Biology & Philosophy (2023) 38:45
https://doi.org/10.1007/s10539-023-09933-x

An account of conserved functions and how biologists use them to integrate cell and evolutionary biology

Beckett Sterner[1] · **Steve Elliott**[2] · **Jeremy G. Wideman**[1]

Received: 1 October 2022 / Accepted: 18 September 2023 / Published online: 11 October 2023

Abstract

We characterize a type of functional explanation that addresses why a homologous trait originating deep in the evolutionary history of a group remains widespread and largely unchanged across the group's lineages. We argue that biologists regularly provide this type of explanation when they attribute conserved functions to phenotypic and genetic traits. The concept of conserved function applies broadly to many biological domains, and we illustrate its importance using examples of molecular sequence alignments at the intersection of evolution and cell biology. We use these examples to show how the study of conserved functions can integrate knowledge of a trait's causal effects on fitness and its history of natural selection without invoking adaptation. We also show how conserved function provides a novel basis for addressing objections against evolutionary functions raised by Robert Cummins.

Keywords Functional explanation · Function prediction · Conserved sequence · Conserved trait · Purifying selection · Negative selection · Sequence alignment · Species design · Role function · Evolutionary function

Introduction

Researchers across the life sciences use the concept of conserved function, (e.g. Dolinski and Botstein 2007; Weinhold et al. 2008; Malhis et al. 2019), but it has been overlooked in philosophers' longstanding program to understand the many senses of 'function' and their epistemic usefulness across the life sciences (Garson 2016). We aim to make several contributions to this overall program. First we characterize the previously unanalyzed concept of conserved function used by biologists. Next, we illustrate how biologists apply the concept when characterizing

✉ Beckett Sterner
bsterne1@asu.edu

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

[2] Center for Biology and Society, Arizona State University, Tempe, USA

phenomena, predicting protein functions, measuring molecular evolution, and discovering cellular signaling mechanisms. Third, we argue that this account of conserved function has several consequences for understanding the epistemology of biology. Among these consequences, we argue that uses of the conserved function concept illustrate how fields such as cell biology and evolutionary biology can be epistemically integrated even as their research primarily investigates role functions or evolutionary functions, respectively. Put differently, the examples we present give new evidence for integrative pluralism about functions (Cusimano and Sterner (2019). Furthermore, we argue that understanding the epistemology of conserved functions helps address many longstanding challenges for concepts of evolutionary functions raised by Kitcher (1993) and Cummins (2002).

There is broad consensus among philosophers of biology that multiple meanings of 'function' are valuable for science because they operate in different types of explanation (Godfrey-Smith 1994). Nonetheless, maintaining multiple meanings sometimes leads to miscommunications and flawed reasoning, such as when different studies of genome function in humans arrive at incompatible conclusions about the prevalence of "junk DNA" (Gerstein et al. 2007; Graur et al. 2013; Germain et al. 2014; Brzović and Šustar 2020; Linquist et al. 2020) or when the current benefits of a trait are a poor guide to its evolutionary origins (Autumn et al. 2002; Brunet et al. 2021). Most of the recent literature on function has focused on refining existing definitions and clarifying their relationships in order to illuminate and possibly avoid these problems, with some occasional novel views proposed (Saborido 2014; Garson 2016). Relatively less attention has been paid to developing a picture of how pluralism about function may contribute positively to biological research practices (Currie 2015; Cusimano and Sterner 2019).

We argue that characterizing the concept of conserved function illuminates how biologists explain phylogenetic distributions of homologous traits across groups of lineages rather than within single lineages. Importantly, we show how biologists use conserved functions to explain the presence of a homologous trait across a group's lineages in terms of negative (or "purifying") natural selection rather than adaptation, highlighting the importance of a type of selection philosophers have yet to incorporate into the epistemology of functional explanations. By making explicit the roles of lineage groups in conserved function explanations, we clarify how molecular and cell biology produce knowledge about role functions that enables macroevolutionary generalizations about life. We also show how phylogenetic knowledge about conserved functions can drive discoveries about mechanisms and role functions in particular species.

In the next section, we start by reviewing the prominence of conserved functions for biologists and how the concept relates to recent work in philosophy of biology. In Sect. "Characterizing a concept of conserved function", we present an account of the concept and show how it invokes negative selection across a set of lineages, rather than positive selection within single lineages. In Sect. "Conserved function underwrites epistemic integration in the life sciences", we describe four examples of how molecular, cellular, and evolutionary biologists use conserved functions for a variety of epistemic aims. We argue that these cases illustrate and support integrative rather than disjunctive pluralism about functions. In Sect. "New insights on classic issues

in the function debate" we show how our account addresses or dissolves longstanding challenges for evolutionary functions raised by Kitcher and Cummins.

A missing form of evolutionary explanation in theories of function

We pursue three positions in this section. First, we review the standard distinction made between role functions and evolutionary functions. Second, biologists use a concept of conserved function. Third, there remains an opportunity for philosophers to address this usage and its relation to the standard picture.

The standard picture

Biologists use the word "function" to mean different things (Garson 2016). Many biologists primarily care about causal roles (what-does-it-do questions) while others care more about selected effects (why-is-it-there questions) (Godfrey-Smith 1993; Amundson and Lauder 1994; Griffiths 2006; Garson 2016). In basic terms, sometimes biologists use 'function' to refer to what some part of a system contributes to an overall capacity of that system, regardless of how the part and system came to be; and sometimes they use 'function' to refer to the historical causes for why a system has come to have some part or property, specifically due to the beneficial effects it caused in the past. Philosophers generally label the former meaning a "role function" and the latter meaning an "evolutionary function" or "etiological function."

Biologists invoke these concepts of biological functions to explain phenomena. For example, we might ask, "Why does a sample of SARS-Cov-2 viruses from England all have a guanine (G) at nucleotide position 345 in their spike protein genes?" An evolutionary explanation might be: the ancestral nucleotide (i.e. prior to the start of the pandemic) at that position was adenine (A), but a mutation occurred at this nucleotide position that increased the contagiousness of the virus, and this effect contributed to the spread and ultimate near-fixation of the mutation in England. This explanation is evolutionary in form: it refers to historical events and processes that are causally responsible for the relevant observation.

In contrast, role functions are relevant to a different kind of question. For instance, "How do mutations in the spike protein increase the virus's contagiousness?" The explanation might then be that a change in nucleotide sequence led to an amino acid change in the protein expressed by the gene that increased the spike protein's biochemical affinity for binding to cell receptors in human hosts. This tells us how the new amino acid encoded at that position causally contributes to the virus's overall capacity to infect cells. It doesn't address why the mutation is common rather than rare in the population, however.

Biologists construct parallel questions about the presence of whole genes, gene copy numbers, protein complexes, and even entire chromosomes in organisms of a species. Cusimano and Sterner (2019) discuss function in the context of explaining gene duplications, for instance, and Doolittle et al. have considered protein complexes as well (Linquist et al. 2020; Brunet et al. 2021). There

are also other types of functional explanation discussed in the philosophical literature, e.g. propensity or organizational functions, which we won't address here (Garson 2016).

The prominence of research about conserved functions

Conserved functions are regularly at the center of high-profile research in biology. Biologists are (implicitly or explicitly) deeply concerned with how certain causal roles of traits have stayed relatively constant among species, lineages, kingdoms, or sometimes even domains. This is borne out by the fact that they often study structures and processes that are highly conserved. Illustrative evidence comes from the words 'evolutionarily conserved' appearing in the titles of even the most high-profile cell and molecular biology articles of the past couple decades (e.g. Kinchen and Ravichandran 2010; Tuller et al. 2010; Neely et al. 2010; Rousseau and Bertolotti 2016; Tu et al. 2018; Sreelatha et al. 2018). In total, the word 'conserved' appears in the titles of 83 cell or molecular biology articles published in *Science* alone between 2013 and 2023. In both cases, the word 'conserved' implies incontrovertible biological importance of the conserved gene, protein, or trait to the organisms that bear it.

Biologists use conserved functions to talk about traits that show a pattern of maintenance within a clade where a properly *phylogenetic* explanation is required, i.e. one that centrally invokes inheritance from a common ancestor to explain a pattern observed across multiple lineages.

Biologists also commonly use 'conserved trait' or 'conserved sequence' to refer to a homologous character state that has remained constant across a group of lineages and that was acquired through inheritance from a common ancestor. An example at the cellular level would be having mitochondria, which are cell organelles present in nearly all eukaryotic lineages today. Later we'll discuss another example, the two-component signaling system in bacteria, where many species rely on two types of proteins to detect and regulate cell movement along chemical gradients in their environment. Alan Love has also analyzed the concept of conserved genetic mechanisms in developmental biology, which he defines as "shared, derived traits composed of particular constituents, organized in a specific way, and found in delimitable spatiotemporal contexts where they manifest a stereotypical behavior or phenomenon" (Love 2017, 337).

In contrast, convergent and parallel evolution are more familiar types of multilineage patterns for philosophers. In both cases, species independently evolve similar characteristics, for example because they experienced similar selection pressures. Biologists refer to similar traits with independent origins as homoplasies to distinguish them from homologies, which are traits shared among a group of lineages due to common ancestry (Novick 2018). Importantly, the criteria distinguishing a homoplasy from a homology are different than explaining why a homology that originated in a distant common ancestor is still observed to be present in some or all of its descendent lineages today.

An opportunity to characterize a concept of conserved function

To date the philosophical literature has overlooked the relevance of phylogeny to functional explanation, and the term 'conserved function' has not been addressed in the philosophical literature. All philosophical discussions of evolutionary functions so far focus on evolution within single lineages, i.e. on how function attributions can serve to explain the existence of a trait in a particular species over time.

The closest discussion is recent work on maintenance functions (Elliott et al. 2014; Linquist et al. 2020; Linquist 2022; but see also Brzović and Šustar 2020). As part of their discussion of the debate over ENCODE and human genome function, Brzović and Sustar (2020) provide a helpful definition of conserved sequence in a footnote: "Evolutionarily conserved regions are sequences which are similar or identical across different taxa. That is, sequences that persist in the genome despite random mutations and deletions or chromosomal rearrangements. Such sequences are more similar across taxa than would be expected in, for instance, the assumption of neutral evolution" (Brzović and Sustar 2020, 3). When biologists talk of a conserved sequence, they are referring at root to the presence of identical or similar nucleotide sequences at the same (i.e. homologous) positions in the genomes of two or more lineages. In addition, recent work by (Elliott et al. 2014; Linquist et al. 2020; Brunet et al. 2021) has argued for the utility of distinguishing maintenance functions, which rely only on a history of negative selection, from origin functions, which presuppose some adaptive process when the trait first became common or fixed. Their view of maintenance functions can in principle apply to monophyletic or paraphyletic groups of lineages as well, but they do not develop this point explicitly. In addition, conserved functions as we characterize them do not presuppose adaptation.

We will be concerned with the reasons why we observe one pattern of character states out of many other possibilities across a set of lineages. For example, why do all animals have mitochondria? We take a pragmatic approach to explanation (Fraassen 1977), in the sense that characterizing conserved functions provides answers to biologists' why questions regarding conserved sequences or traits. We assume in our discussion that any traits of interest are genuine evolutionary homologies and set aside uncertainties about convergence versus homologous origins. To answer the question, then, it is not sufficient to explain the maintenance or existence of mitochondria in any single lineage. Nor is it sufficient to explain the mitochondrion's origin in some earlier ancestral lineage of animals. The trait might have become fixed in the ancestral lineage by chance or by positive selection, but neither option necessarily determines what happens after the descendent lineages have speciated. Inheritance through common descent is clearly relevant, but it only tells us that the trait can evolve under many possible evolutionary processes. We would expect a very different phylogenetic pattern for a trait whose loss has zero effect on fitness versus a lethal or sterilizing effect, for example. Note that any evolutionary outcome demands an explanation here, even a "random" pattern consistent with Brownian motion.

While it is a necessary condition for a trait to be conserved that it show little to no variation, this pattern alone is consistent with several alternative explanations besides a conserved function. In principle, for instance, there might be little variation in a phenotypic trait due to genetic or environmental redundancy or robustness.

It also could be that variation did exist but chance extinction events eliminated those lineages before they could be observed.

What's key to conserved function, as we'll see in the next section, is that variation arising within each lineage has been eliminated by natural selection due to the negative fitness effects of deviating from the observed conserved trait. Distinguishing between these alternative hypotheses is a challenging empirical and methodological problem for evolutionary biology, but we will focus on characterizing and illustrating the use of conserved function as a concept.

Characterizing a concept of conserved function

In this section, we propose a definition of conserved function and clarify its relationship to some related ideas. To do so, we first need to add further nuance to how philosophers of biology typically understand natural selection. In particular, evolutionary biologists commonly distinguish between positive and negative selection, and a trait can acquire an evolutionary function through either form of selection. Hence evolutionary functions may exist without there having been selection *for* a trait in the standard sense of the trait having increased in frequency in a population because it provided a fitness benefit relative to other traits present at the time.

We first introduce our proposed definition of conserved function and then explain some of the key ideas it uses. We state the definition in a way that is meant to apply across levels of evolutionary individuality, e.g. to organisms as well as genes:

> A trait has a conserved function in a monophyletic or paraphyletic group if and only if:

1. The trait is homologous in those lineages, i.e. has been inherited from a single origin in a common ancestor
2. Since the group's origin, the trait has causally contributed to the survival and reproduction of members of the lineages in the same way for each lineage, i.e. by realizing the same type of role function
3. Heritable variation in the trait occurred in each lineage
4. The trait has stayed constant because natural selection has acted against loss or modification.

This definition of conserved function serves to distinguish it from other meanings of function in three main respects. First, in terms of the scope of explanatory target: biologists use the concept of conserved functions to explain patterns of similar traits among multiple lineages, while they use adaptive functions to explain the fixation of a trait over time within single lineages. Second, in terms of process: the definition requires that negative selection has been the dominant evolutionary process responsible for a trait's continued existence or state instead of positive selection. Third, in terms of time: the definition requires negative selection to have acted on variation in the trait since group's common ancestor, in contrast to other definitions that only

address particular subperiods of time in the trait's history, such as a lineage's recent history of selection (Godfrey-Smith 1994).

We include both monophyletic and paraphyletic groups as valid for conserved functions. Paraphyletic groups allow for the ancestral trait to be lost in one or more members of the clade while still ruling out multiple independent trait origins (which would then pick out a polyphyletic group). Conserved function attributions may therefore entail a range of weak to strong generalizations based on whether they explain the preservation of the trait in a small proportion or all of the clade being considered (Mitchell 2000).

We next discuss the four conditions that comprise the definition. Condition 1 requires that the trait be an evolutionary homology shared among the lineages, but there's no assumption made about whether its origin or initial fixation in the common ancestral lineage involved positive selection. A genetic mutation in a protein may have become fixed in the ancestral population by drift, for example, while a later mutation made it essential for the protein's structural stability.

Condition 2 is important for individuating conserved functions. It requires there to be a shared type of causal effect, i.e. the same type of role function at work, for which the trait has been conserved among the set of lineages. This condition eliminates putative explanations of the trait's phylogenetic distribution that give different causes for the trait's preservation in different parts of the tree. For example, if a gene at some point in a clade's history evolved a new function in a paraphyletic group while it maintained the original function in other parts of the clade, then the gene would have a conserved function in that paraphyletic subset but lack a conserved function in the whole clade. Similarly, if biologists use the concept of role functions to individuate conserved functions, they must decide on a taxonomy of role functions, which poses substantial practical challenges. Notably, whether traits have the same or different causal role functions may depend on the level of abstraction biologists use to describe their effects (Inkpen et al. 2017; Love 2017). We believe the need for a shared taxonomy of role types represents a genuine difficulty for using conserved functions in biology, however, rather than an artifact of our analysis.

Condition 3 is about the trait having actually to be preserved by selection against loss or modification. If no heritable variation actually occurs, then the fact that the trait is uniformly present in the set of lineages doesn't need further explanation.[1] This is an important way in which conserved function differs from Sustar and Brzovic's definition of "weak etiological function," for which it suffices if a trait made a "contribution to the containing organism's and its ancestors' fitness, even if there was no variation of the trait" (Brzović and Šustar 2020, 5).

Condition 4 finally takes us to the heart of the concept of conserved function, which is the action of negative selection to preserve the homologous trait (condition

[1] This phenomenon is likely to be rare, but hypothetically it could arise by several mechanisms. A genetic locus might never experience a DNA mutation by chance, for example, or a trait might be sufficiently causally overdetermined by the environment or other organismal traits that it is robust to all of the mutations that actually occurred in the clade's history. Note that these examples are different from lethal mutations that do occur but lead to the organism's rapid death, so that the mutation is never observed in adults.

1) against modification or loss (condition 3) due to the role it plays across the set of lineages (condition 2). Conserved function explanations do not assume that the allele's fixation or high frequency in the population was ever due to selection in an adaptive sense. The role of natural selection in conserved function explanation is therefore not the classical idea of selection *for* a trait where an allele initially has a low frequency in the population and then increases in frequency due to its positive fitness effect relative to other competing alleles. In order to clarify what is assumed, we need to briefly survey the broader typology biologists use to describe selection.

Evolutionary biologists generally categorize selection into a number of sub-types that differ based on the kinds of data being analyzed and the results of such data. Importantly, selection is measured by either reference to phenotypes or genes. It also matters whether the trait is discrete or continuous valued. For example, the DNA nucleotide at a particular position in the genome (A, G, C, or T) is a discrete genetic trait, and the presence or absence of eyes in cave-dwelling animals is a discrete phenotypic trait. Genetic traits are generally discrete-valued, while many phenotypic traits, such as body size or protein expression levels, are continuous-valued.

Keeping this in mind, there are at least three kinds of selection that can act on genetic loci: positive, negative (also often called purifying selection), and balancing selection. In positive selection, a new beneficial allele is generated by a random mutation within an individual in a population. This beneficial allele results in carriers having more offspring and therefore the allele increases in frequency in the population. According to population genetic theory, a beneficial allele in a large population will increase in frequency to fixation, as long as its benefit is greater than the combined effects of drift, mutation pressure, and recombination (Lynch 2007). In negative selection, when a beneficial allele is fixed in a population, any variation at that locus is quickly purged because organisms without the beneficial allele experience lower rates of survival and reproduction. Importantly, there is no empirical difference from a population genetics perspective between a positively selected allele that is approaching fixation and a fixed allele under purifying selection—only when we consider their deeper history do they become clearly different. Balancing selection occurs when more than one allele is maintained at a locus in a population because both alleles are beneficial in certain circumstances (e.g., a sickle-cell causing allele which also provides malaria resistance in human heterozygotes). The same typology also applies for discrete phenotypic traits.

When it comes to continuous phenotypes, four different kinds of selection can be detected: stabilizing, directional, disruptive, and balancing selection. Stabilizing selection of traits is analogous to purifying selection of alleles (as both indicate stasis), whereas directional selection is analogous to positive selection (as both indicate change). In stabilizing selection, a particular trait is held to a mean within a population whereby deviation is selected against. In directional selection a trait further from the mean of the population is favored. For a continuous trait like height, evidence for stabilizing selection would be a stable range of heights present in a population over time. Evidence for directional selection would be a change in height such that the average height in a population was higher or lower dependent upon the selection pressure. Disruptive (also called diversifying) selection occurs when two extremes are preferred over intermediate phenotypes. In our height example,

perhaps only very tall and very short individuals have selective benefits whereas intermediates are for some reason selected against. Balancing selection of traits can be thought of as an extended kind of diversifying selection whereby several discrete phenotypes are beneficial for possibly different reasons.

Condition 4 therefore explains the presence and maintenance of a trait in a group through the action of negative or stabilizing selection in each lineage depending on whether the trait is discrete or continuous-valued. The importance of distinguishing selection resulting in stability versus change can be illustrated using the controversy over the ENCODE project and its implications for "junk DNA" in the human genome (Linquist et al. 2020; Brunet et al. 2021). ENCODE project researchers sought to measure molecular activities associated with DNA segments across the human genome, e.g. the transcription of DNA into RNA molecules of any variety (not limited to mRNA from protein-coding genes). Based on their results, they claimed about 80% of the genome was functional, which starkly contradicted previous estimates of 10%.

Many biologists contested ENCODE's results, arguing that the lower estimate was based on evidence of DNA sequence conservation due to negative selection, while ENCODE's estimate was based solely on evidence for causal role functions without taking into account evolutionary history. In addition, neither estimate tried to account for portions of the genome that are currently or were previously under positive selection. This distinction led Linquist et al. to distinguish between maintenance and origin functions: "Traits or genetic elements that are merely under purifying selection have what we call maintenance functions whereas those that have historically been under directional selection have origin functions." (Linquist et al. 2020, 1). Quantifying the portion of the human genome under any form of selection would almost certainly lead to a third estimate intermediate between 10 and 80%, illustrating the empirical significance of distinguishing negative versus positive selection as a basis for evolutionary functions.

Note, however, that Linquist et al.'s conception of maintenance function does not require negative selection to preserve the same causal role function across the lineages. We also note that conserved functions are conditioned on the loss or modification of the trait having a negative fitness effect, so that preservation across all members of the clade may break down when one or more lineages experience a major change in environment or mode of life, e.g. when a microbe acquires an endosymbiotic lifestyle or a lineage acquires an evolutionary novelty leading to an adaptive radiation. In this case, one can limit to scope of the conserved function claim to a paraphyletic group that excludes the exceptions.

Conserved function underwrites epistemic integration in the life sciences

Having characterized the concept of conserved function, we illustrate how biologists use it, and we show its value for broader philosophical questions. Here we look at what a concept of conserved function can add to our picture of how different meanings of function are productively related in scientific practice rather than

merely a source of confusion (Sterner 2022). Perhaps the most common view among philosophers is what Cusimano and Sterner (2019) call "disjunctive pluralism," where each legitimate meaning of function serves a different epistemic goal, e.g. a different type of explanation, and scientists use these to answer different research questions. Disjunctive pluralism may come in several forms based on how these epistemic goals are thought to be distributed across biology. Between-discipline pluralism, for example, captures the common view that cell biologists investigate role functions while evolutionary biologists investigate evolutionary functions (Garson 2016). Alternatively, within-discipline pluralism recognizes compelling examples of etiological functions in disciplines outside evolution, such as neuroscience and immunology, and therefore "seeks out and emphasizes the plurality of functions inside any branch of biology and psychology" (Garson 2018, 17).

In contrast to this view, Cusimano and Sterner argue for an "integrative pluralism" about function based on the breakdown of a simple one-to-one relationship between epistemic purposes and meanings of function. For example, they show how biologists' explanations of evolutionary change in protein functions integrates knowledge about evolutionary, propensity, and role functions of a system at different compositional levels. This supports Garson's defense of within-discipline pluralism but goes further to address the importance of pluralism within individual research problems.

We argue that biologists' use of conserved function illustrates and provides further evidence for integrative pluralism: the joint requirement of a history of negative selection and a shared causal role in the definition of conserved function provides a basis for theoretical and methodological integration between evolutionary and cell biology. We illustrate this point using examples from computational sequence alignment, especially the derivation and application of amino acid substitution matrices in protein sequence alignment. Sequence alignment more broadly has become an ubiquitous and essential tool for research in any domain of biology dealing with DNA, RNA, or proteins. After describing the use of conserved functions in justifying key assumptions and methods of protein sequence alignment, we show how they were central to the discovery of the two-component signaling mechanism in bacteria.

The overarching idea is that knowledge of evolutionary function and role function are inferentially tightly coupled through cycles of explanation and prediction based on conserved functions. Research projects investigating the current activities and evolutionary history of genetic sequences can therefore drive a virtuous, iterative cycle of discovery using both computational and experimental methods (O'Malley et al. 2010; O'Malley 2011). The connection is usefully summarized in a recent paper by two biologists (Giudicelli and Roest Crollius 2021):

> "Sequences driving evolutionary conserved functions are expected to be themselves evolutionary conserved. Conversely, since genomes result from hundreds of million years of evolution along which virtually every base has had an opportunity to vary, genomic elements that have resisted variation (i.e. conserved elements) have a high probability to be functional elements."

Note that we understand function prediction here in the broad sense of inferring currently unknown facts, not just future states of affairs. For reasons of space, we set aside the special case of forward-looking functions (Garson 2016), which typically require additional information about the fitness effects of mutations that are less commonly available.

Function prediction using protein sequence alignment

Function prediction is an important part of the historical rise of bioinformatics (Stevens 2013), molecular evolution (Suárez-Díaz 2021), and big data in biology (Leonelli 2008; Strasser 2012). The emerging field of bioinformatics arguably proved its merits in the 1980s and 90 s, even before whole genome sequencing, by delivering new computational methods for predicting molecular function based on the "alignment" of DNA or amino acid sequences. Early online services, for example, became indispensable tools for molecular and cell biologists to search for similar genes or proteins and identify new hypotheses about in vivo molecular activities that could be tested experimentally. These tools relied on both lab experiments characterizing the biochemical properties of molecules and theoretical knowledge about molecular evolution.

The functional annotation of species' genomes, including the human genome, remains far from complete and a central challenge for interdisciplinary research. The public story of the human genome project, for instance, typically ends in 2003 with the announcement of the first draft genome (Stevens 2013), but this draft actually excluded large portions of the genome, had many gaps, and provided limited annotation of protein-coding regions. Most of the puzzle pieces for understanding what genomes do in other species are also missing, even in model organisms such as *Escherichia coli*.

The inferential norms and strategies of sequence alignment, though, have been largely overlooked in the philosophical literature. For historical perspectives, see (Strasser and de Chadarevian 2011; Stevens 2017). As a contribution toward filling this gap, we highlight how protein sequence alignment can be used for comparative reasoning about function and mechanisms across species.

We start by considering an early and influential theoretical approach to protein function prediction that uses evolutionary relatedness as a proxy for similarity of function among protein-coding genes. As biologist Chris Ponting summarizes, "An assumption often made is that the functions of homologues have remained essentially unchanged since the time of their last common ancestor" (Ponting 2001, 19). The validity of this assumption has a direct connection with the ubiquity of conserved functions across protein families.[2] Recall that attributing a conserved function to homologous proteins in a clade explains why the underlying protein-coding genes are present in each lineage and why they encode similar amino acid

[2] For simplicity we will treat gene families as if they only contained orthologs, i.e. gene copies located in different species that share a common ancestor, and ignore gene duplications that have occurred within a species (paralogs).

sequences: negative selection has acted on the genes in each lineage to preserve the protein against loss or modification of its relevant biochemical activities.

Conserved function thus provides a useful theoretical scenario for predicting a newly sequenced protein's molecular function based on knowledge of existing protein sequences and functions. The stability of the conditions, which gives rise to negative selection across the protein family (Condition 4), warrants the inference that the new sequence has also been subject to purifying selection for the same causal role (Condition 2). Without this fact, neutral evolution or adaptation would be plausible and potentially make a change in function more likely. The new sequence's evolutionary relationship to known sequences will be identifiable based on its similarity to other members of the family (Condition 1). And of signal practical importance, this similarity can be operationalized by aligning other members of the family to quantify rates and patterns of mutations (Condition 3).

While using sequence similarity alone to predict function is a broadly effective heuristic, it will fail systematically in the absence of a conserved function. As Ponting points out, "A better view is that an evolutionary relationship implies functional similarity but that this may be true to a greater or lesser extent" (Ponting 2001, 19). For example, the assumption that similar sequences entail similar functions will be satisfied only transiently in cases for a protein family whose common ancestor evolved under positive selection but whose members are now evolving neutrally. Proteins evolving neutrally accumulate mutations randomly with respect to any selective forces that shaped their ancestor, so biochemically crucial amino acids are no more likely to be preserved over time than irrelevant ones. The historical traces of positive selection and common descent in the family will therefore be detectable at first but progressively erased with time.

Evidence for conserved functions

So far we've been considering how conserved functions can warrant function predictions, but we haven't yet touched on how biologists provide evidence to support conserved function claims. The distinctive etiology of conserved functions has a couple further implications of epistemic importance here. In particular, proteins with a conserved function will: (1) tend to show reduced average rates of amino acid change than expected under neutral or adaptive scenarios, and (2) show varying rates of change along their amino acid sequence based on the strength of purifying selection and the specific character of the biochemical properties it preserves at each position.

In the context of protein evolution, the existence of heritable variation at the sequence level can generally be taken for granted (Condition 3), but the other criteria for conserved functions require more specific empirical support. Sequence alignment provides evidence for both homology (Condition 1) and the historical activity of negative selection (Condition 4). Sequences evolving neutrally, for example, are expected to show lower sequence identity on average and shorter "runs" of contiguous, identical sites. Preservation of biochemical activity across species (Condition 2) is best supported by experimental studies, or to a lesser degree biophysical modeling of the protein's structure and chemistry.

Amino Acid Substitution Matrices

We next discuss amino acid substitution matrices, developed by Margaret Dayhoff in the 1970s and then by others after the 1990s (Strasser 2010, 2011; Stevens 2013). Models of DNA mutation have already been addressed elsewhere (Dietrich 1994; Dietrich and Skipper 2007; Suárez-Díaz 2009, 2013, 2021). Here we describe how biologists calculated amino acid substitution matrices from the alignments of protein families with known evolutionary homologies, and how biologists then generalized the method to predict protein homologies and functions in other protein families.

Dayhoff originally calculated amino acid substitution matrices to measure the average rate at which different amino acids replaced each other in protein sequences (Condition 3) with nearly neutral effects on fitnesses, i.e. which were not "rejected by natural selection" (Condition 4) (Dayhoff 1969, 77). Her dataset included sequence alignments of multiple protein families whose members have the same known molecular functions (Condition 2), and her results constituted a path-breaking, quantitative picture of molecular evolution. Her calculations assume the protein sequences being aligned are so similar that their evolutionary homology (at the whole protein and amino acid position levels) was not in doubt (Condition 1), and that any differences of amino acids across sequences resulted from single substitution events, i.e. that parsimony applies. Her basic procedure was then to count how often each amino acid type has changed in the aligned sequences and divide that by the number of times the amino acid type occurred in the sequence overall. This percentage estimates the probability an instance of the amino acid type will change in a small period of time. Figure 1 shows an early matrix Dayhoff derived. Her work led to the widespread use of substitution matrices known as PAM matrices. She also initiated a new subfield of research on improved matrices, including for specialized contexts such as membrane proteins that experience very different chemical environments compared to the cytosol. The characterization of conserved functions was an important component of the derivation and theoretical justification for the PAM matrices and hence for the historical development of bioinformatics more generally.[3]

Discovery of two-component signaling mechanism

The integrative use of evolutionary and role functions in sequence alignment methods can also serve to discover mechanistic generalizations (Craver and Darden 2013; Love 2017). We illustrate this point with the discovery of two-component signaling (TCS) systems in bacteria. TCS systems are composed at minimum of two proteins (Fig. 2), a histidine kinase and a response receiver, and they detect and transmit chemical signals across cell membranes or other internal cell compartments (Stock et al. 2000; Bourret and Silversmith 2010). Most commonly, the histidine kinase is a trans-membrane protein that binds to an

[3] We note, though, that sequence alignment methods do not inherently rely on conserved functions, e.g. in profile-based methods that can detect short functional motifs based on alignments of unrelated proteins with similar biochemical activities (Bairoch and Bucher 1994).

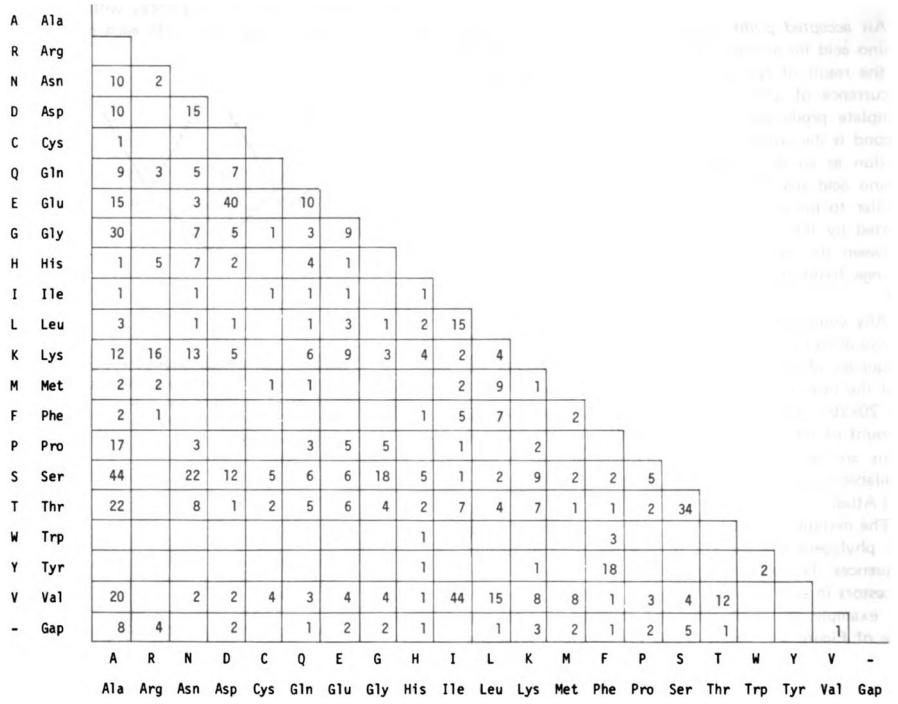

Fig. 1 Counts of accepted point mutations observed in sequence alignments of proteins from ten protein families: cytochrome c, hemoglobin α, hemoglobin β, myoglobin, virus coat protein, chymotrypsinogen glyceraldehyde 3-phosphate, dehydrogenase, clupeine, insulin A and B, and ferredoxin. These data were published as intermediate results for the calculation of amino acid substitution rates in the PAM matrices. Figure from (Dayhoff 1969, 76)

external substrate and modifies the response receiver through allosteric and phosophorylation reactions. The response receivers are located inside the membrane and either directly or indirectly regulate gene expression based on their average degrees of phosphorylation within the cell. Based on sequence conservation, histidine kinases constitute a family of homologous proteins, and response receivers form a second homologous family. TCS systems are almost universally present in bacteria and also in many archaea and eukaryotes, likely due to horizontal gene transfer (Capra and Laub 2012). Bacteria use TCS systems for a wide array of functions such as chemotaxis, quorum sensing, and metabolic regulation.

The discovery of TCS systems was a landmark achievement in biologists' understanding of cell signaling, unifying previously fragmented knowledge from multiple species and leading to decades of mechanistic and evolutionary research (Gupta and Gupta 2021). The key insights were presented in several papers that combined genetic sequencing, lab experiments, and sequence alignment to connect information about the molecular activities and interactions of proteins from several species (Stock et al. 1985; Nixon et al. 1986; Ninfa and Magasanik 1986).

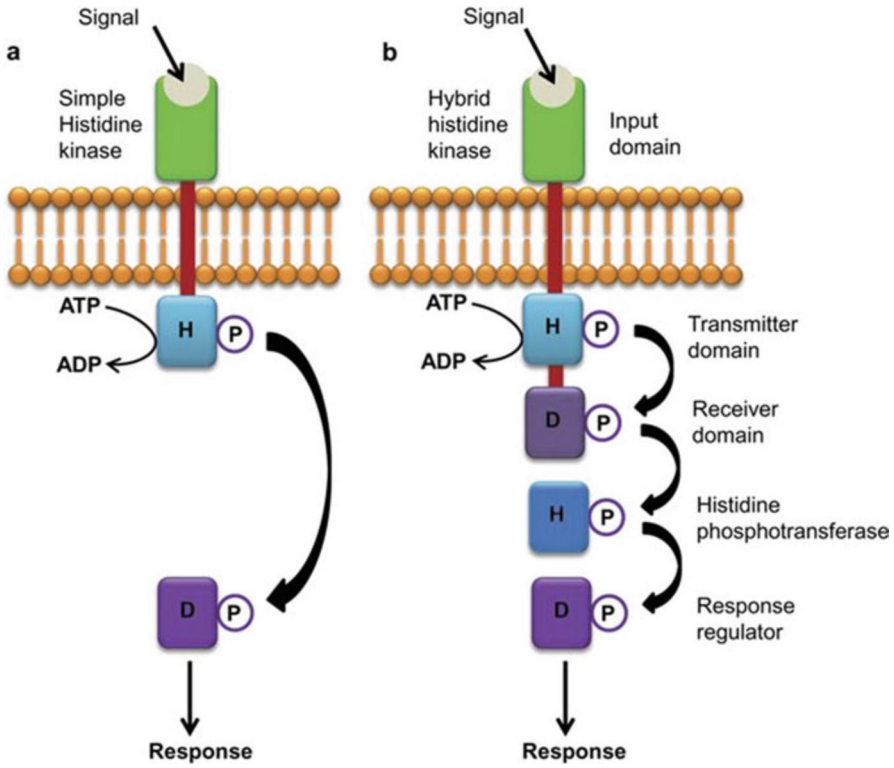

Fig. 2 Illustrations of typical simple **a** and complex **b** two-component signal transduction systems that span cell membranes. All two-component signaling systems contain a histidine kinase—which is composed of the outward facing domain (green box), transmembrane domain (red line), and transmitter domain (blue box labelled H) — and a response regulator (purple box labeled D). The mechanism operates by changes in protein conformation and phosphorylation levels induced by external molecules binding to histidine kinase, leading to downstream phosphorylation changes in the response receiver. Complex systems include a longer internal interaction chain and other modifications to the histidine kinase protein. Figure from (Gupta and Gupta 2021, 560)

Stock et al. (1985) reported the first DNA sequence of CheY, a protein known to be involved in *Salmonella typhimurium* chemotaxis. They applied the FASTP sequence alignment method, which used the PAM 250 matrix, to Dayhoff's Atlas of Protein Sequences. They detected unexpected homologies between CheY and four other proteins in *S. typhimurium, Escherichia coli*, and *Bacillus subtilis* (Conditions 1, 3). Importantly, the homologous proteins were also involved in other cellular functions such as osmoregulation and sporulation, indicating a common evolutionary origin and mode of action. The authors concluded,

> "Despite an apparent diversity in effector functions, all five proteins are similar in that they modulate cellular behavior in response to changing environmental conditions. The sequence homologies argue strongly for an evolutionary relationship and raise the possibility that a common mechanism of information processing may be operating in all these systems" (Stock et al. 1985, 7993).

Nixon et al. (1986) expanded on these results using Dayhoff's updated sequence atlas and a similar repertoire of methods to detect homology, including genetic sequencing and sequence alignment. Their research focused on two protein-coding genes in *Bradyrhizobium* sp. [*Parasponia*] strain RP501 that were known to be co-located in the same operon. After sequencing the protein-coding genes, they used sequence alignments to discover new homologs in other species and reported that "these observations prompted us to examine a number of other putative regulatory genes that are part of two-gene systems and that are involved in responses to environmental stimuli" (Nixon et al. 1986, 7852). After comparing the biochemical activities and properties of these proteins (Condition 2), they arrived at a general mechanistic model for signal transduction: "We propose that these regulatory genes comprise two-component regulatory systems that evolved from a common ancestral system that involved transduction of information about the status of the environment by one protein domain (the C-terminal regions conserved among ntrB, envZ, etc.) to a second one (the N-terminal region conserved among ntrC, ompR, etc.)" (Nixon et al. 1986, 7850).

Further research eventually detected TCS systems in almost all bacterial lineages, with exceptions in obligate endosymbionts such as *Mycoplasma* and *Amoebophilus*. Capra and Laub (2012) reported substantial variation in the number of TCS systems possessed by bacterial species, and that species with more copies of the TCS system tend to inhabit rapidly changing or diverse environments, likely because they benefit from having additional, specialized environmental signaling pathways. Biologists have shown that TCS systems exhibit a mix of vertical and horizontal evolutionary relationships, with some bacterial clades maintaining a stable repertoire of specialized histidine kinase and response receiver gene copies from a common ancestor while other lineages appear to have acquired copies horizontally. Due to lateral gene transfer and derived role functions, the widespread phylogenetic distribution of TCS systems therefore does not seem to be explained by a single conserved function at the level of all bacterial cell lineages. However, all TCS systems, even those in archaea and eukaryotes, show clear sequence conservation at the level of the histidine kinase and response receiver protein lineages, which provides evidence for Condition 4 (Capra and Laub 2012). It may be, then, that these two genes have conserved functions for the TCS system when we understand the TCS system as a clade of *multi-gene* lineages in its own right.

TCS systems therefore highlight how biologists integrate knowledge of evolutionary and role functions in discovering mechanisms and explaining how signal transduction works in cases across all three domains of life. Disjunctive pluralism about functions is inadequate for understanding the examples we presented: it is false that biologists investigate and use different types of function in isolation. We showed how the rise of bioinformatics provided crucial epistemic methods (sequence alignment) and resources (sequence databases) that served to bridge between biological disciplines and facilitate knowledge exchange. Integrative pluralism about function therefore provides a better basis for understanding how and why biologists use multiple senses of function in their research practices.

New insights on classic issues in the function debate

The previous section showed how the study of conserved functions drives beneficial epistemic integration among fields that may otherwise seem to be working in parallel. In this section, we build on this result to address some unresolved and long-standing issues in the functions literature. First, we show how the concept of conserved function provides an alternative to the concept of adaptive design as a basis for guiding the appropriate use of role functions in biological research. This result harkens back to Kitcher's (1993) attempt to use the concept of design to unify different meanings of function, and it indicates a different theoretical strategy that remains sympathetic to his motivating concerns. Second, we show how the concept of conserved functions avoid some of the pointed objections that Cummins (2002) raised against evolutionary functions based on how they require positive selection.

Justifying norms for using role functions without presupposing adaptive design

One of the first and most persistent complaints against the causal role account of function has been its over-permissiveness. An appealing response has been to look for a way to recognize the value of role functions but restrict their proper scope of application (Linquist 2022). A number of philosophers have explored whether the idea of a "species design," grounded in a history of accumulated adaptations, could fill this gap (Boorse 1977; Kitcher 1993; Wouters 2007).

For brevity, we focus on Kitcher's account: "The function of X is what X is designed to do, and what X is designed to do is that for which X was selected (Kitcher 1993, 383). Kitcher further distinguishes two ways in which a trait may have a function, direct and indirect. The direct case arises "when the entity is present because of selection for a particular property (that is, its presence is completely explained in terms of selection for that property)" (Kitcher 1993, 389). The indirect case occurs "when organisms experience selection pressure that demands some complex response of them and one of their parts, traits, or behaviors makes a needed causal contribution to that response" (Kitcher 1993, 383).

Peter Godfrey-Smith (1993) refuted Kitcher's account before it could even be published with an example of developmental constraint in the fruit fly species *Drosophila melanogaster*. The example shows how a constrained developmental tendency may sometimes improve organismal fitness and sometimes reduce it depending on the environmental context, contradicting a simple adaptive interpretation of species design. Nonetheless, the basic question remains: how is evolution relevant to biologists' choice of role functions to study?

As Kitcher generally had positive selection in mind (i.e. selection for a trait), his direct and indirect cases fail to capture the full scope of conserved functions. In particular, they overlook cases in which conserved functions arise purely through the entrenchment of chance mutations by negative selection. Key implications of biological design—especially the within-species constancy of and inter-dependency of traits—therefore have no simple relationship to historical adaptation or fitness optimality. For example, the existence of a single conserved function in a clade does

not entail a characteristic species design in the sense of an interlocking system of characters that typically vary only within well-delimited ranges. As we've seen, a hypothesis of conserved function has implications for the causal workings of extant and past organisms. Nonetheless, any single conserved function is restricted to the scope of a particular homologous trait, and one need not invoke the broader concept of a general bodily design. Similarly, conserved functions imply neither global optimality for the relevant traits, nor do they presuppose the general premise that organisms in fact thrive in their evolved habitats (c.f. Kitcher, 381). Instead, conserved functions only require that modifications or losses actually arising in the past have been eliminated due to their deleterious fitness effects.

We suggest that conserved functions highlight key pieces of evolutionary theory that philosophers have overlooked when trying to characterize why biologists focus on particular kinds of role functions and not others: phylogenetic generalizations about the distribution of traits and their role functions. Philosophers defending the practical value of role functions for biology, including Kitcher, have typically drawn their insights from the practices of ecology, physiology, or anatomy (e.g. Kitcher 1993; Amundson and Lauder 1994; Wouters 2007). Yet these fields, as well as cell and molecular biology, are not merely interested in the causal workings of individual species. They more often prize generalizable understanding of what traits do in all the species that have them (Love 2017).

This shared interest in phylogenetic generalization motivates the study of role functions in evolutionary contexts. Sequence alignment methods, as we saw, adopt evolutionary premises about function conservation across taxa and provide tools for discovering the role functions of newly sequenced proteins. Moreover, predictions about protein role functions help test evolutionary models. New research on role functions, for example, can be used to: test generalizations about the negative fitness effects of losing or modifying a trait in natural conditions; determine the degrees to which observed deviations from conservation in particular species can be explained through changes in ecology, developmental redundancy, or other factors; and indicate historical hypotheses about the role of positive versus negative selection in ancestral lineages (Autumn et al. 2002).

Explaining the existence and phylogenetic distribution of complex traits

We turn to a critique of evolutionary functions by Cummins (2002) that to our knowledge has not received an adequate reply in the literature. Cummins' argues that the existence of complex traits such as eyes, hearts, or wings cannot be explained by a history of selection for the high-level functions we typically attribute to them, e.g. vision, pumping blood, or flight. The reason is that natural selection never acts directly on variation in the existence of complex traits as whole units (i.e. the presence or absence of a whole heart) to drive adaptation. Instead, complex traits have multi-stage evolutionary histories in which their component parts or properties underwent positive selection for other or more specific effects, such as light sensitivity or improved blood circulation. At no point, then, did the trait as a whole unit experience selection for the high-level functions we typically attribute to them.

Cummins' point is not to deny the importance of adaptation per se, but to deny the appropriateness of grounding the concept of function in natural selection. More specifically, his target is what he calls neo-teleology: "the substantive thesis that, in some important sorts of cases at least, a thing's function—the effect we identify as its function—is a clue to its existence" (Cummins 2002, 161). He further distinguishes strong and weak versions. "The strong version holds that any biological trait that has a function was selected for because it performed that function. The weak version holds only that some traits were selected because of their functions" (Cummins 2002, 164). His core argument against strong neo-teleology is then that "most, perhaps all, complex structures such as hearts, eyes, and wings patently have functions but were not selected because of (the effects that count as) their functions" (Cummins 2002, 165). For both versions, Cummins is right both that complex functions do not generally burst onto the scene with a single mutation, and that therefore they cannot be explained by simple etiological stories of adaptation. Moreover, once complex traits such as hearts or wings exist, their maintenance against loss is better explained by purifying rather than positive selection.

While valid enough for positive selection, Cummins critique of neo-teleology fails to distinguish the different types of selection we discussed above. In particular, attributing a conserved function does not depend on a scenario for which a novel variant introduces a categorically novel type of effect, e.g. pumping blood or flight, such that this effect was previously absent and then becomes fixed in the population through positive selection.

Once we consider conserved (and maintenance) functions, Cummins's critique of all concepts of evolutionary functions turns out to be over-broad. Purifying selection is a sufficient basis for attributing an evolutionary function to a trait when losses or modifications of the trait are eliminated because they show reduced fitness due to poorer performance at the associated role function. This etiology is consistent with a complex trait emerging gradually through multiple evolutionary processes and then acquiring a conserved function different from any prior evolutionary functions it may have held, if any. When biologists characterize a function as conserved, they do so relative to a trait and a phylogeny. They indicate, sometimes only implicitly, which role function is relevant to explanations about the trait's persistence within the phylogeny. As a result, one cannot eliminate or reduce the existence of the evolutionary function to an epiphenomenal correlation of the trait's phylogenetic history and role functions.

Conclusion

We have presented the first philosophical account of conserved function, identified its distinctive epistemic merit for explaining the phylogenetic persistence of a homologous trait within a group of lineages, and highlighted its significance for both practical and theoretical debates in biology. We argued that philosophers have overlooked concepts, principles, and explanations of signal importance to biology because they have conceived of evolutionary functions as applying primarily to traits in single lineages. Moreover, we showed how research investigating different aspects

of conserved functions drives practical and theoretical integration between evolutionary and molecular and cell biology through a common interest in discovering phylogenetic generalizations about traits. We then argued this integration resolves some lingering conflicts between evolutionary and causal role views of function, supporting an integrative pluralist account of function in biology.

Future work could profitably expand on the importance of phylogeny for comparative research on biological function, for example in interdisciplinary fields such as bioinformatics, evolutionary development (Love 2017; Novick 2019), and evolutionary cell biology (Lynch et al. 2014). We have noted already how conserved mechanisms, sequences, and other traits are central to evolutionary developmental explanations, and our analysis contributes to showing how biological categories such as body plans may be consistent with evolutionary theory (Novick 2019). The iterative use of molecular sequence alignments and functional inference we described can also illuminate how the emerging field of evolutionary cell biology combines historical and experimental methods to understand key events in the history of the eukaryotic cell (O'Malley et al. 2019).

Acknowledgements We would like to thank participants in the Science of Purpose Initiative reading group and Ford Doolittle's past and present lab members for their insightful feedback and criticism of previous drafts. Our thanks also to the referees for their constructive criticism that substantially improved the paper.

Funding BS was supported by John Templeton Foundation Grant 62220. JGW was supported by National Science Foundation Grant DBI-2119963.

Declarations

Conflict of interest None declared by authors BS, SE, and JGW.

References

Amundson R, Lauder GV (1994) Function without purpose. Biol Philos 9(4):443–469

Autumn K, Ryan MJ, Wake DB (2002) Integrating historical and mechanistic biology enhances the study of adaptation. Q Rev Biol 77(4):383–408. https://doi.org/10.1086/344413

Bairoch A, Bucher P (1994) PROSITE: recent developments. Nucleic Acids Res 22(17):3583–3589

Boorse C (1977) Health as a theoretical concept. Philos Sci 44:542–573

Bourret RB, Silversmith RE (2010) Two-component signal transduction. Curr Opin Microbiol 13(2):113–115. https://doi.org/10.1016/j.mib.2010.02.003

Brunet TDP, Ford Doolittle W, Bielawski JP (2021) The role of purifying selection in the origin and maintenance of complex function. Stud Hist Philos Sci Part A 87(June):125–135. https://doi.org/10.1016/j.shpsa.2021.03.005

Brzović Z, Šustar P (2020) Postgenomics Function Monism. Stud Hist Philos Biol Biomed Sci 80(April):101243. https://doi.org/10.1016/j.shpsc.2019.101243

Capra EJ, Laub MT (2012) The evolution of two-component signal transduction systems. Annu Rev Microbiol 66:325–347. https://doi.org/10.1146/annurev-micro-092611-150039

Craver CF, Lindley D (2013) In search of mechanisms: discoveries across the life sciences. University of Chicago Press, Chicago

Cummins R (2002) Neo-Teleology. In: Ariew A, Cummins R, Perlman M (eds) Functions: New Essays in the Philosophy of Psychology and Biology. Oxford University Press, Oxford

Currie A (2015) Marsupial lions and methodological omnivory: function, success and reconstruction in paleobiology. Biol Philos 30(2):187–209. https://doi.org/10.1007/s10539-014-9470-y

Cusimano S, Sterner B (2019) Integrative pluralism for biological function. Biol Philos 34(6):55. https://doi.org/10.1007/s10539-019-9717-8
Dayhoff MO (ed) (1969) Atlas of protein sequence and structure. The National Biomedical Research Foundation, Silver Spring, Maryland
Dietrich MR (1994) The origins of the neutral theory of molecular evolution. J Hist Biol 27(1):21–59. https://doi.org/10.1007/BF01058626
Dietrich MR, Skipper RA (2007) Manipulating underdetermination in scientific controversy: the case of the molecular clock. Perspect Sci 15(3):295–326. https://doi.org/10.1162/posc.2007.15.3.295
Dolinski K, Botstein D (2007) Orthology and functional conservation in eukaryotes. Annu Rev Genet 41(1):465–507. https://doi.org/10.1146/annurev.genet.40.110405.090439
Elliott TA, Linquist S, Ryan Gregory T (2014) Conceptual and empirical challenges of ascribing functions to transposable elements. Am Nat 184(1):14–24. https://doi.org/10.1086/676588
Fraassen Van BC (1977) The pragmatics of explanation. Am Philos Q 14(2):143–150
Garson J (2016) A critical overview of biological functions. SpringerBriefs in Philosophy. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-32020-5
Garson, J (2018) How to be a function pluralist. Br J Philos Sci 69(4):1101–1122. https://doi.org/10.1093/bjps/axx007
Germain P-L, Ratti E, Boem F (2014) Junk or functional DNA? ENCODE and the function controversy. Biol Philos 29(6):807–831. https://doi.org/10.1007/s10539-014-9441-3
Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Jiang Du, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) What is a gene, Post-ENCODE? History and updated definition. Genome Res 17(6):669–681. https://doi.org/10.1101/gr.6339607
Giudicelli F, Crollius HR (2021) On the importance of evolutionary constraint for regulatory sequence identification. Brief Funct Genomics. https://doi.org/10.1093/bfgp/elab015
Godfrey-Smith P (1994) A modern history theory of functions. Noûs 28(3):344. https://doi.org/10.2307/2216063
Godfrey-Smith P (1993) Functions: consensus without unity. na. http://www.matthewnoahsmith.net/s/godfrey-smithfunctionsconsensuswithoutunity-kka9.pdf
Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E (2013) On the immortality of television sets: 'function' in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol 5(3):578–590. https://doi.org/10.1093/gbe/evt028
Griffiths PE (2006) Function, homology, and character individuation. Philos Sci 73(1):1–25
Gupta R, Gupta N (2021) Fundamentals of bacterial physiology and metabolism. Springer Singapore, Singapore. https://doi.org/10.1007/978-981-16-0723-3
Inkpen S, Andrew GM, Douglas TDP, Brunet KL, Ford Doolittle W, Langille MGI (2017) The coupling of taxonomy and function in microbiomes. Biol Philos 32(6):1225–1243. https://doi.org/10.1007/s10539-017-9602-2
Kinchen JM, Ravichandran KS (2010) Identification of two evolutionarily conserved genes regulating processing of engulfed apoptotic cells. Nature 464(7289):778–782
Kitcher P (1993) Function and design. Midwest Stud Philos 18(1):379–397
Leonelli S (2008) Bio-ontologies as tools for integration in biology. Biol Theory 3(1):7–11. https://doi.org/10.1162/biot.2008.3.1.7
Linquist S (2022) Causal-role myopia and the functional investigation of junk DNA. Biol Philos 37(4):28. https://doi.org/10.1007/s10539-022-09853-2
Linquist S, Ford Doolittle W, Palazzo AF (2020) Getting clear about the F-Word in genomics. PLoS Genet 16(4):e1008702. https://doi.org/10.1371/journal.pgen.1008702
Love AC (2017) Developmental mechanisms. In: Stuart G, Phyllis I (eds) The routledge handbook of mechanisms and mechanical philosophy, 1st edn. Routledge, New York, pp 332–347
Lynch M (2007) The origins of genome architecture. Sinauer, Sunderland
Lynch M, Field MC, Goodson HV, Malik HS, Pereira-Leal JB, Roos DS, Turkewitz AP, Sazer S (2014) Evolutionary cell biology: two origins, one objective. Proc Natl Acad Sci 111(48):16990–16994. https://doi.org/10.1073/pnas.1415861111
Malhis N, Jones SJM, Gsponer J (2019) Improved measures for evolutionary conservation that exploit taxonomy distances. Nat Commun 10(1):1556. https://doi.org/10.1038/s41467-019-09583-2
Mitchell SD (2000) Dimensions of scientific law. Philos Sci 67(2):242–265
Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, Griffin RS, Belfer I, Dai F, Smith SB (2010) A genome-wide drosophila screen for heat nociception identifies A2δ3 as an evolutionarily conserved pain gene. Cell 143(4):628–638

Ninfa AJ, Magasanik B (1986) Covalent modification of the GlnG product, NRI, by the GlnL product, NRII, regulates the transcription of the GlnALG operon in escherichia coli. Proc Natl Acad Sci 83(16):5909–5913. https://doi.org/10.1073/pnas.83.16.5909

Nixon BT, Ronson CW, Ausubel FM (1986) Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes NtrB and NtrC. Proc Natl Acad Sci 83(20):7850–7854. https://doi.org/10.1073/pnas.83.20.7850

Novick A (2018) The fine structure of 'homology.' Biol Philos 33(1):6. https://doi.org/10.1007/s10539-018-9617-3

Novick A (2019) Cuvierian functionalism. Philos Theory Pract Biol. https://doi.org/10.3998/ptpbio.16039257.0011.005

O'Malley MA (2011) Exploration, Iterativity and Kludging in Synthetic Biology. C R Chim 14(4):406–412. https://doi.org/10.1016/j.crci.2010.06.021

O'Malley MA, Elliott KC, Burian RM (2010) From genetic to genomic regulation: iterativity in MicroRNA research. Stud Hist Philos Biol Biomed Sci 41(4):407–417. https://doi.org/10.1016/j.shpsc.2010.10.011

O'Malley MA, Leger MM, Wideman JG, Ruiz-Trillo I (2019) Concepts of the last eukaryotic common ancestor. Nat Ecol Evol 3(3):338–344

Ponting CP (2001) Issues in predicting protein function from sequence. Brief Bioinform 2(1):19–29. https://doi.org/10.1093/bib/2.1.19

Rousseau A, Bertolotti A (2016) An evolutionarily conserved pathway controls proteasome homeostasis. Nature 536(7615):184–189

Saborido C (2014) New directions in the philosophy of biology: a new taxonomy of functions. New directions in the philosophy of science. Springer, Berlin, pp 235–251. https://doi.org/10.1007/978-3-319-04382-1_16

Sreelatha A, Yee SS, Lopez VA, Park BC, Kinch LN, Pilch S, Servage KA, Zhang J, Jiou J, Karasiewicz-Urbańska M (2018) Protein AMPylation by an evolutionarily conserved pseudokinase. Cell 175(3):809–821

Sterner B (2022) Explaining ambiguity in scientific language. Synthese 200(5):354. https://doi.org/10.1007/s11229-022-03792-x

Stevens H (2013) Life out of sequence: a data-driven history of bioinformatics. University of Chicago Press, Chicago

Stevens H (2017) A feeling for the algorithm: working knowledge and big data in biology. Osiris 32(1):151–174

Stock A, Koshland DE, Stock J (1985) Homologies between the salmonella typhimurium CheY protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis, and sporulation. Proc Natl Acad Sci 82(23):7989–7993. https://doi.org/10.1073/pnas.82.23.7989

Stock A, Robinson V, Goudreau P (2000) Two-component signal transduction. Annu Rev Biochem 69:183–215. https://doi.org/10.1146/annurev.biochem.69.1.183

Strasser BJ (2010) Collecting, comparing, and computing sequences: the making of Margaret O. Dayhoff's atlas of protein sequence and structure, 1954–1965. J Hist Biol 43(4):623–660. https://doi.org/10.1007/s10739-009-9221-0

Strasser BJ (2011) The experimenter's museum: genbank, natural history, and the moral economies of biomedicine. Isis 102(1):60–96. https://doi.org/10.1086/658657

Strasser BJ (2012) Data-driven sciences: from wonder cabinets to electronic databases. Stud Hist Philos Biol Biomed Sci 43(1):85–87. https://doi.org/10.1016/j.shpsc.2011.10.009

Strasser BJ, de Chadarevian S (2011) The comparative and the exemplary: revisiting the early history of molecular biology. Hist Sci 49(3):317–336. https://doi.org/10.1177/007327531104900305

Suárez-Díaz E (2009) Molecular evolution: concepts and the origin of disciplines. Stud Hist Philos Biol Biomed Sci 40(1):43–53. https://doi.org/10.1016/j.shpsc.2008.12.006

Suárez-Díaz E (2013) The long and winding road of molecular data in phylogenetic analysis. J Hist Biol 47(3):443–478. https://doi.org/10.1007/s10739-013-9373-9

Suárez-Díaz E (2021) The historiography of molecular evolution. In: Dietrich MR, Borrello ME, Harman O (eds) Handbook of the historiography of biology, historiographies of science. Springer, Cham, pp 59–80

Tu Y-H, Cooper AJ, Teng B, Chang RB, Artiga DJ, Turner HN, Mulhall EM, Ye W, Smith AD, Liman ER (2018) An evolutionarily conserved gene family encodes proton-selective ion channels. Science 359(6379):1047–1050

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141(2):344–354

Weinhold N, Sander O, Domingues FS, Lengauer T, Sommer I (2008) Local function conservation in sequence and structure space. PLoS Comput Biol 4(7):e1000105. https://doi.org/10.1371/journal.pcbi.1000105

Wouters AG (2007) Design explanation: determining the constraints on what can be alive. Erkenntnis 67(1):65–80. https://doi.org/10.1007/s10670-007-9045-2