to model rather than presuppose the effects of vertical and horizontal transmission on the evolvability of population lineages. Moreover, removing the dependence of population lineages on assumptions about how individuals reproduce is an important step toward generalizing evolutionary theory beyond the Modern Synthesis's emphasis on pure sexual or asexual reproduction.

For example, imagine a pair of 'species' that are obligate mutualists and form units composed of one individual from each species.[4] We can model a generation using two stages: a group stage, defined by a demarcator type that includes one individual from each species; and an individual stage, where new individuals produced by the groups briefly live on their own. Let's assume that new groups draw their members randomly from the total populations of the two species in the individual stage, and that there is no migration from or into other populations. If we define closure and mixture with respect to consecutive group-level demarcators, we will still have closure and mixture despite the absence of direct vertical transmission. We could apply the same setup to represent mosaic inheritance for parts of a multipartite virus genome (Manrubia and Lázaro 2016) or the genomes of a cell and its endosymbionts (Curtis et al. 2013).

---

[3] See (Clarke 2016) for a critical response to their argument.

[4] I put 'species' in quotes here to indicate that I'm using the term informally rather than presupposing a particular account of species.

Expectation 2: population lineages are maintained by gene flow

An evolutionary view of species understands them as distinctive natural units because each species has a common evolutionary trajectory (Simpson 1961; Hull 1976). Articulating this idea presents two immediate challenges for evolutionary biology, though: what does it mean to share an evolutionary trajectory, and how do we explain the actual existence of groups with this property? Philosophers Matthew Barker and Robert Wilson have articulated two possible strategies for explaining the evolutionary unity of species: integrative cohesion, such that interactions between many or all of a species' components facilitate their causal unification into a whole, and response cohesion, where a species' components are disposed to respond in the same way to an intervention (Barker and Wilson 2010, pp 64–65).

Historically, the presence of gene flow has been the dominant explanation among biologists for why conspecifics tend to share an evolutionary trajectory, which Barker and Wilson (2010) label "The View." As they characterize it, The View holds that species are cohesive because their members share a common evolutionary trajectory as a result of possessing similar phenotypic and genotypic properties. These similarities in turn are explained by the homogenizing effects of gene flow among populations in the species.

Barker and Wilson attack the adequacy of The View by targeting two key claims it endorses: "(1) The phenomenon that stands in need of explanation is integrative species cohesion. (2) What explains integrative species cohesion, ultimately, is gene flow via its causal influence on the response cohesion of the species population" (Barker and Wilson 2010, p 66). I agree with Barker and Wilson that integrative and response cohesion are different targets for explanation and that the empirical evidence doesn't support gene flow as a uniquely important cause of either form of cohesion. However, their criticism of The View extends further to question whether integrative cohesion even qualifies as a coherent explanandum once we recognize the limitations of gene flow.

Barker and Wilson raise several relevant points against integrative cohesion as a phenomenon in need of explanation. Typical species exhibit a "gappiness" due to geographic, ecological, behavioral, and other barriers that precludes them from being adequately integrated (Barker and Wilson 2010, p 67). As a result, what interactions do occur between conspecific individuals or populations fail to combine in a way that explains the overall behavior of the species. Moreover, weakening The View so as to simply individuate species by the presence of gene flow would make explaining their cohesion using gene flow vacuous.

The mixture criterion offers an alternative account of integrative cohesion suited to these concerns. The existence of gaps in the causal interactions among individuals is not a problem in itself—it would occur stochastically even in panmictic populations. What matters is whether gaps persist long enough to undermine the shared evolutionary trajectory of the population. How fast insulated subgroups must turn over to ensure evolutionary unity among a population is in principle an empirical problem for any set of real individuals. Moreover, the turnover rate does have implications for the homogeneity of individuals: a higher rate generally implies more sharing of material across any partition we choose of the

population, which has the same effect as gene flow in generating future cohesion. While the turnover rate has explanatory force, it is only one factor among many that may contribute to explaining why a population exhibits mixture at that rate. Integrative cohesion therefore remains a substantive target for explanation even when we characterize it in terms of mixture.

It is also important to note that mixture does not presuppose a notion of material flow from one spatial location to another. For example, mixture accommodates reproductive integration sustained by regular selective sweeps in purely asexual populations. A selective sweep occurs when a novel mutation has a sufficiently strong fitness benefit for natural selection to rapidly drive it to fixation in a population. In a purely asexual context, there are no mechanisms for hybridization, so a selective sweep effectively culls all but one reproductive lineage from the population. Since insulated subgroups must be strict subsets of the population at each time step, regular fixation events would place an upper ceiling on their duration.

A corollary benefit of this result is the unification of de Queiroz's treatment of sexual and asexual species in his general lineage concept. He notes that if sexual reproduction is the only way that individual genealogies may combine to form population lineages, then asexual reproducers are automatically excluded from forming species. "But perhaps there are other processes that unite collections of asexual organism lineages to form higher level lineages… that are comparable to those formed by sexual organism lineages in certain evolutionarily significant respects" (de Queiroz 1998, p 60). The mixture criterion provides an umbrella under which various processes, such as sexual reproduction and natural selection, may combine to generate higher-level integrative cohesion.

Expectation 3: every biological individual is a member of one and only one population lineage

This statement is sometimes taken to be an a priori condition on any species definition. For example, "we think that a species definition should and will find general acceptance only if it partitions genealogical networks of individual organisms exhaustively into mutually exclusive and historically continuous parts" (Kornet et al. 1995, p 111). Others take exclusivity to be a positive virtue for an account of population lineages, though not necessarily on a priori grounds (Millstein 2015). The existence of horizontal gene transfer should remove any apparent necessity to this view: living things can have many parents, each for different parts of their genomes, which come from very different taxa.

Imagine a group of bacteria containing multiple ecotypes that nonetheless regularly share DNA, such as for antibiotic resistance. In this hierarchically structured scenario, each ecotype contains distinctive genes that are adapted for some local niche but are deleterious for members of other ecotypes. We expect these genes, then, to be transferred only rarely across ecotypes, while the group as a whole will still share other genes that carry ecotype-independent benefits.

We can say that P is a *nested mosaic population lineage* for entity types E if:

1. It is a population lineage for a subset E_1 of types in E.
2. There is another population series Q that contains P as a subset.
3. Q is a population lineage over $E_2 = E–E_1$.

The population series P is nested because all the demarcator tokens it contains are also members of Q. However, Q can be a distinct population lineage because the production and transfer of entities in E_1 is separable from the production and transfer of entities in E_2. As a result, P can exhibit closure and mixture over E_1 even as its members are also participating in the closure and mixture of Q for E_2. This example does not exhaust the possible forms of mosaic but illustrates the formalism's capacity to identify clear population lineages where individuals are members of more than one population lineage.

Expectation 4: extensive hybridization of population lineages implies the loss of their identities

Hybridization between two populations poses a challenge for species concepts based on reproductive isolation. If two groups must be strictly isolated, then any hybridization leads the groups to lose their individual identities as units of evolution (Kornet 1993). In practice, species often seem to interbreed a moderate amount without merging into a single unit. How to draw a line, if one exists, between groups as forming two distinct units of evolution instead of one?

The question is especially urgent for understanding units of evolution in bacterial populations with high levels of lateral gene transfer. Some biologists have sought to recover hidden patterns of vertical inheritance by looking for 'core' genes that are particularly important to individual functioning, such as genes involved in protein synthesis, and are therefore less likely to be laterally transferred among divergent groups (Rivera et al. 1998; Jain et al. 1999). Laura Franklin-Hall has criticized this strategy, however, as carrying unwanted commitments to an essentialist view of species: "Because of divergences among the phylogenies of different organismal parts… there are no particular lineages [of parts] to which we can appeal when delimiting species—unless we are willing to accept that certain organism parts are essential to organism identity" (Franklin 2007, p 71).

This misconstrues what biologists are doing: they select certain genes as identifying units of evolution because their functional importance makes them very unlikely to be transferred horizontally, not because their function privileges their history above the history of other genes. We can use the demarcator formalism to define an extended notion of population lineage in order to show how which entity types end up preserving the population lineage's identity is a contingent and empirical matter of fact.

P is an *extended population lineage* if (1) P contains exactly one insulated subgroup that persists and exhibits mixture, and (2) P is connected. Now we need to define what it means for a insulated subgroup to persist and be connected. For entity types E, a insulated subgroup H *persists* from time t_0 to t_1 if:

1. H is closed over E,

Table 3 Criteria for whether the history of a population series P is connected

	Finite	Infinite
Deterministic	P has no splits of length T or longer	P contains only finite splits
Probabilistic	The probability of splits T or longer is approximately zero	The probability of splits T or longer converges to zero as T increases to infinity

2. H existed in P at time t_0-1,
3. H exists at each time step from t_0 to t_1.

Now let's define *connected*. I will do this in an analogous manner to mixture: first we'll define what a split is in a population series and use the distribution of splits to specify when the population series is connected. A population sequence P *contains a split* over E if there exists a series of strict subsets S_t of P_t for all $t = 1$ to T such that both S_t and its complement are closed for E. We can give an analogous table for P being connected as we did for P exhibiting mixture (see Table 3).

How does this address the issue of identity in the face of hybridization? Imagine we have two population lineages that start off independently but begin to exchange genes. Some individuals descended from each population over time may never participate in material transfer, and some gene types may never be transferred between the descendants of each population lineage. Minimally, then, we can look for a persistent part of a population lineage that is closed for at least some genes. In particular, a persistent insulated subgroup that exhibits mixture maintains the key nature of a population lineage (closure and mixture) while also preserving its genealogical identity in terms of material transfer and production (closure and persistence). A population lineage is then extended in the sense that it can contain members outside the persistent insulated subgroup but remain connected to it by regularly receiving material entities. In practice, the entity types in E that characterize the persistent insulated subgroup can be determined empirically by inspecting or inferring their genealogical histories rather than selecting the entity types in advance.

Conclusion

The mixture account of population lineages offers a new, structural approach to defining population lineages. It breaks with traditional conceptions by placing no special importance on sexual reproduction or random mating. Instead, the account characterizes a population lineage in terms of its closure to external inputs and the frequent turnover of subgroups that are closed to the rest of the population. By relying on pattern instead of process, the account generalizes what it means for a population to be reproductively integrated over time. Many cases that have proved problematic under existing theories of population lineages, such as symbiosis and lateral gene transfer, can be straightforwardly accommodated under the mixture account.

The mixture account also has implications for modeling practices in population genetics, especially landscape genetics (Manel et al. 2003; Landguth et al. 2015), statistical phylogeography (Knowles 2009), and coalescent-based species delimitation (Liu et al. 2009). Since the account defines population lineages in terms of pattern rather than process, we can separate the existence of a population lineage from any single model of mating interactions. The account therefore transforms the relationship between random mating models and population lineages from being definitional to empirical. In other words, we can now investigate empirically how non-trivial population structure affects the rate of mixture among individuals by measuring changes to the distribution of insulated subgroup durations.

Coalescent-based models for species delimitation, for example, rely on the assumption of random mating to detect when the history of a gene diverges from what we expect for a single interbreeding population (Liu et al. 2009). The number of genes that diverge from this expectation in turn provides support for a branching event in phylogenetic history. One problem with this assumption is that actual populations often have substantial internal structure, caused for instance by heterogeneous ecological landscapes, that can bias estimates of their divergence. Another issue is that even truly random mating in a single population can generate outcomes that appear to be better explained by the existence of multiple isolated populations. This false result becomes more probable for shorter timescales and smaller populations. Coalescent-based species delimitation, however, lacks a principled way to place a lower cutoff on when divergences in gene histories are reliable evidence of branching events. In these regards, the mixture account offers new avenues for unpacking the effects that random mating has as an idealization on the empirical adequacy of population genetic models.

References

Avers CJ (1989) Process and pattern in evolution. Oxford University Press, New York

Bapteste E, Dupré J (2013) Towards a processual microbial ontology. Biol Philos 28(2):379–404

Barker MJ, Wilson RA (2010) Cohesion, gene flow, and the nature of species. J Philos 107(2):61–79

Barraclough TG, Nee S (2001) Phylogenetics and speciation. Trends Ecol Evol 16(7):391–399

Bouchard F (2008) Causal processes, fitness, and the differential persistence of lineages. Philos Sci 75(5):560–570

Boucher SC (2015) Functionalism and structuralism as philosophical stances: Van Fraassen meets the philosophy of biology. Biol Philos 30(3):383–403

Brigandt I (2017) Bodily parts in the structure-function dialectic. In: Lidgard S, Nyhart LK (eds) Biological individuality: integrating scientific, philosophical, and historical perspectives. University of Chicago Press, Chicago, pp 249–274

Chiu L, Gilbert SF (2015) The birth of the holobiont. Biosemiotics 8(2):191–210

Church SA, Taylor DR (2002) The evolution of reproductive isolation in spatially structured populations. Evolution 56(9):1859–1862

Clarke E (2013) The multiple realizability of biological individuals. J Philos CX 8:413–435

Clarke E (2016) Levels of selection in biofilms. Biol Philos 31:191–212

Curtis BA, Tanifuji G, Burki F et al (2013) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492(7427):59–65

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. In: Howard DJ, Berlocher SH (eds) Endless forms: species and speciation. Oxford University Press, New York, pp 57–75

de Queiroz K (1999) The general lineage concept of species and the defining properties of the species category. In: Wilson RA (ed) Species: new interdisciplinary essays. MIT Press, Cambridge, pp 49–89

Doolittle WF (2010) Eradicating typological thinking in prokaryotic systematics and evolution. Cold Spring Harb Symp Quant Biol 74:197–204

Doolittle WF, Booth A (2016) It's the song, not the singer. Biol Philos 32(1):5–24

Doolittle WF, Zhaxybayeva O (2010) Metagenomics and the units of biological organization. Bioscience 60(2):102–112

Ereshefsky M, Pedroso M (2012) Biological individuality: the case of biofilms. Biol Philos 28(2):331–349

Fontaneto D, Barraclough TG (2015) Do species exist in asexuals? Theory and evidence from Bdelloid rotifers. Integr Comp Biol 55(2):253–263

Franklin LR (2007) Bacteria, sex, and systematics. Philos Sci 74(1):69–95

Godfrey-Smith P (2009) Darwinian populations and natural selection. Oxford University Press, Oxford

Godfrey-Smith P (2011) Agents and acacias: replies to Dennett, Sterelny, and Queller. Biol Philos 26(4):501–515

Godfrey-Smith P (2016) Individuality, subjectivity, and minimal cognition. Biol Philos 31(6):775–796

Grant PR, Grant BR (2014) Evolutionary biology: speciation undone. Nature 507(7491):178–179

Griesemer JR (2000) Reproduction and the reduction of genetics. In: Beurton P, Falk R, Rheinberger HJ (eds) The concept of the gene in development and evolution. Cambridge University Press, Cambridge, pp 240–285

Griesemer JR (2005) The informational gene and the substantial body. In: Jones MR, Cartwright N (eds) Idealization XII: correcting the model. Rodopi, New York, pp 59–115

Griesemer JR (2014a) Reproduction and scaffolded developmental processes. In: Minelli A, Pradeu T (eds) Towards a theory of development. Oxford University Press, New York, pp 183–202

Griesemer JR (2014b) Reproduction and the scaffolded development of hybrids. In: Caporael LR, Griesemer JR, Wimsatt WC (eds) Developing scaffolds in evolution, culture, and cognition. MIT Press, Cambridge, pp 23–56

Hull DL (1976) Are species really individuals? Syst Zool 25(2):174–191

Hull DL (1980) Individuality and selection. Ann Rev Ecol Syst 11(1):311–332

Jablonka E, Lamb MJ (2005) Evolution in four dimensions. MIT Press, Cambridge

Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci 96(7):3801–3806

Kendig C (2014) Towards a multidimensional metaconception of species. Ratio 27(2):155–172

Knowles LL (2009) Statistical phylogeography. Ann Rev Ecol Evol Syst 40(1):593–612

Kornet DJ (1993) Permanent splits as speciation events: a formal reconstruction of the internodal species concept. J Theor Biol 164(4):407–435. doi:10.1006/jtbi.1993.1164

Kornet DJ, McAllister JW (2005) The composite species concept. In: Reydon TAC, Hemerik L (eds) Current themes in theoretical biology. Springer, Dordrecht, pp 95–127

Kornet DJ, Metz JAJ, Schellinx HAJM (1995) Internodons as equivalence classes in genealogical networks. J Math Biol 34(1):110–122

Landguth E, Cushman SA, Balkenhol N (2015) Simulation modeling in landscape genetics. In: Balkenhol N, Cushman SA, Storfer A, Waits L (eds) landscape genetics. Wiley, Chichester, pp 101–113

Laubichler MD, Maienschein J (eds) (2007) From embryology to evo-devo. MIT Press, Cambridge

Lewontin RC (1970) The units of selection. Ann Rev Ecol Syst 1:1–18

Liu L, Yu L, Kubatko L et al (2009) Coalescent methods for estimating phylogenetic trees. Mol Phylogenet Evol 53(1):320–328

Mallet J, Meyer A, Nosil P, Feder JL (2009) Space, sympatry and speciation. J Evol Biol 22(11):2332–2341

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18(4):189–197

Manrubia S, Lázaro E (2016) Getting to know viral evolutionary strategies. In: Domingo E, Schuster P (eds) Quasispecies: from theory to experimental systems. Springer, Berlin, pp 201–217

Matthewson J (2015) Defining paradigm Darwinian populations. Philos Sci 82(2):178–197

Millstein RL (2009) Populations as individuals. Biol Theory 4(3):267–273

Millstein RL (2010) The concepts of population and metapopulation in evolutionary biology and ecology. In: Bell MA, Futuyma DJ, Eanes WF, Levinton JS (eds) Evolution since Darwin. Sinauer, Sunderland, pp 61–86

Millstein RL (2015) Thinking about populations and races in time. Stud His Philos Biol Biomed Sci 52:5–11

Odling-Smee J, Laland KN, Feldman MW (2003) Niche construction. Princeton University Press, Princeton

O'Malley MA (2014) Philosophy of microbiology. Cambridge University Press, Cambridge

Rivera MC, Jain R, Moore JE, Lake JA (1998) Genomic evidence for two functionally distinct gene classes. Proc Natl Acad Sci 95(11):6239–6244

Seehausen O, Butlin RK, Keller I et al (2014) Genomics and the origin of species. Nat Rev Genet 15(3):176–192

Simpson GG (1961) Principles of animal taxonomy. Columbia University Press, New York

Sterelny K (2011) Darwinian spaces: Peter Godfrey-Smith on selection and evolution. Biol Philos 26(4):489–500

Sterner B (2015) Pathways to pluralism about biological individuality. Biol Philos 30(5):609–628

Sterner B (2017) Individuality and the control of life cycles. In: Lidgard S, Nyhart LK (eds) Biological individuality: integrating scientific, philosophical, and historical perspectives. University of Chicago Press, Chicago, pp 84–108

Wakeley J (2009) Coalescent theory. Roberts & Company Publishers, Greenwood Village CO