1 Defining Microbiome Function

2

3 Jonathan L Klassen *

4 Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

5

6 * Correspondence:

7 Dr. Jonathan L Klassen

8 jonathan.klassen@uconn.edu

Abstract

Why does a microbe associate with a host? What function does it perform? Such questions are difficult to unequivocally address and remain hotly debated. This is partially because scientists often use different philosophical definitions of "function" ambiguously and interchangeably, as exemplified by the controversy surrounding the ENCODE project. Here, I argue that research studying host-associated microbial communities and their genomes (i.e., microbiomes) faces similar pitfalls and that unclear or misapplied conceptions of function underpin many controversies in this field. In particular, experiments that support phenomenological models of function can inappropriately be used to support functional models that instead require specific measurements of evolutionary selection. Microbiome research also requires uniquely clear definitions of "who the function is for", in contrast to most single-organism systems where this is implicit. I illustrate how obscuring either of these issues can lead to substantial confusion and misinterpretation of microbiome function, using the varied conceptions of the holobiont as a current and cogent example. Using clear functional definitions and appropriate types of evidence are essential to effectively communicate microbiome research and foster host health.

Introduction

Microbiology has been revolutionized by observation that most macroorganisms are colonized by diverse communities of microbes, prompting many studies that seek to understand why such diversity exists[1]. Many initial studies of these microbial communities were primarily descriptive, prompted by the availability of new tools that revealed a much greater diversity of symbionts than was evident from using older culture-based approaches. This research paradigm has been astoundingly successful and continues to reveal the immense taxonomic and genomic diversity of the microbial world (i.e., the microbiome). However, describing microbial diversity is only the first stage of a broader program to understand not only which microbes have colonized a host but also their rationale for being there, i.e., their function[2]. The importance of this second priority is emphasized by a Google Scholar search for "microbiome AND function" that yielded ~109,000 results (searched July 11, 2017). Defining function is therefore a key goal of microbiome research.

All discussions of an entity's function seek to understand the rationale for that entity occurring at a particular place and time. Studies of microbiome function therefore attempt to describe the significance of an association between microbes and a symbiotic host, focusing on host and microbial traits as the entities that bear such functions. Classically, microbiologists have used phenotypic tests to identify traits in cultured microbes. This approach has recently been extended using (meta)genomics to characterize genes within a host or microbiome as trait-encoding entities that might bear functions[3]. Unfortunately, these methods alone cannot provide a complete rationale for a microbe existing within a microbial community because they cannot entirely describe why a particular trait exists at a given time and place. Such "why" questions can confusingly be answered in multiple ways, causing "function" to mean different things to

different people depending on their perspective[4]. This ambiguity can cause different meanings of function to be inappropriately interchanged, leading to false claims about why microbes colonize their hosts and the consequences of these relationships.

In this essay, I will distinguish between different definitions of "function" and apply these definitions to microbiome research. Although I will primarily use examples from the human gut microbiome, these concepts can be applied to any host-microbe symbiosis. I will also discuss the consequences of confusing different definitions of function and argue that conceptual precision is crucial to avoid misdirected microbiome research.

The Multiple Meanings of "Function"

For over 40 years, philosophers have debated how to define the term "function" in ways that are both logically robust and that match how biologists actually use this term[4–10]. These debates have produced two unique and non-overlapping definitions of biological function, typically labelled as "causal role" (CR) and "selected effect" (SE) functions, respectively (Figure 1). Philosophers agree that both CR and SE definitions of function are valid and reflect different conceptions of function that predominate in different biological fields[4]. Microbiome researchers must therefore be aware of these philosophical distinctions to avoid unintentionally confusing CR and SE functions when describing host-microbe symbioses.

A trait's CR function is defined by how a larger system changes when that trait is removed, analogous to how an electrical component is defined in a circuit diagram[4,6]. For example, pumping blood is a CR function of the mammalian heart because heart failure stops circulation (the larger system). Loss-of-function experiments such as gene knock-outs and amino acid substitutions are common methods that microbiologists and molecular biologists use in a

similar manner to identify CR functions by observing how an organism or protein changes when one of its parts is modified.

Importantly, CR functions depend strongly on the nature of the system in which they are defined. For example, making a thumping noise is a CR function of the mammalian heart (in addition to pumping blood) because disrupting this sound will corrupt a system used to determine whether or not a mammal is alive. This does not supersede the heart's CR function of pumping blood, but rather allows multiple CR functions to be valid depending on the frame of reference in which they are considered (here, blood circulation versus a diagnostic test for being alive). Such frames of reference are less ambiguous for molecular biology experiments such as gene knock-outs, where the entire organism clearly comprises the relevant frame of reference. In summary, Causal Role functions are strictly phenomenological and mechanistically describe how parts contribute to a larger system that can be defined in multiple, non-exclusive ways.

In contrast to CR functions, the SE function of a trait is defined by the evolutionary rationale for that trait being maintained in an organism over time via selection[5,7,8]. For example, the SE function of the mammalian heart is to pump blood because heart failure causes a strong reduction in mammal fitness that is subject to negative selection. This contrasts with other attributes of the heart that do not impact fitness, such as making a thumping noise. Similarly, defining the SE function of a gene within a microbial genome requires understanding why selection has maintained that gene within that microbe. Consider a conserved gene that encodes for a cellulase enzyme. The conservation of this gene and its corresponding biochemical activity in both the native host and its relatives indicates that selection has acted on this gene to provide glucose for these hosts as an SE function. However, glucose provision would not be a SE function of this gene immediately after it is cloned into a heterologous host because evolution

has had no opportunity to select for this function in the heterologous host. This remains true even if the cellulase has measurable biochemical activity because it is selection, not activity, that defines an SE function. As another example, a gene that is no longer used by a pathogen undergoing genome reduction does not provide a SE function for that pathogen because this gene is no longer under selection, even if it has not yet been purged from the pathogen's genome. A trait with an SE function therefore both has been and currently remains adaptive for its host because the continued existence of that trait is due to evolutionary selection acting on fitness benefits that this trait confers upon its host[11].

Because defining an SE function requires measuring active selection on a trait, such traits must be expressed and cause a phenotype on which selection can act. All SE functions are therefore also CR functions that can be analyzed phenomenologically following the CR paradigm (Figure 1). This can be seen in the above example where a cellulase in its native host exhibits both biochemical activity (a measurement of CR function) and evolutionary conservation (a measurement of SE function). Active selection also implies that traits with SE functions are under *direct* selection, versus *indirect* selection caused by covariance with a second trait under direct selection[12,13]. This can be understood by comparing the consequences of removing a target of direct vs. indirect selection, i.e., analyzing their respective CR functions. Imagine a non-essential gene (*geneA*) whose chromosomal location is immediately adjacent to a second gene (*geneB*) that is essential for host survival. Selection will directly act on *geneB* because it is essential. Selection may also act indirectly on *geneA* because recombination involving *geneA* risks disrupting the essential *geneB*. Stated differently, variation in the trait encoded by *geneA* correlates with host fitness (as expected for an SE function) only because *geneA* is chromosomally linked to *geneB*, the actual target of direct selection. Following the

logic used to identify CR functions, a precise knockout of *geneA* without any polar effects will not affect the function of *geneB.* However, a knockout of *geneB* will be lethal. Only the trait encoded by *geneB* therefore has both SE and CR functions, whereas the trait encoded by *geneA* lacks a CR function and therefore also a true SE function, despite its genetic linkage to *geneB*. This example highlights how SE functions can only be defined by measuring direct selection on a trait, which satisfies both the CR and SE definitions of function.

There is therefore an asymmetry between CR and SE functions: although all SE functions are also CR functions, the converse is untrue (Figure 1). Indeed, CR functions are useful for exactly this reason, and can be applied in situations where developmental complexity and/or epistasis makes precise measurement of selection difficult. This asymmetry has led some evolutionary biologists to emphasize the importance of SE functions over CR functions, e.g., to avoid "spandrel" traits that originated as unselected byproducts of processes unrelated to those traits' current CR functions[14]. In an alternative approach, other biologists consider evolutionary considerations to be completely separate from mechanistic ones and only label mechanisms as functions[15,16]. These differences have led to substantial ambiguity and controversy regarding how to interpret experiments that describe trait function (e.g., Box 1). Clear definitions of CR and SE functions are therefore needed to avoid logical fallacies and wasted scientific effort, including in microbiome research.

Causal Role Functions of the Microbiome

Causal Role functions conceptualize the function of a trait as what happens to a system when that trait is removed[6]. When a gene's CR function is determined by deleting that gene and observing changes in host phenotype, the host organism defines a system of which the studied gene is a part. Similarly, many microbiome studies compare the phenotypes of symbiotic hosts

that possess a microbiome to axenic hosts that do not[17] and thereby consider the host as a system of which the microbiome is a part. These phenotypic differences between colonized and axenic hosts define the CR functions of the microbiome for that host[6]. The CR functions of a single microbial gene or species can be determined similarly by comparing the phenotypes of hosts that contain those genes or species to those that lack them (e.g., ref 18). Given the widespread nature of such phenotypes, nearly all microbiomes can be said to provide CR functions for their symbiotic hosts[1].

This conception of microbiomes as providing CR functions for their hosts treats microbes as component parts of their host. Indeed, there is a strong parallel between defining CR function by removing a mouse's heart and defining CR function by removing that mouse's microbiome, leading some to speculate that the microbiome can be thought of as a neglected host "organ"[19]. However, such a definition should be used cautiously given that the microbiome includes cells and genomes from multiple species instead of from a single species as in the traditional definition of an organ[20]. The CR approach to function makes no statement about microbiome assembly, persistence, or prevalence in other environments – all that matters is that the host phenotype changes when microbes are removed. It is therefore valid (if not always relevant) to conceive of symbiotic hosts as "holobionts" where the microbiome performs CR functions for a host as part of a single system that contains both the host and its microbial symbionts[21]. Structural definitions that treat symbiotic microbes as an "organ-like" component of their host or as part of a holobiont that is centered on the host are therefore consistent with the logical structures used by philosophers to define CR functions.

The above discussion has focused on the host as a larger system in which microbes are components with CR functions, mirroring the philosophical tradition of considering traits as

functional parts of larger organisms[4,6]. However, microbes are autonomous entities that might themselves be the beneficiaries of functions, analogous to selfish genetic elements that reside within a host genome[9]. For example, removing a host from an obligate intracellular microbial symbiont would cause that symbiont to become non-viable, giving this host the CR function of providing a home for that microbe. Although logical, this situation is likely to be poorly generalizable because most symbiotic microbes are also viable without their hosts[22] and can display identical phenotypes in both host and non-host environments (e.g., anaerobiosis). Thus, a "holobiont" centered on a microbe instead of a host may be logically valid but of limited use except for highly-intertwined relationships, e.g., obligate intracellular symbionts and pathogens[23].

In summary, microbiomes and their constituent parts provide CR functions to their hosts that are defined by how host phenotypes change when microbes or their parts are removed. Such CR functions of a microbiome are consistent with microbiomes being "organ-like" entities that are part of a larger system defined by the host, i.e., a holobiont. In some cases hosts may also be conceived of as providing CR functions for a microbe but the generalizability of this process remains unclear.

Selected Effect Functions of the Microbiome

The Selected Effect function of a trait is defined by why that trait exists as an adaptation for its host[5,7,8]. This contrasts with CR functions that describe mechanisms but not how those mechanisms came to exist or if they are adaptive. Selection on a trait is measured as the correlation between the fitness that a trait confers upon a host and that host's reproductive success, such that adaptive traits promote host fitness and therefore increase the absolute abundance of that host over time (Box 2). Microbiome experiments often work similarly by

measuring changes in microbial community composition that correlate with changes in host health. This assumes that microbial community composition can be considered a trait that is possessed by the host, and treats host health as a surrogate for reproductive fitness (which can be difficult to measure, especially for long-lived hosts such as humans). Whether these assumptions hold may be system-specific and require explicit tests for validation. Furthermore, whether microbes themselves should most appropriately be conceived of as traits that are possessed by the host or a means to realize some other trait such as nutrient acquisition remains unclear.

Mindful of these caveats, how might we measure selection on traits encoded by the microbiome as extended phenotypes of their host (i.e., selection acting on the host to maintain traits that are provided via its microbiome, regardless of selection on the microbes)? Selection depends both on how trait values change between host generations and the fidelity of intergenerational trait transmission (Box 2). Host mechanisms that maintain host-microbe interactions leading to the persistence of traits that are provided via the microbiome will therefore be selected if the improvements to host fitness that are provided by these traits outweigh the effects of imperfect microbial transmission between host generations, such as during horizontal or mixed-mode transmission[24]. Microbiome variability within a host's lifespan[25] can further disrupt the heritability of traits that are provided via the microbiome and weaken the potential for selection to act on these traits. On the other hand, the high redundancy of traits within a microbiome[26] may allow traits to be heritable without vertical transmission if hosts can continuously acquire microbes that provide the same trait via cultural practices such as cohabitation with family members and/or maintaining a consistent living environment[27]. Horizontal gene transfer is also common in host-microbe symbioses and might be another means for traits to be maintained in a microbiome without strict vertical transmission of symbionts[28].

208 Determining the extent to which selection on the host can include traits that are provided via the
209 microbiome despite imperfect transmission between host generations will be a fruitful area for
210 future research.

211 Even if we accept that selection on hosts can include traits that are provided via the
212 microbiome, disentangling direct and indirect modes of selection (that is, selection acting
213 directly on a trait or indirectly on some second trait that is linked to the first; see above) remains
214 problematic because traits that are provided via the microbiome can elicit multiple host
215 phenotypes. For example, butyrate produced by mammalian microbiomes both provides energy
216 for the host and regulates pathways that maintain low levels of nitrate and oxygen in the gut
217 lumen[29]. In this example, host selection might act on the accumulation of butyrate as an energy
218 source (regardless of its originating from microbial metabolism or some other source), the
219 presence of particular microbes in the host gut to provide butyrate, or the presence of a particular
220 host signaling pathway that has the side effect of enriching for butyrate-producing microbes.
221 Here, the evolutionary path that was followed to achieve the present state remains unclear. Such
222 mechanistic complexity in host-microbiome systems therefore makes it difficult to acquire
223 evidence for direct selection acting on any particular trait, as required to define SE functions that
224 are provided via the microbiome.

225 The preceding paragraphs (and the corresponding equations in Box 2) consider selection
226 acting on a host that possesses traits that are provided via a microbiome. However, other possible
227 targets of selection often exist that must be distinguished from selection acting on host traits that
228 include the microbiome. Consider a pilus that promotes microbial adherence to a host. Selection
229 may act on this pilus to maintain an interaction between a specific host-microbe pair, giving the
230 pilus an SE function of mediating this specific host-microbe interaction. The host is not under

231 selection to maintain this interaction (selection is acting at the level of the microbe, of which the
232 pilus is a part) but is instead an indispensable part of the interaction trait, i.e., without the host
233 there is no possibility of an interaction and therefore no selection on the pilus as a trait to
234 maintain that interaction. Thus, the host is in some sense a part of an SE trait of the microbe
235 because selection is acting on an interaction that includes both partners, i.e., the host provides the
236 necessary context in which selection on the microbe acts[23]. Other forces might alternatively
237 explain the existence of this pilus without any host involvement. For example, the high
238 replication rate of a microbe relative to its host might select for microbial adherence to its host to
239 out-compete non-adherent microbes that are washed away by flow through the gut. This selective
240 advantage would only occur in the presence of flow and would exist in the guts of alternative
241 hosts and/or equivalent non-host environments where similar flows occur. Here, the host only
242 provides an environmental context for pilus function and is not part of a specific interaction on
243 which selection acts. Finally, a microbe might spend a considerable portion of its life cycle
244 outside of a host such that selection acts primarily in that context[22], e.g., a pilus that is under
245 selection to adhere to an abiotic surface may also incidentally adhere to a host. Here, the host is
246 entirely dispensable and plays no role in selection on the pilus. Determining when a microbiome-
247 encoded trait is under selection as part of a microbe or a host, and the potential overlap between
248 these modes of selection, remain significant challenges when identifying traits that provide SE
249 functions in host-microbe symbioses.

250 In summary, traits that are provided via the microbiome might provide SE functions for
251 their hosts depending on the fitness effects that they confer on these hosts and their heritability
252 between host generations. Host-microbe symbioses can therefore be considered using standard
253 evolutionary models that are agnostic to transmission mode, similar to traits that are transmitted

culturally[13,30]. Note that this discussion only considers selection acting at the level of hosts or microbes individually, and that additional tests are required to identify more integrated modes of selection[13]. Although imperfect symbiont transmission may obstruct host selection, this may be offset by trait redundancy within a microbiome. Critically, such functional models must differentiate between microbial traits that provide SE functions for the microbe versus those that provide SE functions for the host[23]. Experiments are currently lacking that explicitly measure selection, determine the targets of such selection, and contrast the strength of such selection to non-adaptive forces that may alternatively drive host-microbe relationships[31]. Some studies have used a constant laboratory environment to demonstrate the potential for selection to act on the host during such relationships[32–34], and the next step will be to determine if such selection also occurs in the wild and its importance relative to non-adaptive forces. However, the complexity of such measurements may make it difficult to precisely and unambiguously identify SE functions in host-microbe symbioses, even if these microbiomes clearly provide their hosts with CR functions. Microbiome research may instead need to follow other fields such as anatomy that primarily use CR definitions of function because it remains intractable to demonstrate specific SE functions in such complex systems[4].

Is "Function" Misused in Microbiome Research, and does it Matter?

Because "function" has different and non-synonymous meanings in biology, there is a high potential to confuse functional definitions and inappropriately conclude that an entity has a SE function based on phenomenological evidence that can only be used to define a CR function. This is exemplified by the controversy surrounding the ENCODE project (Box 1). Although controversies of a similar magnitude have yet to erupt among microbiologists, CR and SE functions may still be similarly confused in microbiome research. For example, the hologenome

concept of evolution rightly describes how microbes modulate many host phenotypes as CR functions[20,21,35,36], and how symbiotic partners need to be considered when describing the evolution of host or microbial traits that mediate symbiotic interactions[20,23]. However, caution must be exercised when considering more complicated models of SE function (such as those that treat the holobiont/hologenome as a distinct level of selection[36,37]) without explicitly disambiguating direct and indirect selection resulting from linkage to other traits and/or lower-levels of selection[13,38,39]. Hypotheses of host-microbe coevolution may be a second area where CR functions are frequently misappropriated as SE functions. Although microbes clearly possess many traits with CR functions that allow them to co-exist with symbiotic hosts, evidence for the precisely defined SE functions of these traits on which selection acts remains minimal. Because evidence for the reciprocal evolution of host traits with SE functions that maintain symbiotic relationships (as necessitated by coevolution[40]) is even sparser, the widespread assertion that microbes and their hosts have coevolved may be another example of valid CR functions being misconstrued as SE functions without appropriate supporting evidence. Thus, examples exist where microbiome research at least implicitly confuses CR and SE definitions of function.

I suggest that confusing CR and SE functions of the microbiome matters for three reasons. First, using imprecise definitions of function weakens communication between researchers because different people use the same term to mean different things. This is particularly true when attempting to test precisely defined models, because different data are required to falsify different types of functional hypotheses (e.g., phenotypic vs. evolutionary). Without precise definitions, it can be too easy to test a functional hypothesis using experiments that are logically incompatible with that hypothesis, leading to confusion that inhibits progress. Second, using imprecise definitions of function weakens public trust in science by obscuring the

evidence-based link between theory and interpretation. This is particularly true in evolutionary biology, where some special interest groups are quick to exploit weak evolutionary claims, particularly those that involve humans (and by extension, their microbiomes). Microbiome researchers should therefore take care to avoid such vulnerabilities. Finally, confusing CR and SE functions might lead to the misapplication of therapeutics designed to alter host health via the microbiome. For example, the consequences of potentially losing human gut microbial diversity during industrialization[41] are likely to be much more severe if that diversity provides SE functions for the host versus CR functions, because only the former will necessarily have heritable fitness consequences for the host. Similarly, supplying traits that benefit a host via probiotics will likely require different degrees of personalization to achieve stable host colonization depending on whether those traits provide CR or SE functions and whether those SE functions are directed towards the host or the microbe. Specific definitions of function are therefore practically needed to understand the long term consequences of microbiome changes for the host and how any potential consequences of these changes might be modified.

In conclusion, microbiome researchers have much to gain from the extensive philosophical research that concretely defines the various meanings of the term "function". Current practices in microbiome research can easily conflate CR and SE functions, thereby creating explanatory models of the microbiome that are based on weak evidence. Such misconstructions have acute implications for communicating microbiome research and guiding efforts to improve host health. I therefore join others[42] in advocating for stronger cross-disciplinary training in the basics of logic and philosophy as an essential means to advance rigorous and reproducible microbiome research.

Box 1: Function and the ENCODE project

In a technical tour-de-force, the Encyclopedia of DNA Elements (ENCODE) project consortium deployed a wide variety of methods to characterize every functional element present in the human genome, defining "functional element" as "a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA) or displays a reproducible biochemical signature (for example, protein binding, or a specific chromatin structure)"[43]. This is clearly a CR approach to function where the identified genomic regions have functions relating to biochemical activities such as transcription or protein binding. The ENCODE authors found that >80% of the human genome was "functional" in this CR sense[43], in contrast to a large body of literature indicating that at most 25% (and probably much less) of the human genome is under selection and therefore has an SE function[10,44]. Thus, two different philosophical approaches generated wildly different estimates of what percentage of the human genome was, in some sense, functional.

Unfortunately, the ENCODE authors did not explicitly differentiate between CR and SE approaches to function, and instead implied that the >80% of the human genome that they described as functional (in the CR sense) superseded previous estimates measuring the extent of SE function based on natural selection[43]. This elicited a strong reaction from evolutionary biologists, who clearly differentiated between these different philosophical definitions and advocated for the primacy of SE approaches[9,10,45–47] (see also ref 48 for a response). These evolutionary biologists considered most of the biochemical events observed by the ENCODE team to be incidental in nature and/or of little value to the survival and reproduction of the host, and concluded that the observed CR functions could not supersede previous perspectives based on SE definitions of function.

346 The analyses produced by the ENCODE consortium and the responses that they
347 generated highlight how poorly differentiating between CR and SE functions can obscure
348 scientific insight and lead to logical confusion that inhibits research advances. Clearly, the
349 ENCODE project identified many genuine biochemical activities, and the reasons why these
350 activities occur demand further explanation. However, it is a logical fallacy to infer that these CR
351 functions exist because they provide some benefit to their host as selected effects, at least
352 without further experiments designed to specifically test this hypothesis. Put another way,
353 different experimental parameters need to be explicitly measured to identify SE functions (e.g.,
354 the intensity of natural selection) versus CR functions (e.g., a biochemical activity), and these
355 measurements cannot be substituted for each other. Doing otherwise can lead to conclusions that
356 are, at best, logically tenuous or, at worst, incorrect.

Box 2. Measuring selection

Although natural selection is often discussed by microbiome researchers, only rarely is the strength of such selection actually measured. However, methods to measure selection are readily available and commonly applied throughout evolutionary biology. Among several approaches, one of the most prevalent uses the Price equation[49] to describe how trait or gene frequency changes over time:

$$\Delta\bar{z} = Cov(\omega, z') + E(\Delta z) \quad (1)$$

(The formulation using relative fitness is shown — see ref 13 for other variants.) Here, $\Delta\bar{z}$ represents the change in the average value of a trait between generations, $Cov(\omega, z')$ describes the covaration between relative host fitness (ω) and the average trait values possessed by its offspring (z'), and $E(\Delta z)$ describes the intergenerational change of this trait that is not due to selection. Phrased differently, $Cov(\omega, z')$ represents the intergenerational change of a trait due to selection and $E(\Delta z)$ represents the extent of biased transmission of this trait between host generations. The Price equation is therefore a concise method of partitioning intergenerational trait variation into one component that depends on selection and another that does not. This approach has been recognized as being particularly relevant for analyzing the eco-evolution of host-microbe symbioses[50].

As a simplistic example of this approach, consider an experiment that tests if a host associates with a microbe due to selection on the host to maintain that interaction or due to non-adaptive forces that promote microbial colonization regardless of selection. Following Equation 1, this experiment might measure the abundance of the target microbe in hosts from multiple generations to derive the average change in that microbe's abundance between host generations

($\Delta\bar{z}$) and the abundance of that microbe in host offspring (z'). By also measuring the fitness of these host offspring relative to their parent hosts (ω, e.g., by comparing their relative fecundity), the extent to which selection acts on the host to maintain the targeted host-microbe relationship ($Cov(\omega, z')$) can be quantified. Finally, the importance of non-adaptive factors can be described by subtracting $Cov(\omega, z')$ from $\Delta\bar{z}$. This approach will be strongest when multiple host generations can be observed so that $Cov(\omega, z')$ can be estimated accurately, highlighting the usefulness of model hosts with relatively short lifespans (e.g., insects) for such experiments. It is also worth stressing that measuring fitness is only a part of measuring selection and not synonymous with it, as is sometimes inappropriately assumed. Although this example considers microbial abundance as a trait of a host, other variants can easily be envisioned using genes as the focal trait or microbes as the host organism. Such variation will undoubtedly leverage the many variants of the Price equation approach that have been deployed throughout eco-evolutionary research (e.g., ref 51).

Because z' describes the average trait values possessed by a host's offspring, the Price approach to measuring selection is explicitly intergenerational and thus directly relates to the heritability of the considered traits, i.e., the correlation between trait values possessed by offspring and those possessed by their parents. Importantly, this approach is formally agnostic to the origin of these traits (host or microbial) or whether the same trait is shared between different microbes. However, it does require defining a single host whose fitness is altered by the trait under consideration, i.e., either a host or a microbe. Extensions of this approach to include multiple levels of selection exist but require discriminating between direct selection acting at one level and indirect selection acting at one level that is caused by direct selection acting at another

401 level[13]. Approaches such as the Price equation are therefore important frameworks for measuring
402 and modeling selection that can accommodate both host- and microbiome-encoded traits.

403 An alternative approach to identifying selection is to observe the historical effects of such
404 evolutionary processes. For example, adaptive traits are often conserved among
405 phylogenetically-related organisms because the historical loss of those traits generated fitness
406 costs that eventually drove the organisms bearing those costs extinct[52]. Similarly, a low ratio of
407 non-synonymous to synonymous substitutions can indicate negative selection acting to remove
408 substitutions in a protein-coding gene that would deleteriously alter the amino acid composition
409 of that protein[53]. Although such patterns can identify a past history of selection acting on a trait,
410 they can only imperfectly infer if selection is currently acting on that trait as required to define
411 an SE function. These methods (and approaches based on the Price equation) also cannot
412 discriminate between direct and indirect modes of selection (as is also required to define an SE
413 function) unless specific frameworks are used that can differentiate between these possibilities[12].
414 Even using such frameworks, it is formally impossible to exclude the possibility that unmeasured
415 covariates might artificially cause the observed measurement of selection instead of direct
416 selection acting on the trait of interest. The methodological difficulties of measuring current and
417 direct selection remain significant obstacles to defining SE functions that are encoded by the
418 microbiome.

419

References

1. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 3229–3236 (2013).

2. Casadevall, A. & Fang, F. C. Descriptive science. *Infect. Immun.* **76,** 3835–3836 (2008).

3. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: a phylogenetic perspective. *Science (N. Y.)* **350,** aac9323 (2015).

4. Amundson, R. & Lauder, G. V. Function without purpose: the uses of causal role function in evolutionary biology. *Biol. Philos.* **9,** 443–469 (1994).

5. Wright, L. Functions. *Philos. Rev.* **82,** 139–168 (1973).

6. Cummins, R. Functional analysis. *J. Philos.* **72,** 741–765 (1975).

7. Millikan, R. G. In defense of proper functions. *Philos. Sci.* **56,** 288–302 (1989).

8. Neander, K. The teleological notion of 'function'. *Australas. J. Philos.* **69,** 454–468 (1991).

9. Doolittle, W. F., Brunet, T. D. P., Linquist, S. & Gregory, T. R. Distinguishing between 'function' and 'effect' in genome biology. *Genome Biol. Evol.* **6,** 1234–1237 (2014).

10. Doolittle, W. F. & Brunet, T. D. P. On causal roles and selected effects: our genome is mostly junk. *BMC Biol.* **5,** 116 (2017).

11. Godfrey-Smith, P. A modern history theory of functions. *Nous* **28,** 344–362 (1994).

12. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution (N. Y).* **37,** 1210–1226 (1983).

13. Okasha, S. *Evolution and the Levels of Selection*. (Oxford University Press, 2006).

14. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. B Biol. Sci.* **205,** 581–598 (1979).

15. Mayr, E. Cause and effect in biology. *Science (N. Y.)* **134,** 1501–1506 (1961).

16. Tinbergen, N. On aims and methods of Ethology. *Z. Tierpsychol.* **20,** 410–433 (1963).

17. Smith, K., McCoy, K. D. & Macpherson, A. J. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin. Immunol.* **19,** 59–69 (2007).

18. Mahowald, M. A. *et al.* Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5859–5864 (2009).

19. Baquero, F. & Nombela, C. The microbiome as a human organ. *Clin. Microbiol. Infect.* **18,** 2–4 (2012).

20. Bordenstein, S. R. & Theis, K. R. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol.* **13,** e1002226 (2015).

21. Theis, K. R. *et al.* Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. *mSystems* **1,** e00028-16 (2016).

22. Mushegian, A. A. & Ebert, D. Rethinking 'mutualism' in diverse host-symbiont communities. *BioEssays* **38,** 100–108 (2016).

23. Kopac, S. M. & Klassen, J. L. Can they make it on their own? Hosts, microbes, and the

holobiont niche. *Front. Microbiol.* **7,** 1647 (2016).

24. Ebert, D. The epidemiology and evolution of symbionts with mixed-mode transmission. *Annu. Rev. Ecol. Evol. Syst.* **44,** 623–643 (2013).

25. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15,** R89 (2014).

26. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).

27. Archie, E. A. & Tung, J. Social behavior and the microbiome. *Curr. Opin. Behav. Sci.* **6,** 28–34 (2015).

28. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480,** 241–244 (2011).

29. Byndloss, M. X. *et al.* Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science (N. Y.)* **357,** 570–575 (2017).

30. Fitzpatrick, B. M. Symbiote transmission and maintenance of extra-genomic associations. *Front. Microbiol.* **5,** 46 (2014).

31. Lynch, M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 8597–8604 (2007).

32. Sharon, G. *et al.* Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20051–6 (2010).

33. Brucker, R. M. & Bordenstein, S. R. The hologenomic basis of speciation: gut bacteria cuase hybrid lethality in the genus *Nasonia*. *Science (N. Y.)* **341,** 667–669 (2013).

34. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLOS Biol.* **14,** e2000225 (2016).

35. Zilber-Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32,** 723–735 (2008).

36. Rosenberg, E. & Zilber-Rosenberg, I. *The Hologenome Concept: Human, Animal and Plant Microbiota*. (Spring International Publishing, 2013).

37. Roughgarden, J., Gilbert, S. F., Rosenberg, E., Zilber-Rosenberg, I. & Lloyd, E. A. Holobionts as units of selection and a model of their population dynamics and evolution. *Biol. Theory* **13,** 44–65 (2018).

38. Moran, N. A. & Sloan, D. B. The hologenome concept: helpful or hollow? *PLoS Biol.* **13,** e1002311 (2015).

39. Douglas, A. E. & Werren, J. H. Holes in the hologenome: why host-microbial symbioses are not holobionts. *mBio* **7,** e02099-15 (2016).

40. Janzen, D. H. When is it coevolution? *Evolution (N. Y).* **34,** 611–612 (1980).

41. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).

42. Casadevall, A. & Fang, F. C. Rigorous science: a how-to guide. *mBio* **7,** e01902-16 (2016).

43. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome.

503 *Nature* **489,** 57–74 (2012).

504 44. Graur, D. An upper limit on the functional fraction of the human genome. *Genome Biol.*
505 *Evol.* **9,** 1880–1885 (2017).

506 45. Eddy, S. R. The C-value paradox, junk DNA and ENCODE. *Curr. Biol.* **22,** R898–R899
507 (2012).

508 46. Doolittle, W. F. Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci. U. S.*
509 *A.* **110,** 5294–5300 (2013).

510 47. Brunet, T. D. P. & Doolittle, W. F. Getting 'function' right. *Proc. Natl. Acad. Sci. U. S. A.*
511 **111,** E3365 (2014).

512 48. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc. Natl.*
513 *Acad. Sci. U. S. A.* **111,** 6131–6138 (2014).

514 49. Price, G. R. Selection and covariance. *Nature* **227,** 520–521 (1970).

515 50. Webster, N. S. & Reusch, T. B. H. Microbial contributions to the persistence of coral
516 reefs. *ISME J.* **11,** 2167–2174 (2017).

517 51. Govaert, L., Pantel, J. H. & De Meester, L. Eco-evolutionary partitioning metrics:
518 Assessing the importance of ecological and evolutionary contributions to population and
519 community change. *Ecol. Lett.* **19,** 839–853 (2016).

520 52. Rocha, E. P. C. Evolutionary patterns in prokaryotic genomes. *Curr. Opin. Microbiol.* **11,**
521 454–460 (2008).

522 53. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.*
523 **18,** 486 (2002).

Correspondence

Jonathan L Klassen

jonathan.klassen@uconn.edu

University of Connecticut, Department of Molecular and Cell Biology, Storrs, CT, USA

Acknowledgements

I thank the members of the Klassen lab and J. P. Gogarten for their helpful feedback on earlier versions of this manuscript. Funding for this work was provided by NSF IOS-1656475 and the University of Connecticut.

Author Contributions

JL Klassen conceptualized and wrote this manuscript.

Competing Interests statement

Yes there is potential Competing Interest.

Funding for this work was provided by NSF IOS-1656475 and the University of Connecticut. These funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Figure Legends and Tables

Figure 1. The relationship between Causal Role (CR) and Selected Effect (SE) functions. Causal Role functions are defined by the change in the phenotype of a larger system when a part of that system is removed, whereas SE functions are defined by natural selection acting on such a part, resulting in the observed phenotype. Only a subset of CR functions are also SE functions, just as circles comprise a small subset of all possible ellipses.