ANNUAL REVIEWS

Annual Review of Microbiology

Transporter Proteins as Ecological Assets and Features of Microbial Eukaryotic Pangenomes

David S. Milner,[1] Luis Javier Galindo,[1] Nicholas A.T. Irwin,[1,2] and Thomas A. Richards[1]

[1]Department of Biology, University of Oxford, Oxford, United Kingdom; email: thomas.richards@biology.ox.ac.uk

[2]Merton College, University of Oxford, Oxford, United Kingdom

Annu. Rev. Microbiol. 2023. 77:45–66

First published as a Review in Advance on March 21, 2023

The *Annual Review of Microbiology* is online at micro.annualreviews.org

https://doi.org/10.1146/annurev-micro-032421-115538

Keywords

accessory genome, core genome, adaptation, virulence, horizontal gene transfer, gene duplication

Abstract

Here we review two connected themes in evolutionary microbiology: (*a*) the nature of gene repertoire variation within species groups (pangenomes) and (*b*) the concept of metabolite transporters as accessory proteins capable of providing niche-defining "bolt-on" phenotypes. We discuss the need for improved sampling and understanding of pangenome variation in eukaryotic microbes. We then review the factors that shape the repertoire of accessory genes within pangenomes. As part of this discussion, we outline how gene duplication is a key factor in both eukaryotic pangenome variation and transporter gene family evolution. We go on to outline how, through functional characterization of transporter-encoding genes, in combination with analyses of how transporter genes are gained and lost from accessory genomes, we can reveal much about the niche range, the ecology, and the evolution of virulence of microbes. We advocate for the coordinated systematic study of eukaryotic pangenomes through genome sequencing and the functional analysis of genes found within the accessory gene repertoire.

Contents

INTRODUCTION: PANGENOMES

Pangenome: the combined set of all genes present across all strains of a given species

Core genes: genes found within the genome of all strains of a given species

Orthologous: describes genes in different species that have evolved from a common ancestral gene and are not distinguished by gene duplication

Accessory gene: gene found within some, but not all, strains of a given species

How microbial organisms adapt to, and colonize, new and varying ecological niches has implications for a number of fields, including the study of host-pathogen interactions and other symbiotic systems. Even within the boundaries of a given species, extensive differences in coding content can be observed (**Table 1**), allowing identification of important variations in gene repertoires between strains (3, 45, 56, 70, 79, 82, 98, 104). These genome comparisons have led to the concept of the pangenome, where the gene content within a given species complex comprises the core genes (orthologous gene families present within all sampled strains of that species) and the accessory—or variable—genes (orthologous gene families, and their paralogs, inconsistently distributed across the genomes of different strains within a species complex) (7). Constructive debate is ongoing regarding whether pangenomes are adaptive, neutral, or (probably) both (1, 58, 94). Several models have suggested that gene acquisitions can be adaptive (58), leading to strains differentially acquiring relative advantages in variant microniches. As such, comparisons of core and accessory genes can provide important clues about how different strains adapt to their environment, interact with others (10, 64), and, in some cases, generate virulence (e.g., 3, 56, 67, 81).

Recent studies have highlighted the enrichment of genes within the core and accessory gene repertoires, across a number of taxa (e.g., 3, 45, 56, 79, 82, 98, 104). These datasets have been used to identify a relationship between pangenome gene distribution (accessory versus core) and categories of protein function (e.g., 67, 79, 107). In this review article, we explore the hypothesis that subsets of membrane-localized metabolite transporter-encoding genes of eukaryotic microbes show a mosaic pattern of gain/loss and therefore enrichment in the accessory gene set. This pattern suggests that genes that encode transporter proteins represent examples of accessory systems with adaptive implications in certain species and specific environments. We hypothesize that the study of these proteins, specifically the presence of these genes within the accessory genome, will inform us about key factors shaping how eukaryotic microbial lineages are evolving, and the environmental and metabolic conditions that are driving evolutionary variation. We discuss how this

Table 1 Eukaryotic pangenome datasets explored in this review article

Species	Number of strains	Lifestyle	Core/accessory gene clusters[a]	Reference
Aspergillus fumigatus	12	Osmotroph	8,073/3,002 (1,294–1,964)	56
Candida albicans	34	Osmotroph/ parasite	5,432/1,893 (487–622)	56
Cryptococcus neoformans var. grubii	25	Osmotroph/ parasite	5,486/2,698 (964–1,590)	56
Emiliania huxleyi	14	Mixotroph	20,055/8,949	79
Saccharomyces cerevisiae	100	Osmotroph	4,900/2,850 (518–967)[b]	56
	1,360	Osmotroph	5,293/1,784 (mean = 430)[b,c]	45
Zymoseptoria tritici	19	Osmotroph/ parasite	9,193/6,282[d]	3
Seminavis robusta	49	Mixotroph	2,8120/9,683	70

[a]Core/accessory genome sizes reported by each study, with distribution of accessory genome size across strains included in parentheses where known. For the 1,360-strain *S. cerevisiae* study, the mean accessory genome size is noted in parentheses.
[b]Note the differences in total pangenome size between the two *S. cerevisiae* analyses (discussed in main text).
[c]Li et al. (45) divide the *S. cerevisiae* accessory genome into the accessory pool (gene groups that are present in ≤5% of strains; 1,121 gene groups in this category) and character genes (gene groups that are present in 5–95% of strains; 663 gene groups in this category). These two categories have been combined into the accessory gene cluster value stated here. Li et al. also use an extended core (genes that are present in >95% of strains) rather than a core genome.
[d]Badet et al. (3) divide the *Z. tritici* accessory genome into accessory orthogroups (4,690 groups) and genes found in a single species only (1,592 singletons).

is important for our understanding of strains that have gained virulence traits and outline support for the idea that gain of accessory genes that encode transporter proteins may underpin pathogen evolution. As part of this latter point, we summarize data showing that many viruses of eukaryotes have acquired transporter-encoding genes to augment host interactions, with these viral genes then also acting as an additional source of accessory genome variation for infected hosts.

OUR UNDERSTANDING OF PANGENOME STRUCTURES IS POORLY DEVELOPED FOR EUKARYOTIC MICROBES

The pangenome concept is firmly established for prokaryotic microbes, with numerous studies demonstrating genome variation across related bacterial strains (reviewed in 58). The concept of species boundaries within prokaryotes is blurred (17), yet it is clear that genome repertoire varies between strains of the same species and that this variation is meaningful for the ecology and evolution of prokaryotic species.

In contrast, eukaryotic pangenome sampling is relatively limited both in the range of taxa for which pangenome sequencing initiatives have been undertaken and in the number of strains sampled for genome sequencing within a species complex. **Table 1** details seven microbial eukaryotic species that have been subject to pangenome sequencing initiatives and for which there are publicly available datasets where the gene repertoires have been divided into core and accessory genomes. We note there are multiple additional population genome studies of eukaryotic microbes (e.g., 5, 78, 93, 103, 106), as well as pangenome datasets that span species boundaries (e.g., 25, 107), that have not been analyzed in further detail in this review article. The list (**Table 1**) includes mixotrophic algae and fungi [organisms that obtain nutrients by osmotrophy (85)], some of which are also parasites. Typically fewer than 50 genomes are sequenced in these studies, with the notable exception of *Saccharomyces cerevisiae*, where 1,360 genomes were studied. In many cases, both for the datasets included in **Table 1** and in population genome studies, the pangenomes analyses are derived from one reference genome combined with strain-level resequencing, with the variant

Mixotrophic: describes organisms that can use autotrophic and heterotrophic modes to obtain energy and/or additional metabolites

Osmotrophy: feeding method by which organisms take up nutrients by absorption through the membrane; facilitated by transporter proteins

strains assembled onto the reference genome. It is not clear whether such methods systematically underestimate pangenome diversity, thereby generating artifacts in gene presence/absence analyses. For example, such approaches may fail to recover differentially distributed chromosomes not present in the reference genome assembly.

As many of these genome datasets are minimal in terms of strain sampling, the true core genome for each species complex is likely to contain fewer genes than current sequencing efforts indicate. This is because the sample sizes are limited and will, therefore, produce data patterns that overestimate the core gene families (45). Furthermore, patterns of differential gene loss can render candidate orthologous cluster groups where the inferred cluster is composed of paralogs with mosaic ancestry that has been produced by differential gene loss (hidden paralogy). This is a problem that is often exacerbated by bioinformatic clustering approaches (a challenge tackled in Reference 16; see also differences between *Saccharomyces cerevisiae* pangenome estimates in **Table 1** resulting from different clustering approaches). Of course, differential mosaic loss of recent paralogues with identical functions is probably functionally unimportant, while differential mosaic loss of older paralogs with variant functions, which are not resolved by ortholog clustering programs, is problematic. Both sources of artifact (hidden paralogy and undersampling) can lead to an underestimation of the accessory genome and the complexity of the evolutionary history of pangenomes (99). Nevertheless, our inventories of core genomes will likely contract with increased genome sampling, while our inventories of genes within the accessory genome will expand as variant strains are sequenced. This is true for many prokaryotic pangenome sequencing initiatives but is especially true for eukaryotic datasets where sampling is limited.

Horizontal gene transfer (HGT): transfer of genetic material between reproductively isolated genomes

Conditionally dispensable chromosomes: chromosomes that are found in some fungi/eukaryotes and are required for virulence but not growth

Gene duplication: duplication of a region of DNA, generating one or more copies of a gene

For prokaryotes, a range of work has shown that accessory genomes are largely defined by horizontal gene transfer (HGT) [sometimes called lateral gene transfer (LGT)] acquisitions and mobile genetic elements (58). For eukaryotes, there is evidence for differential patterns of HGT across species (reviewed in References 32, 83) and, indeed, transfer of conditionally dispensable chromosomes (27, 28, 30, 51) (also called accessory, supernumerary, or B chromosomes). Yet, examples of these patterns of variation within species complexes (e.g., 24) are currently thought of as rare in eukaryotes. In contrast, current work has demonstrated that the major factors in fungal pangenome evolution are gene duplication and differential loss (56).

Regardless of the mode of acquisition, the gain and loss of genes can play a role in determining the niche of each strain (19, 20) and in some cases could lead to speciation by allowing colonization of alternative niches (i.e., allopatric speciation) or generating mating and recombination incompatibility (i.e., sympatric speciation). In such cases, modification of the accessory genome would effectively generate evolutionary paths that subdivide pangenomes: an evolutionary phenomenon that is poorly understood. Indeed, these gain or loss events must impact speciation events in prokaryotes and some eukaryotes when they drive the differential distribution of DNA uptake and exchange mechanisms (e.g., transformation and conjugation), restriction endonuclease and associated DNA methylation systems (61, 69), or mating compatibility loci (33), which can all act to disrupt gene flow and recombination, in turn leading to speciation. Study of pangenome dynamics in microbes is needed to understand how such patterns of variation feed into the processes that determine ecological ranges and speciation.

GENE ESSENTIALITY VERSUS BOLT-ON TRAITS

Assuming that we can sample enough strains to accurately define a core genome, the contents of a core genome may approximate to the set of genes that are essential (see the sidebar titled Essentiality) for cellular and life-cycle functions across the native environments of that species. Therefore, the core genome could be seen to represent genes that are rarely lost, because they are

ESSENTIALITY

An essential gene is one that is required for the survival of an organism. However, these genes may only be essential under given environmental conditions [i.e., conditionally essential, rather than absolutely essential (109)], within certain species (95) or strains (88), or because they encode components of protein complexes that themselves are only essential in certain species or conditions (90). Furthermore, essential genes may also be evolvable (48), with adaptive mutations permitting cell viability after gene loss. Together, these factors may explain why some essential genes can be found in accessory genomes (e.g., 76), as the terms essentiality and accessory are always condition dependent. Historically, laboratory experiments have provided a poor basis for determining both essential and accessory conditions, because laboratory experiments do not accurately reflect natural effective population sizes or the environmental heterogeneity of free-living microbial lineages.

functionally important, but also genes that are rarely replaced (even by gene duplication), because of compatibility issues. Alternatively, a gene could be required for growth (essential) but replaceable, which can explain why essential genes can be part of the accessory genome (see the sidebar titled Essentiality). In eukaryotes, nonessential genes are often overrepresented in the accessory genome [e.g., as seen in *S. cerevisiae* (56)], and therefore accessory genes may provide strains with niche-specific functions, such as antimicrobial resistance, or pathways encoded by metabolic gene clusters (96), which can, for example, function in overcoming competitors through toxin production, or in processing environment-specific metabolites. Such traits highlight the plastic nature of essentiality (specifically, conditionally essential genes). Understanding the relationship between truly essential and conditionally essential genes is an important challenge.

What factors characterize the core gene repertoire? The products of core genes are predicted to be components of complex protein-protein interaction networks or multi-protein complexes (9, 36, 60) (**Figure 1**), with eukaryotic core metabolic genes tending to have ancient archaeal ancestry, rather than bacterial ancestry (11). As core genes are integral to the function of protein complexes and wider protein networks, they are more likely to be indispensable (37). High protein connectivity also appears to determine a lower tolerance to gene duplication resulting in a reduced duplicability among these gene families (73), and restricting the distribution of such gene families to within the core genome, at least in *S. cerevisiae* (see also Reference 110 for further complexities in this pattern evident across different lineages). Duplication is thought to be suppressed in this way because it can lead to gene dosage effects, where asymmetric expansions of the protein components of interacting complexes perturb core system functions: This has been called the dosage-balance model (34, 41, 73). Interestingly, this problem is minimized when whole-genome duplications generate increased complexity, because proteome expansion is symmetric across interdependent proteins (31).

Experiments have tested the feasibility of parallel additions to protein networks among seemingly essential, core genes. For example, the study of elongation factor (EF-1α/EFL) and methionine adenosyltransferase (MAT/MATX) paralog distribution has led to the suggestion that some paralog pairs tend to show low tolerance for co-occurrence in the same proteome network (40, 97). These genes are ancient paralog pairs that tend not to occupy the same genome, suggesting an ancient gene duplication followed by loss, or HGT coupled with replacement. It is not clear whether these gene distribution patterns hint at mutual exclusivity (due to deleterious protein-protein interactions, or an imbalance in protein dosage within protein complexes, for example) or just differential loss due to redundancy. Szabová et al. (97) tested these scenarios directly by engineering *Trypanosoma brucei* to encode both paralogs, and then tuning out the native paralog

Protein-protein interaction network: network representing the physical association of proteins in an organism

Protein connectivity: the number of physical interactions a protein has with other proteins; interactions can differ across time and environment

Dosage-balance model: predicts gene duplications that increase the dosage of one protein could be rescued by overproduction of partner proteins

Whole-genome duplication: duplication events that copy the entire genome of an organism

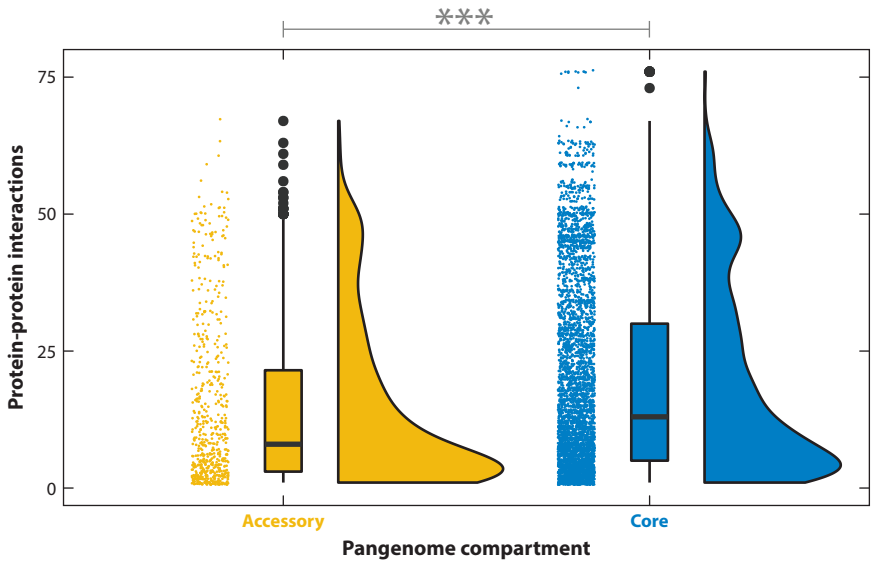

Figure 1

Protein-protein connectivity in the *Saccharomyces cerevisiae* S288C pangenome. Pangenome data were constructed from the genome sequences of 100 strains by McCarthy & Fitzpatrick (56), and connectivity data from BIOGRID (72) were collated by Cotton & McInerney (11). Core proteins have significantly higher connectivity (mean, 19.02 interactions) than accessory proteins (mean, 14.54 interactions). Statistical significance calculated using a Kruskal-Wallis test (triple asterisks indicate $p < 0.001$).

using RNA interference. These experiments showed that MAT/MATX could complement function but EF-1α/EFL could not, illustrating that there are no hard rules and that, in some cases, additions and replacements to seemingly core functions can be tolerated, but in others they cannot.

Collectively these observations infer that genes encoding proteins that (*a*) have low protein-protein interaction complexity, (*b*) are not essential across all niches occupied by the strain, and (*c*) are not part of multicomponent complexes are likely to be readily lost, gained, replaced, or expanded and are therefore key drivers of accessory genome variation. We name such genes and the phenotypes they encode bolt-on traits. Such gene complements should be predictable. For example, many metabolite transporter functions fit the characteristics of (*a*) nonessentiality, (*b*) being encoded by a single gene, and (*c*) having low protein-protein interaction networks (84).

EUKARYOTES HAVE HIGHLY VARIANT TROPHIC MECHANISMS THAT MAY INFLUENCE PANGENOME CHARACTERISTICS

Eukaryotes possess a striking range of mechanisms for obtaining metabolites. These can be classified as (*a*) phototrophic, (*b*) osmotrophic, and (*c*) phagotrophic. The latter two categories are broadly termed heterotrophic. It is also important to note that within the heterotrophic classification a large diversity of parasitic lifestyles is displayed, with for example, obligate intracellular parasites, obligate biotrophs, necrotrophs, and facultatively parasitic heterotrophs. Many of these categories show distinct, but noisy, patterns of genome evolution (for some cross comparisons, see for example References 4, 18, and 39). Furthermore, in real terms, phototrophic eukaryotes often possess a range of heterotrophic functions (12, 59) and are therefore mixotrophic. As such, boundaries of taxonomic classification based on trophic mechanism [for example, favored by Whittaker (105)] break down rapidly when we consider the wider functions of microbial cells. Current

Phototrophic: describes organisms that utilize light to generate an electron transport chain to produce chemical energy

Phagotrophic: describes organisms that feed by engulfing and ingesting food particulates or other organisms

Heterotrophic: describes organisms that cannot produce their own key metabolites, instead relying on the consumption of organic matter produced by others

understanding of eukaryotic pangenomes across these different feeding ecologies is especially poor. It is therefore difficult to understand how pangenome dynamics can vary between species with variant ecologies. Understanding such patterns is critical for understanding pangenome evolutionary dynamics and will likely reveal much about how adaptation occurs across different lifestyles.

TRANSPORTER PROTEINS DEFINE MICROBIAL NICHES

Multiple aspects of many metabolic networks require uptake of precursor molecules into the cell. In some ways every cell on this planet is heterotrophic, and as such, in the absence of diffusion or permeation across cell membranes, must rely on the function of transporter proteins either on the surface of the cell or within the endosome/phagosome (a privatized space for processing metabolites). Transporters are composed of proteins that form transmembrane helices that fold into the plasma membrane to form a pore or carrier protein that docks with a target substrate and facilitates movement across the membrane. This can be a passive or an active process (requiring ATP). In some cases [e.g., ABC transporters (80)], the transporter is composed of different protein domains, whereas in other cases transporters can be composed of multiple copies of related proteins [e.g., the *S. cerevisiae* ammonium transporters, Mep1–3 (54)] or a single predominant protein [e.g., the urea transporter in *Candida albicans*, Dur3 (65)].

Because transporter function and the supply of primary substrates lie at the root of nearly all metabolic networks, the complement of transporters encoded by a microbial cell underpins and shapes (*a*) the biochemical network of the cell, (*b*) the environment a microbe can colonize, and (*c*) how a microbe competes for resources with others. All these points are important for defining the ecology of microbes. The importance of point *c* has been demonstrated by Brown et al. (8). *S. cerevisiae* cells can utilize hexose sugars for both respiration and fermentation (71, 101) and can also tolerate a broad range of sugar concentrations. Accordingly, *S. cerevisiae* encodes a suite of hexose transporters with different uptake kinetics (6) that are expressed depending on sensing of the carbon-source availability in the extracellular environment (87). In this way the diversity of hexose transporters, and how they are transcribed, determines the ecology of variant *S. cerevisiae* strains. Experimental evolution of *S. cerevisiae* strains has demonstrated that the hexose transporter gene repertoire can be expanded, leading to additions to the accessory genome, increased hexose transporter transcription levels, and enhanced hexose uptake kinetics. This evolutionary outcome occurs in response to changes in hexose availability in the environment (8) and provides the evolved strain with an advantage over the parental strain under low-glucose conditions. **Figure 2** summarizes these experimental results.

A FIRST ASSESSMENT OF THE PREVALENCE OF TRANSPORTER-ENCODING GENES IN EXEMPLAR EUKARYOTIC PANGENOME DATASETS

Accessory genomes have been posited to include genes that can add functions. Such accessory genes have been linked to pathogenicity in fungi, facilitate host range expansion (13), and contribute to the virulence of plant pathogens (2). Within accessory genomes, genes encoding proteins with lower connectivity are enriched (e.g., **Figure 1**), and such genes can be acquired by mechanisms such as HGT, gene duplication (56), and de novo generation (100) and can facilitate adaptation to variant environments (60).

Transporter proteins typically exhibit low protein-protein connectivity (60, 84), can manifest a trait via a single gene acquisition, and can replace entire biochemical synthesis networks by allowing uptake of preformed metabolites (e.g., 15, 29). These factors suggest that accessory genomes

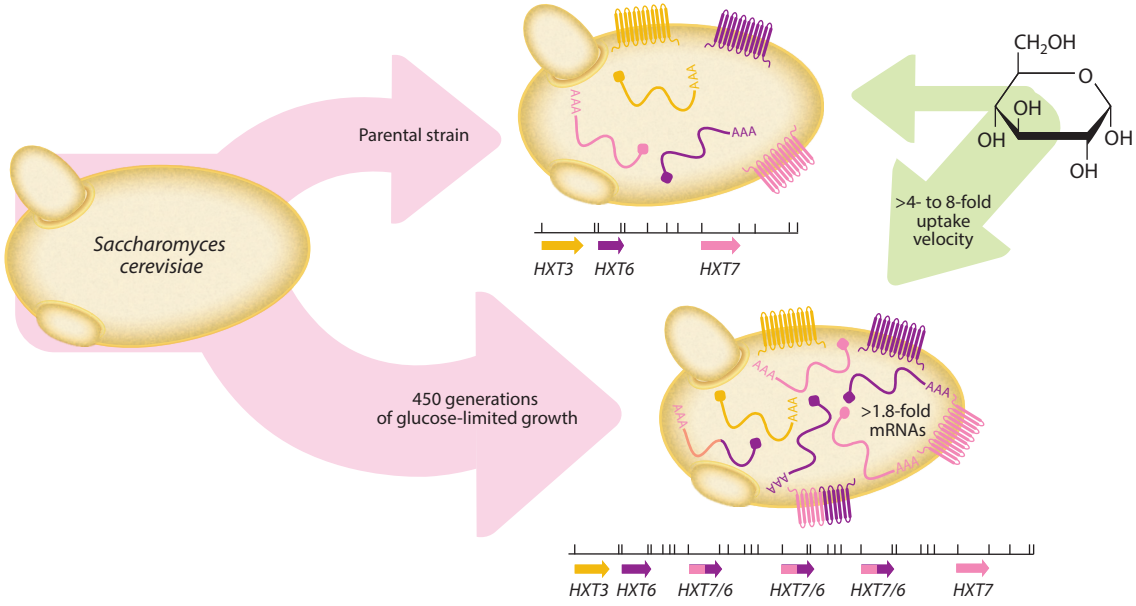

Figure 2

Schematic representation of the results obtained by Brown et al. (8). Pink arrows represent both the parental strain and the strain after 450 generations of glucose-limited growth. HXT transporter proteins (drawn as transmembrane proteins in the yeast cell), mRNAs (drawn within the yeast cell and marked by poly(A) tails), and genes (represented in tandem on chromosome IV) are represented by yellow (HXT3), purple (HXT6), pink (HXT7), and pink/purple (HXT7/6). Glucose uptake is represented by its chemical structure and the uptake velocity of each strain is represented by the size of the green arrows (not to scale).

may be more likely to contain bolt-on transporter traits. Furthermore, it may be more expeditious for an organism to gain, or duplicate, a transporter protein gene that integrates within a simple preexisting cellular network than to gain an entirely novel biosynthetic function. This latter route may take a significant amount of generations and mutations to be integrated into the regulatory network of the cell (43). Unlike the protein complexes generally encoded in the core genome, functions driven by transporter proteins can tolerate gene duplication and the associated increase in protein dosage (discussed above in the example of the *S. cerevisiae* hexose transporters). Indeed, because transporter function is directly related to ecological resource competition, rapid increases in protein dosage due to gene duplication can result in a competitive advantage (8, 60). Even in cases where rapid accumulation of metabolic resources within the cell, brought on by altered transporter function, can limit cellular physiology, such traits can still be selected for, because the rapid accumulation of a metabolite allows such cells to outcompete cellular forms with lower accumulation rates. In such cases, stockpiling of metabolic resources can be a selfish act of an individual to the detriment of overall community productivity (47, 52). In this way transporters can be expanded and then lost, for example, in the absence of resource competition driving selection for increased protein dosage and selfish uptake.

Gene duplication, in addition to protein dosage effects, provides the foundations for subfunctionalization and neofunctionalization (68). In such cases, sister paralogs of transporters can gain a range of variant functions, including (*a*) altered transportation kinetics (38, 54), (*b*) altered substrate preferences (38, 50), and (*c*) variant transcriptional profiles in response to life-cycle or environmental conditions (53). All these different functions are variants of previously established cellular phenotypes, and so they allow complexity to be added in a manner consistent

with the idea of a bolt-on trait. They also allow fine-tuning of metabolic responses to environments and so provide competitive advantage, while further amending the accessory component of the pangenome.

Taken together, these factors suggest that transporter proteins are prime candidates for gene gain and, in turn, loss [dependent on substrate availability and ecological competition (60)], meaning that transporter repertoires appear patchy (21, 60). We have previously discussed how functional ecological dimensions to transporter protein repertoire evolution can drive patterns of gain, replacement, and loss, much like a gene ratchet (60), thereby driving complex patterns of paralog evolution. This concept is summarized in **Figure 3** and shows how variant environments drive different patterns of selection for transporter gene repertoire. In such cases, loss, as well as gain, can determine the breadth of environments a strain can colonize. Such patterns of selection will be further complicated when the gene-environment dynamic includes competition for resources between strains or species and where transporter evolution can act to optimize competition for resources (such additional dynamics are not shown in **Figure 3**).

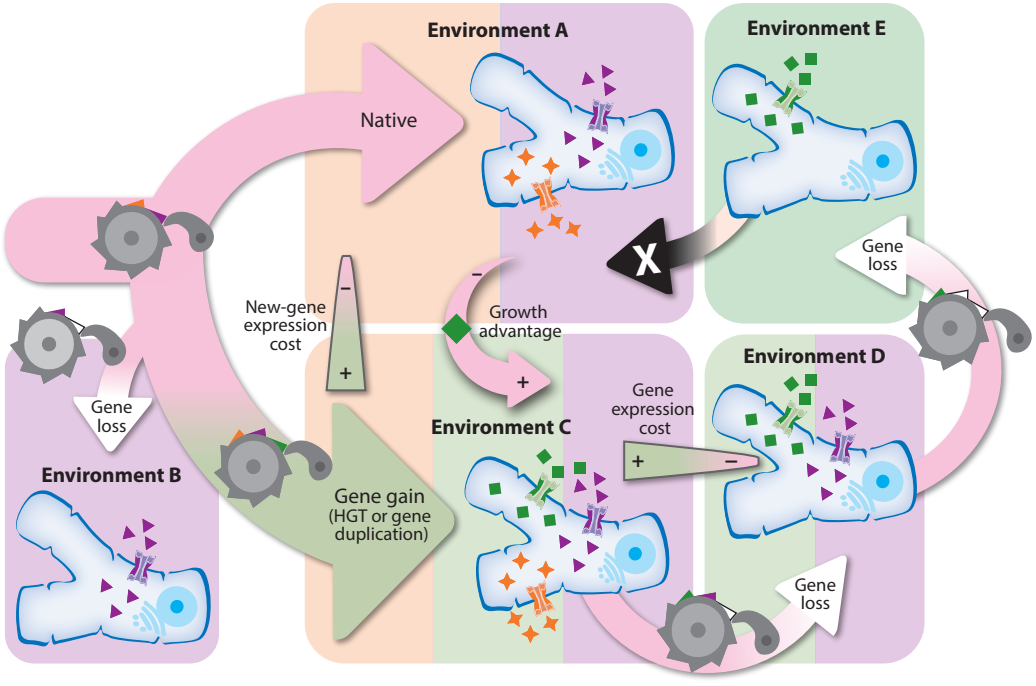

Figure 3

Schematic representation of the model proposed by Milner et al. (60) for the evolution of transporter genes within a related lineage. Light pink arrows represent the native population, pink-white arrows represent a population after gene loss, pink-green arrows represent a population that acquired a gene, and pink-black arrow indicates a population that cannot colonize its ancestral environment. The ratchet illustration represents the loss (*white*) or presence/gain of three different transporter genes (*orange*, *purple*, and *green*). The metabolites transported by each protein are represented by stars (*orange*), triangles (*purple*) or squares (*green*). Boxes represent environments in which each population evolves, and the colors of the boxes represent the metabolites present in them. This illustration of the transporter/environment ratchet model demonstrates how different environments can drive patterns of transporter gene gain/loss, but it does not show the effect of strain/species competition, which would further complicate such evolutionary dynamics. Abbreviation: HGT, horizontal gene transfer.

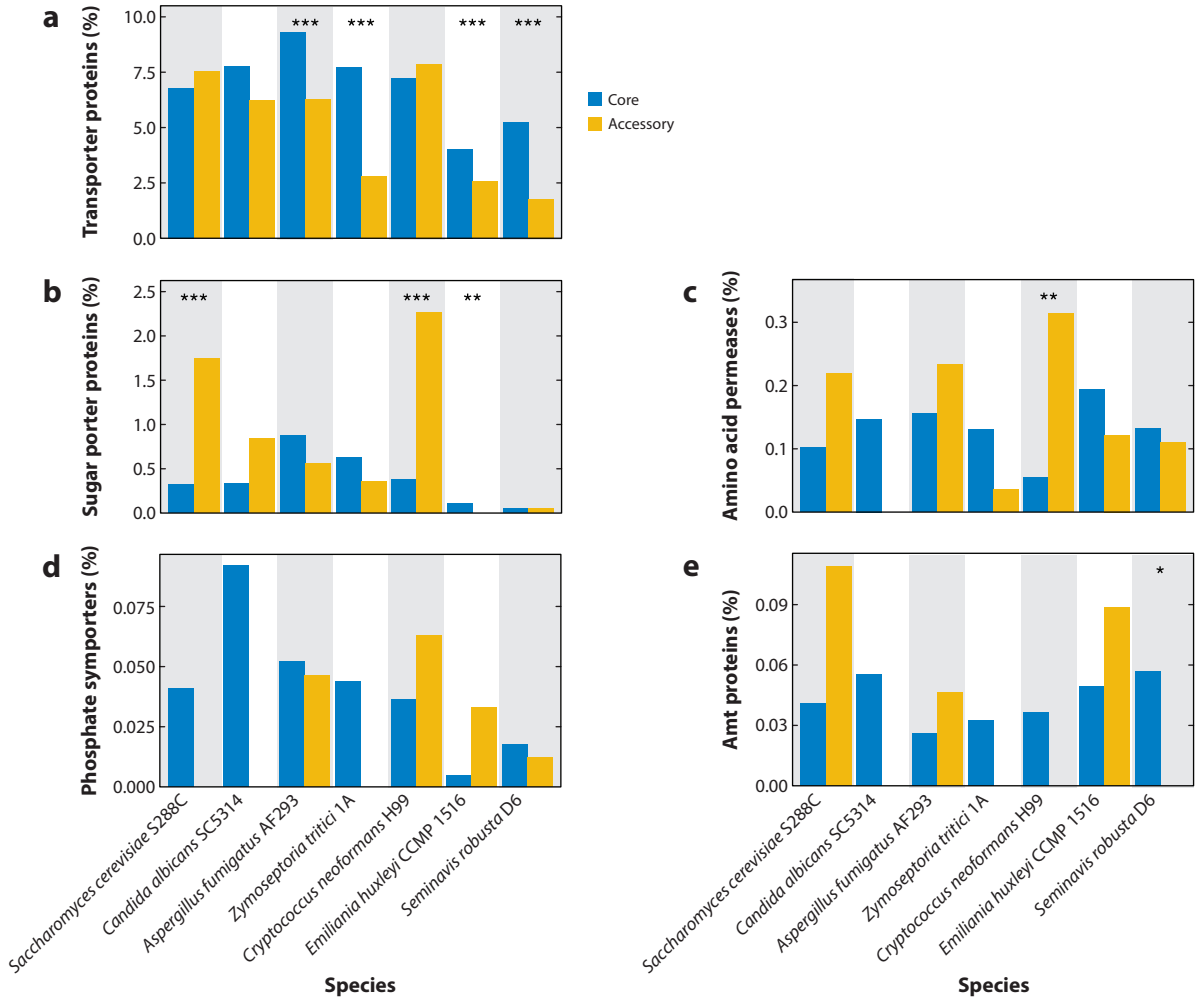

Figure 4

Transporter proteins are enriched in some eukaryotic core genomes, while specific functional classes of transporter proteins are enriched in some accessory genomes. (*a*) Proportion of transporter proteins [i.e., those proteins with ≥4 transmembrane domains, as classified using TMHMM (42), and with a significant ($<1e^{-5}$) TCDB hit (91)] within the core/accessory genomes of each species. (*b*) Proportion of sugar porters (TCDB 2.A.1.1; see Reference 91), (*c*) amino acid permeases (TCDB 2.A.18), (*d*) phosphate:H^+ symporters (TCDB 2.A.1.9), and (*e*) ammonium transporters (Amt proteins) (TCDB 1.A.11) within the core/accessory genomes. All data are represented as a proportion of the total number of genes within each core/accessory dataset. Raw data from References 3, 56, 70, and 79; statistical significance calculated using a χ^2 test (*single asterisks* indicate $p < 0.05$; *double asterisks* indicate $p < 0.01$; *triple asterisks* indicate $p < 0.001$).

EVIDENCE FOR TRANSPORTER FAMILY ELABORATION IN EUKARYOTIC PANGENOME DATASETS

To explore the idea that transporter families are discontinuous in eukaryotic pangenomes, we reviewed the distribution of transporter gene families in published small-scale pangenome datasets from a range of microbial eukaryotes (3, 56, 70, 79) (**Figure 4**; **Table 1**). As discussed above, this data selection is restricted by the lack of available pangenome datasets for eukaryotic microbes, and in many cases it incorporates only a small number of genome comparisons. However, with

accessory genomes likely to contain many strain-specific functions, it is interesting to investigate what transporter-encoding genes are present within the sampled pangenomes. We also conducted this survey because, although the diversity of eukaryotic trophic mechanisms represented is limited, we were interested to study the heterogeneity of transporter classes evident in the accessory genome. This form of variation is informative for understanding the evolution and ecology of different species complexes.

Broad enrichment of genes involved in specific transport processes in accessory genomes has been shown to be the case in some fungi (e.g., 56). However, the widespread expansion, specifically of transporter-encoding genes, in eukaryotic accessory genomes does not seem to be consistent; different taxa show different functional classes of transporter present in the accessory genome (**Figure 4**). This, in part, is explained by the different ecological and trophic characteristics of the organisms compared here, with most of the organisms sampled being heterotrophs (e.g., osmotrophs, 85) while only two mixotrophic algae (*Emiliania huxleyi* and *Seminavis robusta*), which are able to fix carbon de novo, are sampled (**Table 1**). Therefore, unsurprisingly, the osmotrophic consumers show variation in the complement of predicted sugar transporters within the accessory genomes, while the phototrophs, which obtain sugar through photosynthesis, show a stark reduction in predicted sugar transporters, with very few encoded in the accessory genome (**Figure 4*b***).

The lack of a broad enrichment of transporter-encoding genes across accessory genomes may also reflect the specific ecological requirements of these strains, which influence the types of function present in the accessory genome. For example, where a species occupies a niche that is characterized by variable or limiting nutrients (79, 86), we would expect to see a prevalence of variant transporter genes within the accessory genome that encode transporters for these same nutrients. Indeed, transporter proteins for ammonium and phosphate are prevalent in the accessory genome of *E. huxleyi* (e.g., 79) (**Figure 4*d*,*e***), reflecting the need for this alga to compete with other organisms for effective uptake of limited nitrogen and phosphate sources in the marine environment and maybe also varying ecological ranges of different *E. huxleyi* strains (44). On the contrary, such pressures do not seem to be reflected in the accessory genomes of many of the fungi sampled, which often occupy environments where access to nitrogen and phosphate sources is not a limiting factor.

Similarly, among osmotrophic organisms, the accessory genomes of *S. cerevisiae* S288C and *Cryptococcus neoformans* H99 are significantly enriched for sugar transporters (**Figure 4*b***), and the latter also has a significant enrichment of amino acid permeases (**Figure 4*c***). This may reflect the frequent habitat (and, therefore, nutrient) shifts of *C. neoformans* and the need to take up amino acids for survival in host-associated environments (discussed further below). Together, these findings suggest that the enrichment of transporter-encoding genes in accessory genomes is likely to be targeted to specific subsets of uptake functions, which is, in turn, determined by two factors: (*a*) the requirements of the core metabolism and (*b*) the availability of compounds in the environment usable for metabolic function.

In contrast, there are also instances where a higher proportion of transporter proteins are encoded within the core genome. *Yarrowia lipolytica*, for example (57), demonstrates a pattern of enriched transporter repertoire within the core genome, likely reflecting the large-scale utilization of hydrophobic substrates by this nonconventional, oleaginous yeast, which can catabolize alkanes, fatty acids, and triacylglycerols and is dependent on transport of these compounds into the cell (22) using a core set of membrane transporters. Transporter-gene enrichments in core genomes are also seen for *E. huxleyi*, *Zymoseptoria tritici*, *Aspergillus fumigatus*, and *S. robusta* (**Figure 4*a***); this could reflect the larger accessory genomes in these organisms [of the seven pangenomes analyzed in this review (**Table 1**), these four had the largest accessory genomes]. *Z. tritici*, for example, is

predicted to contain a large number of effector proteins in the accessory genome (75), which is skewing these proportions.

Blood-brain barrier: selectively permeable endothelial barrier between the capillaries and tissue of the brain, protecting nervous tissue from infection

In summary, the comparisons of transporter repertoires provide insights into the typical lifestyles of each species/strain and point toward specific metabolites that have necessitated transporter gene expansion within the accessory genome. Functional analysis of the nature of these acquisitions could be important for further understanding how pathogenicity may arise (e.g., 92) and how organisms are able to adapt to varying environmental conditions. Furthermore, it is important to note that global comparisons, as shown in **Figure 4*a***, are likely to hide important patterns of variation within specific functional classes of transporters, species complexes, or related lifestyles.

WHAT ACCESSORY GENOME TRANSPORTER REPERTOIRES CAN TELL US ABOUT ENVIRONMENTAL NICHES AND PATHOGENICITY: SOME EXAMPLES

As mentioned above, within some of the fungal pangenomes sampled (*C. neoformans*, *S. cerevisiae*), sugar transporters are shown to be prevalent in the accessory genomes (**Figure 4*b***), suggesting that expansion of these gene families may have evolved in response to varying levels of sugar availability in the environment or have evolved novel substrate uptake functions. Such additions could be important factors in the expansion of ecological niches of these microbes and, in turn, could facilitate the evolution of pathogenicity.

Metabolism of one such sugar, inositol, is important for traversal across the blood-brain barrier by *C. neoformans* (49), an organism that causes in excess of 180,000 deaths worldwide annually (77). Inositol is an abundant metabolite in the human brain, with a 200-fold greater concentration than in blood plasma (23). Significantly, *C. neoformans* H99 has an expanded inositol transporter (Itr) repertoire (108) and, interestingly, the majority of these *Itr* genes, as well as the two inositol transporters shown to be functional (**Table 2**), are found in the *C. neoformans* accessory genome (data from 56). More critically, the crossing of the blood-brain barrier by this fungal pathogen is dependent on these two transporters (49, 108), suggesting a strong association between accessory genome transporter function and *C. neoformans* virulence. Adding further weight to this argument is that the *Cryptococcus gattii* R265 strain has a reduced *Itr* gene repertoire, with only six homologs (108) (**Table 2**). *C. gattii* R265 is also lacking the

Table 2 Inositol transporters in the *Cryptococcus neoformans* H99 pangenome, and comparison with the *Cryptococcus gattii* R265 Itr repertoire

Gene	*C. neoformans* H99: accessory/core[a]	Number of strains	*C. gattii* R265 present[b]
ITR1	Core	25/25	+
ITR1A	Accessory	7/25	–
ITR2	Accessory	24/25	+
ITR3	Accessory	20/25	+
ITR3A	Accessory	23/25	–
ITR3B	Accessory	19/25	–
ITR3C	Accessory	18/25	–
ITR4	Core	25/25	+
ITR5	Accessory	24/25	+
ITR6	Core	25/25	+

[a]Presence in the accessory/core pangenome for each gene in *C. neoformans* H99.
[b]Presence/absence for each gene in *C. gattii* R265, determined using the data from Reference 56. *Itr* genes shown to be functional by complementation experiments (108) are highlighted in gray.

Table 3 Amino acid/auxin permease (TCDB 2.A.18) and amino acid–polyamine-organocation (TCDB 2.A.3) proteins encoded in the *Cryptococcus neoformans* H99 pangenome[a]

Gene ID	Accessory/core	Predicted transmembrane domains	TCDB prediction/ annotation[b]	Differential expression (46)[c]	
				Mouse lungs	Monkey lungs
AAAP (TCDB 2.A.18)					
CNAG_01904	Core	10	Vacuolar Avt6	—	—
CNAG_03866	Core	11	Vacuolar Avt1	Up	—
CNAG_03988	Core	9	Vacuolar Avt2	—	—
CNAG_05996	Accessory	10	Vacuolar Avt1	—	Up
CNAG_06541	Accessory	11	Neutral Aap1	—	Down
CNAG_05685	Accessory	11	Neutral Aap1	—	—
CNAG_06828	Accessory	10	Vacuolar Avt3	—	—
CNAG_01074	Accessory	11	Neutral Aap1	—	Up
APC (TCDB 2.A.3)					
CNAG_02539	Core	12	Aap1	—	Down
CNAG_07902	Core	12	Aap2	—	—
CNAG_01118	Core	12	Aap3	—	—
CNAG_00597	Accessory	12	Aap4	—	Up
CNAG_07367	Accessory	12	Aap5	—	Up
CNAG_07449	Core	10	Aap6	—	—
CNAG_05345	Core	12	Aap7	—	—
CNAG_00574	Core	11	Aap8	—	—

Abbreviations: AAAP, amino acid/auxin permease; APC, amino acid–polyamine-organocation.

[a]Genes shown to be important for virulence in the invertebrate infection model, *Galleria mellonella* (55), are highlighted in gray.

[b]Annotations for APC proteins are based on Reference 55.

[c]Dashes indicate that these genes were not significantly differentially expressed in the corresponding animal infection model (46).

two, aforementioned, inositol transporters shown to be functional and important for colonization of neural tissues during *C. neoformans* infections (**Table 2**). It is interesting, therefore, that *C. neoformans* H99 infections tend to cause meningoencephalitis, while *C. gattii* R265 infections tend to cause pulmonary infections (66), indicating that *C. neoformans* is equipped to disseminate to, and proliferate within, neural tissues. Together, this comparison between the transporter repertoires of these two species suggests that the expansion of these *Itr* genes in the accessory genome may be one of the reasons for this difference in pathogenicity.

The amino acid/auxin permease (TCDB 2.A.18) family of proteins is also enriched in the *C. neoformans* H99 accessory genome (**Figure 4*c***) (**Table 3**). Two predicted amino acid permeases in this accessory genome (CNAG_01074 and CNAG_05996) are upregulated in the lungs of a monkey infection model (46), suggesting that these may also be important for establishing infection. Two additional accessory genome–encoded amino acid transporters in the amino acid–polyamine-organocation (APC) superfamily (TCDB 2.A.3), Aap4 and Aap5, are also upregulated in this monkey infection model and are required for *C. neoformans* virulence in the infection model, *Galleria mellonella* (55), suggesting that the transition into a primate host necessitates a shift in nutrient uptake and that amino acid transporters encoded by the accessory genome could be a key component in facilitating this shift in niche.

Meningoencephalitis: inflammation of both the meninges—the membranes that protect the brain and spinal cord—and the brain

In *S. cerevisiae*, ammonium (NH_4^+), a nitrogen source that can support optimal yeast growth, is transported by the products of the three nonessential *mep* genes, *mep1*, *mep2*, and *mep3* (54). The *mep1* and *mep3* genes are both in the core genome of *S. cerevisiae* (56), while the gene encoding the highest-affinity transporter, Mep2 (54), is found in the accessory genome [present in 99 of 100

Table 4 Ato transporters in the *Candida albicans* SC5314 pangenome

Gene	Accessory/core	Number of strains
ATO1	Core	34/34
ATO2	Core	34/34
ATO3	Accessory	32/34
ATO4	Core	34/34
ATO5	Core	34/34
ATO6	Core	34/34
ATO7	Core	34/34
ATO8	Core	34/34
ATO9	Accessory	33/34
ATO10	Accessory	7/34

Abbreviation: Ato, ammonium transport outward.

S. cerevisiae strains, but absent in YJM1415 (56) and, in a separate study, present in 1,341 of 1,360 *S. cerevisiae* strains, although Li et al. (45) categorized this gene in the extended core, as this gene is present in >95% of strains]. It is important to note that different studies will define alternative thresholds for categorization into core versus accessory genomes and that sequencing completeness also needs to be considered when allocating genes to the accessory genome. Nevertheless, it is interesting to speculate that this gene, upregulated under limiting concentrations of ammonium (54), may be lost in environmental conditions with replete nitrogen sources, which therefore no longer necessitate retention of this transporter-encoding gene. It also raises the question as to whether accessory genome transporters tend to have higher affinities than homologous proteins in core genomes. To our knowledge, this has not yet been investigated in eukaryotes, yet it would be congruent with the requirement for additional transporter proteins in low/variable-nutrient environments [such as the HXT6/7 duplications described by Brown et al. (8)] and the utilization of a higher-affinity ammonium transporter by a virus to manipulate host metabolism (62).

Expanded repertoires of transporter-encoding genes are also present in the pangenome of the fungal pathogen *C. albicans*. The ammonium transport outward (Ato) transporters are one example of an expanded transporter family (**Table 4**), with members of this family having roles in the alkalinization of the acidic phagolysosome (102), allowing the transition of *C. albicans* from yeast into a hyphal form, and subsequent escape from macrophages (14). Note that this is not a direct metabolite uptake trait but a trait that alters the surrounding environment. Importantly, the number of Ato proteins has been shown to correlate with the level of alkalinization of a species (102), with proteins encoded by genes within the accessory genome shown to be functional [e.g., Ato10 (14)], suggesting a correlation between protein family expansion, transcriptional dosage, and pathogenicity.

Taken together, these results suggest that the accessory genome encodes subsets of transporter proteins important for regulating virulence and modifying the environment of pathogenic infections. In many instances, the transporter gene families show a level of discontinuity that is consistent with gene acquisition ratchets (**Figure 3**).

BOLT-ON TRANSPORTERS ARE EXPLOITED BY VIRUSES

Given the examples of transporter function, and associated virulence, in the fungal pathogens discussed above, we also predict that transporter protein functions are important for virulence in other systems. The genes transferred between eukaryotes and viruses are predicted to encode

diverse functions across viral lineages but are enriched for metabolic processes, such as nutrient uptake (35). Indeed, there is consistent evidence that transporter proteins are encoded by a variety of viruses (26, 35, 62, 63), especially viruses of marine eukaryotic phytoplankton (**Table 5**). Transporter proteins encoded by viruses have also been shown to manipulate eukaryotic hosts during infection (62) by directly altering host uptake kinetics of target metabolites. Growth, and viral

Table 5 Examples of putative transporter proteins encoded in viral genomes[a]

Protein family	Viral recipient	Closest eukaryotic relative	Annotation (PANTHER/eggNOG)
C000111	Unclassified Phycodnaviridae	Chlamydomonadales	Aquaporin transporter
C000111	Nucleocytoviricota	Eukaryota	Aquaporin transporter
C000146	Coccolithovirus	Prymnesiophyceae	Ammonium transporter
C000146	Nucleocytoviricota	*Aureococcus anophagefferens*	Ammonium transporter
C000146	Prasinovirus	*Ostreococcus*	Ammonium transporter
C000146	Unclassified Phycodnaviridae	Mamiellales	Ammonium transporter
C000146	Nucleocytoviricota	Alveolata	Ammonium transporter
C000350	Coccolithovirus	Eukaryota	Nucleotide-sugar transmembrane transporter
C000350	Nucleocytoviricota	Eukaryota	Nucleotide-sugar transmembrane transporter
C000350	Coccolithovirus	Eukaryota	Nucleotide-sugar transmembrane transporter
C000350	Coccolithovirus	*Chrysochromulina tobinii*	Nucleotide-sugar transmembrane transporter
C000489	Klosneuvirus	Eukaryota	Organic solute transporter–related
C000489	Unclassified Mimiviridae	SAR	Organic solute transporter–related
C000560	Nucleocytoviricota	Trypanosomatidae	Phosphate transporter
C000560	Unclassified Phycodnaviridae	Chlamydomonadales	Phosphate transporter
C000560	Unclassified Phycodnaviridae	*Aureococcus anophagefferens*	Phosphate transporter
C000560	Mimiviridae	Archaeplastida	Phosphate transporter
C000560	Nucleocytoviricota	Opisthokonta	Phosphate transporter
C000560	Nucleocytoviricota	Eukaryota	Phosphate transporter
C000628	Nucleocytoviricota	Eukaryota	Folate-biopterin transporter 1
C000628	Medusaviridae	Eukaryota	Folate-biopterin transporter 1
C000731	Unclassified Phycodnaviridae	Eukaryota	Solute carrier family 34 sodium phosphate, member 2–related
C000731	Unclassified Phycodnaviridae	SAR	Solute carrier family 34 sodium phosphate, member 2–related
C000989	Nucleocytoviricota	Eukaryota	Ammonium transporter
C001122	Nucleocytoviricota	*Amoebophrya* sp. AT5.2	Fe^{2+}/Mn^{2+} transporter PCL1
C001122	Megaviricetes	Stramenopiles	Fe^{2+}/Mn^{2+} transporter PCL2
C001122	Iridovirus	*Bodo saltans*	Fe^{2+}/Mn^{2+} transporter PCL3
C001411	Percavirus	Vertebrata	Equilibrative nucleoside transporter
C001494	Raphidovirus	Stramenopiles	Molybdate transporter (major facilitator superfamily)
C002397	Coccolithovirus	Eukaryota	Major facilitator superfamily transporter
C003908	Unclassified Phycodnaviridae	Chlorophyta	Choline transporter–like (SLC family 44)
C004505	Marseillevirus LCMAC202	Eukaryota	Transmembrane family, TMEM144 of transporters

Abbreviation: SAR, stramenopiles, alveolates, and rhizaria.

[a]Data from Reference 35. Phylogenetic trees for all examples available at **https://itol.embl.de/shared/nirwin**.

productivity, in hosts is often limited by access to nutrients, so it is likely that viruses are manipulating transporter function to increase host metabolism, and therefore viral productivity. Tied to this function, it is possible that virus-encoded transporters are acting as bolt-on traits, with their transportation functions being exploited by viruses. The low connectivity of transporters (discussed above) therefore enables viruses to encode single proteins that can manipulate host nutrient uptake and associated metabolic functions, driving viral fitness gains.

Proteins encoded by viral genomes also complicate the concept of eukaryotic pangenomes, in the same way that genes encoded by plasmids and phages complicate the classification of prokaryotic pangenomes (7, 89), with widespread exchange of genes between eukaryotes and their viruses (35) potentially acting as an additional accessory gene reservoir for eukaryotic organisms. Transporter genes, with lower connectivity facilitating a propensity for HGT (35, 60, 84), are then also presumably able to be transferred back from viral to eukaryotic genomes (i.e., reacquired by eukaryotic hosts), with the viral genomes, therefore, contributing to eukaryotic accessory genome diversity. These data (**Table 5**) therefore suggest that viruses can act as reservoirs for transporter-encoding genes within pangenomes. As part of this role, it is not clear how residency within the viral pangenome may drive functional diversification of such traits, potentially as a result of the higher mutation rates of virus-encoded genes (74). However, it is likely that increased rates of variation may feed into diversification of transporter protein function when the transporter-encoding genes are residing in the virus-derived pangenome.

CONCLUDING REMARKS

Eukaryotic microbes have pangenomes, which are shaped by a range of factors including HGT, gene duplication, and gene loss. Currently, these patterns of genomic variation are not clearly defined, largely because our sampling of genomes within eukaryotic species complexes is minimal. A coordinated sampling effort across a range of species with different biologies, ecological ranges, and trophic mechanisms is needed. We note, however, that comparisons are difficult, as such a strategy for pangenome sampling needs to include comparisons of species with both narrow and vast environmental ranges. This latter point is especially challenging because we understand the environmental ranges of very few eukaryotic microbes that are not obligate parasites. We also do not understand the nature of species boundaries for many microbial eukaryotes. Environmental meta-genome sequencing and single-cell genome sequencing may help to solve the former problem, but the requirement for genome completeness may limit the utility of these approaches. Understanding species boundaries for many microbial eukaryotes, however, will likely remain a substantial obstacle.

In many cases pangenome datasets span diverse isolates; for example, in the case of the *Emiliania* pangenome discussed, the genomes sampled from this group span 2% variation in ribosomal DNA marker sequences (79), a level that, in many cases, corresponds to variation within a genus. Comparing and contrasting variation within nonequally defined groups will be challenging. Here, rules are needed to help guide the field with regard to legitimate definitions of pangenomes relative to asexual species, sexual species, and genus groups. Furthermore, pangenome sequencing initiatives need complete, reference-grade genome assemblies to minimize artifactual predictions of the accessory genomes, and highly precise ortholog clustering approaches to avoid underprediction of hidden paralogy and accessory genome evolutionary dynamics.

Armed with greater sampling of eukaryotic pangenomes, we will be able to understand the roles of HGT, gene loss, and gene duplication in eukaryotic pangenome evolution. In the absence of clearly demarked species boundaries, patterns of variation can be related to strain phylogeny. This will help us to understand the relative roles these evolutionary factors play, and importantly,

we will be able to understand what functional constraints retard or drive these mechanisms as they generate accessory genome variation. Such studies can then be combined with functional protein approaches to understand how these different pathways for generating genomic diversity are shaped by the cellular functions of the proteins encoded.

In this article we have advocated the use of pangenome analyses for understanding the evolution and ecology of eukaryotic lineages. We hypothesize that such studies will be important for understanding how lineages evolve in response to their niche, their host, and their competitors, and we discuss how transporter proteins are an important factor for all these considerations. We outline how this is not a general phenomenon for all transporter proteins but rather heterogeneity in the transporter complements in accessory genomes is an important factor for understanding the evolutionary ecology of the strains studied. Related to this, we provide examples in a range of fungi and relate this to cases where transporters have played a role in understanding pathogen evolution. If, as we predict, transporters represent important traits for the ecology of infectious disease–causing microbes, they may also represent potential drug targets, because their function is likely to be important to how microbes colonize host environments and, therefore, how they are able to cause disease.

SUMMARY POINTS

- Eukaryotic pangenomes require systematic study and increased sampling efforts.
- The role of horizontal gene transfer (HGT) and gene duplication needs to be understood across divergent taxonomic and ecological contexts.
- Because of a range of functional considerations, gene duplication is likely to be a key driver of eukaryotic pangenome dynamics.
- An inconsistent repertoire of transporter protein classes will represent a key aspect of many accessory genomes.
- Transporter protein repertoire evolution drives microbial ecology and evolution and is an important consideration for understanding the evolution of virulence.

FUTURE ISSUES

- Many lab-strains, particularly for model systems, have been maintained in stable, nutrient-rich culture conditions for millions of generations, purging genome variation: Are we systematically underestimating accessory genome complexity?
- As accessory genes are acquired, what is the effect on associated protein networks?
- How do functional aspects of acquisition vary between gains by gene duplication and gains by HGT?
- How do aspects of cellular network function vary between gains of novel functions and gains to established functional networks?
- How does transporter protein acquisition relate to functional networks, microbial ecology, and pathogen evolution?

- Do transporter proteins encoded in accessory genomes tend to have higher affinity for substrate (or any other difference in functional transportation kinetic parameters) than those in core genomes, and can such trends inform synthetic biology?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Andreani NA, Hesse E, Vos M. 2017. Prokaryote genome fluidity is dependent on effective population size. *ISME J.* 11(7):1719–21
2. Badet T, Croll D. 2020. The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr. Opin. Plant Biol.* 56:65–73
3. Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol.* 18(1):12
4. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, et al. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330(6010):1549–51
5. Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, et al. 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci. Adv.* 3(7):e1700239
6. Boles E, Hollenberg CP. 1997. The molecular genetics of hexose transport in yeasts. *FEMS Microbiol. Rev.* 21(1):85–111
7. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. 2019. The ecology and evolution of pangenomes. *Curr. Biol.* 29(20):R1094–103
8. **Brown CJ, Todd KM, Rosenzweig RF. 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* 15(8):931–42**
9. **Cohen O, Gophna U, Pupko T. 2011. The complexity hypothesis revisited: Connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* 28(4):1481–89**
10. Cordero OX, Ventouras L-A, DeLong EF, Polz MF. 2012. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *PNAS* 109(49):20059–64
11. Cotton JA, McInerney JO. 2010. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *PNAS* 107(40):17252–55
12. Croft MT, Warren MJ, Smith AG. 2006. Algae need their vitamins. *Eukaryot. Cell.* 5(8):1175–83
13. Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLOS Pathog.* 8(4):e1002608
14. Danhof HA, Lorenz MC. 2015. The *Candida albicans* ATO gene family promotes neutralization of the macrophage phagolysosome. *Infect. Immun.* 83(11):4416–26
15. Dean P, Sendra KM, Williams TA, Watson AK, Major P, et al. 2018. Transporter gene acquisition and innovation in the evolution of Microsporidia intracellular parasites. *Nat. Commun.* 9(1):1709
16. Ding W, Baumdicker F, Neher RA. 2018. panX: pan-genome analysis and exploration. *Nucleic Acids Res.* 46(1):e5
17. Doolittle WF. 2008. Microbial evolution: stalking the wild bacterial species. *Curr. Biol.* 18(13):R565–67
18. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, et al. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309(5733):404–9
19. Etten JV, Bhattacharya D. 2020. Horizontal gene transfer in eukaryotes: not if, but how much? *Trends Genet.* 36(12):915–25
20. Eyres I, Boschetti C, Crisp A, Smith TP, Fontaneto D, et al. 2015. Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. *BMC Biol.* 13:90

8. Foundational work showing expansion in hexose transporters underpins adaptation to low sugar environments in yeast. Should be taught alongside peppered moths and Darwin's finches.

9. HGT, and therefore accessory genome composition, is shaped by how the protein encoded by the transferred gene interacts with the wider proteome network.

21. Fan X, Qiu H, Han W, Wang Y, Xu D, et al. 2020. Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. *Sci. Adv.* 6(18):eaba0111
22. Fickers P, Benetti P-H, Waché Y, Marty A, Mauersberger S, et al. 2005. Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Res.* 5(6–7):527–43
23. Fisher SK, Novak JE, Agranoff BW. 2002. Inositol and higher inositol phosphates in neural tissues: homeostasis, metabolism and functional significance. *J. Neurochem.* 82(4):736–54
24. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, et al. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat. Genet.* 38(8):953–56
25. González-Pech RA, Stephens TG, Chen Y, Mohamed AR, Cheng Y, et al. 2021. Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*. *BMC Biol.* 19(1):73
26. Greiner T, Moroni A, Van Etten JL, Thiel G. 2018. Genes for membrane transport proteins: not so rare in viruses. *Viruses* 10(9):456
27. Hatta R, Ito K, Hosaki Y, Tanaka T, Tanaka A, et al. 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics* 161(1):59–70
28. He C, Rusu AG, Poplawski AM, Irwin JAG, Manners JM. 1998. Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. *Genetics* 150(4):1459–66
29. Heinz E, Hacker C, Dean P, Mifsud J, Goldberg AV, et al. 2014. Plasma membrane-located purine nucleotide transport proteins are key components for host exploitation by microsporidian intracellular parasites. *PLOS Pathog.* 10(12):e1004547
30. Hu J, Chen C, Peever T, Dang H, Lawrence C, Mitchell T. 2012. Genomic characterization of the conditionally dispensable chromosome in *Alternaria arborescens* provides evidence for horizontal gene transfer. *BMC Genom.* 13(1):171
31. Hughes T, Ekman D, Ardawatia H, Elofsson A, Liberles DA. 2007. Evaluating dosage compensation as a cause of duplicate gene retention in *Paramecium tetraurelia*. *Genome Biol.* 8(5):213
32. Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* 16(2):67–79
33. Inderbitzin P, Harkness J, Turgeon BG, Berbee ML. 2005. Lateral transfer of mating system in *Stemphylium*. *PNAS* 102(32):11390–95
34. Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11(2):97–108
35. Irwin NAT, Pittis AA, Richards TA, Keeling PJ. 2022. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat. Microbiol.* 7(2):327–36
36. Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *PNAS* 96(7):3801–6
37. Jeong H, Mason SP, Barabási A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* 411(6833):41–42
38. Jordan P, Choe J-Y, Boles E, Oreb M. 2016. Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. *Sci. Rep.* 6(1):23502
39. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, et al. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414(6862):450–53
40. Keeling PJ, Inagaki Y. 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. *PNAS* 101(43):15380–85
41. Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* 279(1749):5048–57
42. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305(3):567–80
43. Lercher MJ, Pal C. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* 25(3):559–67
44. Lessard EJ, Merico A, Tyrrell T. 2005. Nitrate: phosphate ratios and *Emiliania huxleyi* blooms. *Limnol. Oceanogr.* 50(3):1020–24

24. Evidence of recent HGT between plant-associated fungi that expands the accessory genome and drives pathogenicity.

45. Li G, Ji B, Nielsen J. 2019. The pan-genome of *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 19(7):foz064
46. Li H, Li Y, Sun T, Du W, Li C, et al. 2019. Unveil the transcriptional landscape at the *Cryptococcus*-host axis in mice and nonhuman primates. *PLOS Negl. Trop. Dis.* 13(7):e0007566
47. Lindsay RJ, Pawlowska BJ, Gudelj I. 2019. Privatization of public goods can cause population decline. *Nat. Ecol. Evol.* 3(8):1206–16
48. Liu G, Yong MYJ, Yurieva M, Srinivasan KG, Liu J, et al. 2015. Gene essentiality is a quantitative property linked to cellular evolvability. *Cell* 163(6):1388–99
49. Liu T-B, Kim J-C, Wang Y, Toffaletti DL, Eugenin E, et al. 2013. Brain inositol is a novel stimulator for promoting *Cryptococcus* penetration of the blood-brain barrier. *PLOS Pathog*. 9(4):e1003247
50. Logeman BL, Wood LK, Lee J, Thiele DJ. 2017. Gene duplication and neo-functionalization in the evolutionary and functional divergence of the metazoan copper transporters Ctr1 and Ctr2. *J. Biol. Chem.* 292(27):11531–46
51. **Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464(7287):367–73**
52. MacLean RC, Gudelj I. 2006. Resource competition and social conflict in experimental populations of yeast. *Nature* 441(7092):498–501
53. Major P, Sendra KM, Dean P, Williams TA, Watson AK, et al. 2019. A new family of cell surface located purine transporters in Microsporidia and related fungal endoparasites. *eLife* 8:e47037
54. Marini AM, Soussi-Boudekou S, Vissers S, Andre B. 1997. A family of ammonium transporters in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 17(8):4282–93
55. Martho KFC, de Melo AT, Takahashi JPF, Guerra JM, da Silva Santos DC, et al. 2016. Amino acid permeases and virulence in *Cryptococcus neoformans*. *PLOS ONE* 11(10):e0163919
56. **McCarthy CGP, Fitzpatrick DA. 2019. Pan-genome analyses of model fungal species. *Microb. Genom.* 5(2):e000243**
57. McCarthy CGP, Fitzpatrick DA. 2019. Pangloss: a tool for pan-genome analysis of microbial eukaryotes. *Genes* 10(7):521
58. McInerney JO, McNally A, O'Connell MJ. 2017. Why prokaryotes have pangenomes. *Nat. Microbiol.* 2:17040
59. McKie-Krisberg ZM, Sanders RW. 2014. Phagotrophy by the picoeukaryotic green alga *Micromonas*: implications for Arctic Oceans. *ISME J.* 8(10):1953–61
60. Milner DS, Attah V, Cook E, Maguire F, Savory FR, et al. 2019. Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes. *PNAS* 116(12):5613–22
61. Milner DS, Wideman JG, Stairs CW, Dunn CD, Richards TA. 2021. A functional bacteria-derived restriction modification system in the mitochondrion of a heterotrophic protist. *PLOS Biol.* 19(4):e3001126
62. Monier A, Chambouvet A, Milner DS, Attah V, Terrado R, et al. 2017. Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. *PNAS* 114(36):E7489–98
63. **Monier A, Welsh RM, Gentemann C, Weinstock G, Sodergren E, et al. 2012. Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. *Environ. Microbiol.* 14(1):162–76**
64. Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3(2):e00036-12
65. Navarathna DHMLP, Das A, Morschhäuser J, Nickerson KW, Roberts DD. 2011. Dur3 is the major urea transporter in *Candida albicans* and is co-regulated with the urea amidolyase Dur1,2. *Microbiology* 157(Part 1):270–79
66. Ngamskulrungroj P, Chang Y, Sionov E, Kwon-Chung KJ. 2012. The primary target organ of *Cryptococcus gattii* is different from that of *Cryptococcus neoformans* in a murine model. *mBio*. 3(3):e00103-12
67. Norsigian CJ, Danhof HA, Brand CK, Midani FS, Broddrick JT, et al. 2022. Systems biology approach to functionally assess the *Clostridioides difficile* pangenome reveals genetic diversity with discriminatory power. *PNAS* 119(18):e2119396119
68. Ohno S. 1970. *Evolution by Gene Duplication*. Berlin: Springer

51. Key demonstration that fungi can place major units of their accessory genomes in accessory chromosomes, and that these are mobile between evolutionary lineages.

56. Systematic comparisons of eukaryotic pangenomes reveal that gene duplications are an important factor.

63. Virus genomes carry variant transporter genes, providing a key hint that viruses are acting to drive gene transfer and therefore act to further provision eukaryotic accessory genomes.

69. Oliveira PH, Touchon M, Rocha EPC. 2016. Regulation of genetic flux between bacteria by restriction-modification systems. *PNAS* 113(20):5658–63
70. Osuna-Cruz CM, Bilcke G, Vancaester E, De Decker S, Bones AM, et al. 2020. The *Seminavis robusta* genome provides insights into the evolutionary adaptations of benthic diatoms. *Nat. Commun.* 11(1):3320
71. Otterstedt K, Larsson C, Bill RM, Ståhlberg A, Boles E, et al. 2004. Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. *EMBO Rep.* 5(5):532–37
72. Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, et al. 2021. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30(1):187–200
73. Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424(6945):194–97
74. Peck KM, Lauring AS. 2018. Complexities of viral mutation rates. *Virol. J.* 92(14):e01031-17
75. Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* 16:5
76. Poulsen BE, Yang R, Clatworthy AE, White T, Osmulski SJ, et al. 2019. Defining the core essential genome of *Pseudomonas aeruginosa*. *PNAS* 116(20):10072–80
77. Rajasingham R, Smith RM, Park BJ, Jarvis JN, Govender NP, et al. 2017. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *Lancet Infect Dis.* 17(8):873–81
78. Rastogi A, Vieira FRJ, Deton-Cabanillas A-F, Veluchamy A, Cantrel C, et al. 2020. A genomics approach reveals the global genetic polymorphism, structure, and functional diversity of ten accessions of the marine model diatom *Phaeodactylum tricornutum*. *ISME J.* 14(2):347–63
79. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499(7457):209–13
80. Rees DC, Johnson E, Lewinson O. 2009. ABC transporters: the power to change. *Nat. Rev. Mol. Cell. Biol.* 10(3):218–27
81. Reis AC, Cunha MV. 2021. The open pan-genome architecture and virulence landscape of *Mycobacterium bovis*. *Microb. Genom.* 7(10):000664
82. Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ. 2009. Biogeography of the *Sulfolobus islandicus* pan-genome. *PNAS* 106(21):8605–10
83. Richards TA, Leonard G, Soanes DM, Talbot NJ. 2011. Gene transfer into the fungi. *Fungal Biol. Rev.* 25(2):98–110
84. Richards TA, Talbot NJ. 2013. Horizontal gene transfer in osmotrophs: playing with public goods. *Nat. Rev. Microbiol.* 11(10):720–27
85. Richards TA, Talbot NJ. 2018. Osmotrophy. *Curr. Biol.* 28(20):R1179–80
86. Rokitta SD, Von Dassow P, Rost B, John U. 2014. *Emiliania huxleyi* endures N-limitation with an efficient metabolic budgeting and effective ATP synthesis. *BMC Genom.* 15(1):1051
87. Rolland F, Winderickx J, Thevelein JM. 2002. Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Res.* 2(2):183–201
88. Rousset F, Cabezas-Caballero J, Piastra-Facon F, Fernández-Rodríguez J, Clermont O, et al. 2021. The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat. Microbiol.* 6(3):301–12
89. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, et al. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537(7622):689–93
90. Ryan CJ, Krogan NJ, Cunningham P, Cagney G. 2013. All or nothing: protein complexes flip essentiality between distantly related eukaryotes. *Genome Biol. Evol.* 5(6):1049–59
91. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, et al. 2021. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res.* 49(D1):D461–67
92. Savory FR, Milner DS, Miles DC, Richards TA. 2018. Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. *Mol. Biol. Evol.* 35(8):1887–900
93. Seabolt MH, Roellig DM, Konstantinidis KT. 2022. Genomic comparisons confirm *Giardia duodenalis* sub-assemblage AII as a unique species. *Front. Cell. Infect. Microbiol.* 12:1010244
94. Shapiro BJ. 2017. The population genetics of pangenomes. *Nat. Microbiol.* 2(12):1574

94. Constructive discussion about whether accessory genomes are neutral or adaptive.

95. Sharma S, Markham PF, Browning GF. 2014. Genes found essential in other mycoplasmas are dispensable in *Mycoplasma bovis*. *PLOS ONE* 9(6):e97100
96. Slot JC, Gluck-Thaler E. 2019. Metabolic gene clusters, fungal diversity, and the generation of accessory functions. *Curr. Opin. Genet. Dev.* 58–59:17–24
97. Szabová J, Ruzicka P, Verner Z, Hampl V, Lukes J. 2011. Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: coexistence of mutually exclusive transcripts predates functional rescue. *Mol. Biol. Evol.* 28(8):2371–78
98. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome." *PNAS* 102(39):13950–55
99. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, et al. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21(1):180
100. Vakirlis N, Hebert AS, Opulente DA, Achaz G, Hittinger CT, et al. 2018. A molecular portrait of *de novo* genes in yeasts. *Mol. Biol. Evol.* 35(3):631–45
101. Verduyn C, Zomerdijk TPL, van Dijken JP, Scheffers WA. 1984. Continuous measurement of ethanol production by aerobic yeast suspensions with an enzyme electrode. *Appl. Microbiol. Biotechnol.* 19(3):181–85
102. Vylkova S, Carman AJ, Danhof HA, Collette JR, Zhou H, Lorenz MC. 2011. The fungal pathogen *Candida albicans* autoinduces hyphal morphogenesis by raising extracellular pH. *mBio* 2(3):e00055-11
103. Weir W, Capewell P, Foth B, Clucas C, Pountain A, et al. 2016. Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *eLife* 5:e11473
104. Whelan FJ, Hall RJ, McInerney JO. 2021. Evidence for selection in the abundant accessory gene content of a prokaryote pangenome. *Mol. Biol. Evol.* 38(9):3697–708
105. Whittaker RH. 1969. New concepts of kingdoms or organisms: Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science* 163(3863):150–60
106. Williams TA, Nakjang S, Campbell SE, Freeman MA, Eydal M, et al. 2016. A recent whole-genome duplication divides populations of a globally distributed microsporidian. *Mol. Biol. Evol.* 33(8):2002–15
107. Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324(5924):268–72
108. Xue C, Liu T, Chen L, Li W, Liu I, et al. 2010. Role of an expanded inositol transporter repertoire in *Cryptococcus neoformans* sexual reproduction and virulence. *mBio* 1(1):e00084-10
109. Zhang Z, Ren Q. 2020. Why are essential genes essential? The essentiality of *Saccharomyces* genes. *Microb. Cell* 2(8):280–87
110. Zhu Y, Du P, Nakhleh L. 2012. Gene duplicability-connectivity-complexity across organisms and a neutral evolutionary explanation. *PLOS ONE* 7(9):e44491

108. Important example of how transporter evolution can shape the evolution of a parasitic fungus.

Annual Review of Microbiology
Volume 77, 2023

Contents

Errata

An online log of corrections to *Annual Review of Microbiology* articles may be found at http://www.annualreviews.org/errata/micro