Viral-eukaryotic gene exchange drives infection mode and cellular evolution

Nicholas Irwin (nickatirwin@gmail.com)
University of Oxford https://orcid.org/0000-0002-2904-8214
Alexandros Pittis
European Molecular Laboratory (EMBL) https://orcid.org/0000-0003-4116-9972
Thomas Richards
University of Oxford https://orcid.org/0000-0002-9692-0973
Patrick Keeling
University of British Columbia https://orcid.org/0000-0002-7644-0745

Article

Keywords: Horizontal Gene Transfer, Infection Strategies, Viral Host-manipulation Strategies, Viral Glycosyltransferases

Posted Date: April 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-380297/v1

Version of Record: A version of this preprint was published at Nature Microbiology on December 31st, 2021. See the published version at https://doi.org/10.1038/s41564-021-01026-3.

Abstract

Gene exchange between viruses and their hosts acts as a key facilitator of horizontal gene transfer and is thought to be a major driver of evolutionary change 1–3. Our understanding of this process comes primarily from bacteria and phage co-evolution4, but the mode and functional significance of gene transfers between eukaryotes and their viruses remains more anecdotal. Here we show that viral-eukaryotic gene exchange can define infection strategies and has recurrently influenced eukaryotic evolution. Using a systematic, phylogenetically-informed approach, we characterized viral-eukaryotic gene exchange across diverse taxa, identifying thousands of transfers, and revealing their frequency, taxonomic distribution, and projected functions, across the eukaryotic tree of life. Eukaryote-derived viral genes revealed common viral host-manipulation strategies, including the key cellular pathways and compartments targeted during infection, identifying potential targets for broad-spectrum host-targeted antiviral therapeutics. Furthermore, viral-derived eukaryotic genes exposed a recurring role for viral glycosyltransferases in the diversification of eukaryotic morphology, as viral-derived genes have impacted the evolution of structures as diverse as algal cell walls, trypanosome mitochondria, and animal tissues. These findings illuminate the nature of viral-eukaryotic gene exchange and its impact on the biology of viruses and their eukaryotic hosts, providing novel perspectives for understanding viral infection mechanisms and revealing the influence of viruses on eukaryotic evolution.

Main Text

The exchange of genes between viruses and eukaryotes through horizontal gene transfer (HGT) is a key evolutionary driver capable of facilitating host manipulation and viral resistance [2,3,5]. Host-derived genes are known to be employed by viruses for replication and cellular control [5,6]. This is observed across a diversity of viral lineages which encode cellular-derived informational genes like tRNA synthetases and polymerases, as well as operational genes, such as immune effectors and metabolic enzymes [6–15]. These genes counter host immunity, hijack cellular machinery, and circumvent nutritional bottlenecks, making them key resources for adaptation [5,16].

Conversely, viral-derived genes in eukaryotic genomes are frequently perceived as inconsequential remnants of viral interactions, or even discarded as contamination in genomic analyses. However, these genes can be co-opted and supplement or supplant existing cellular components and functions. For example, core proteins such as histones and E2F transcription factors have been replaced by viral proteins in dinoflagellates and fungi, respectively [17,18], while viral structural proteins, fusogens, and proviruses are utilized for communication, cellular fusion, and antiviral defense, in mammals and other eukaryotes [2,3,19–22]. The co-option of such viral proteins has been found to coincide with cellular innovation and the radiation of major eukaryotic lineages where these genes serve key functions [23,24].

Accordingly, these transfers have important evolutionary, ecological, and health implications, but we nonetheless lack a general understanding of the mode, tempo, and functional patterns of viral-eukaryotic gene exchange due to a lack of systematic analyses across diverse taxa. To reconcile this, we

comprehensively characterized viral-eukaryotic gene transfer in 201 eukaryotic and 108,842 viral taxa by developing a phylogenetic pipeline capable of screening thousands of evolutionary trees for HGT-indicative topologies while accounting for phylogenetic statistics and contamination (Extended Data Fig. 1, 2). These analyses identify 1,333 candidate virus-to-eukaryote and 4,807 eukaryote-to-virus transfers, along with 600 transfers with unknown directionality, affecting 2,841 distinct protein families (Fig. 1a, Supplementary Table 1). Phylogenetically ambiguous or long branching HGTs were considered weakly supported and were excluded in downstream analyses (Fig. 1a, Supplementary Table 1), which, along with limitations in taxon sampling, make these figures a conservative estimate of HGT events.

The resulting HGTs revealed trends regarding the nature of viral-eukaryotic gene exchange. Transfers from eukaryotes to viruses were observed approximately twice as frequently as transfers in the reverse direction (Fig. 1a, b). This imbalance is explained by the higher number of viral recipients compared to donors per eukaryotic taxa (Fig. 1c) and the greater number of genes transferred to each viral recipient relative those received per viral donor (Fig. 1d, e). These data also demonstrate a correlation between gene acquisition and donation ($r_{Pearson}$ = 0.50, $p < 1x10^{-18}$, Fig. 1b), suggesting that viral-eukaryotic gene transfer is reciprocal, likely instigated through specific host-virus interactions as opposed to non-specific (e.g., environmental) uptake, and is biased towards viral acquisition. This may reflect the expanded repertoire of eukaryotic genes compared to their viral counterparts, which would generate greater opportunity for viral gene acquisition during host-interaction.

Identifying the taxonomy of donors and recipients revealed the propensity of certain lineages to participate in HGT. Nucleocytoplasmic large DNA viruses (NCLDV or Nucleocytoviricota, including Phycodnaviridae, Mimiviridae, Iridoviridae, Pithoviridae, Asfarviridae, and Poxviridae) contributed to the majority of genetic exchanges (78%), although lineage-specific associations, such as the acquisition of animal genes by herpes- and poxviruses, were also noted, and highlight the variable host breadth of viral groups (Fig. 1f, g, Extended Data Figure 1c). Amongst eukaryotes, gene exchange was more prevalent in unicellular compared to multicellular organisms, and particularly abundant in unicellular opisthokonts (the protist relatives of animals and fungi), the diverse protist clade known as SAR (Stramenopila, Alveolata, and Rhizaria), and other ecologically important algal groups such as chlorophytes and haptophytes. This included numerous HGTs coinciding with the diversification of SAR and the largest influx of viral genes was detected around the origin of the dinoflagellates (Fig. 1f, g). Elevated gene transfer amongst unicellular eukaryotes may result from more frequent encounters with NCLDV, which are hyper-diverse and abundant in aquatic environments [10], as well as a lack of germline segregation, which likely contributes to the reduced frequency of HGTs observed in animals and plants (Fig. 1g) [25]. However, gene exchange was more common amongst invertebrates compared to vertebrate animals, and our methodology likely under-represents viral gene transfer in animals due to the under-estimation of retroviral acquisitions, which are commonly observed throughout animal lineages but whose detection is limited by the lack of host-free retroviral genome assemblies [26].

We also noted eukaryotic species harboring particularly large numbers of viral genes (Fig. 1e, g). These included species previously described to contain substantial viral genomic insertions from phycodnaviruses (*Ectocarpus siliculosus* and *Tetrabaena socialis*), phycodnaviruses and asfarviruses (*Hyphochytrium catenoides*), or multiple poorly classified viruses (*Acanthamoeba castellanii*), indicating single or few sources (Fig. 1g, Supplementary Table 1) [27–30]. Other species also exhibited elevated numbers of viral genes derived from multiple NCLDV sources (Fig. 1e, g). Whether these large multigene acquisitions retain functional roles, such as in anti-viral virophage production [31], or reflect remnants of past infections, is unclear. However, large insertions were not detected at ancestral nodes (Fig. 1g), suggesting that viral integrations are recurrent, affect diverse eukaryotic lineages, and are generally only transiently retained, but provide an opportunity for the longer-term retention and co-option of individual viral genes given adaptive significance and selection for fixation.

To investigate the functional relevance of these HGTs, we examined the transfer direction and functional enrichments of exchanged protein families. Of the 1,859 families exhibiting HGT with known directionality, the majority (93%) underwent unidirectional transfer (Fig. 2a). Dividing this dataset by direction, genes involved in viral acquisitions were generally transferred unidirectionally (92%), whereas a larger proportion of families undergoing virus-to-eukaryote transfer participated in bidirectional exchange (29%) (Fig. 2a), suggesting that some of these exchanges may involve transduction (cell-virus-cell HGT). By moving across the phylogenies of all families exhibiting eukaryotic acquisitions, from viral donors towards the root, we estimated that 30.5% (n = 259) of viral genes acquired by eukaryotes were originally eukaryotic, whereas fewer (8.2%, n = 70) originated in prokaryotes (Extended Data Fig. 3, Supplementary Table 1). The remainder had unclear origins (24.2%, n = 205) or were not attributable to a cellular lineage (37.1%, n = 315), suggesting that these genes are either viral innovations or ancient viral acquisitions sharing deep cellular homology undetectable in our dataset (Extended Data Fig. 3a). These data demonstrate that over evolutionary time, viruses have a capacity to mediate intra-eukaryotic and inter-domain HGT through transduction. This suggests that viruses act as a gene conduit between eukaryotic lineages, as in prokaryotes, where viral transduction is key in ecological adaptation and genome evolution [1,4,32].

Direction of transfer was also associated with distinct functional biases. Eukaryote-to-virus transfers were enriched in functions associated with cellular activity and house-keeping, such as metabolic proteins, E3-ligases, and tRNA synthetases (Fig. 2b, Supplementary Table 1, Supplementary Table 3). The enrichment of metabolic proteins highlights the role of cellular-derived genes in reprogramming host metabolism during infection, which appears to be achieved through both *de novo* metabolite synthesis pathways and uptake (e.g., metabolic enzymes and/or nutrient transporters), as well as cellular recycling via proteolysis (e.g., proteasomal degradation and autophagy) (Fig. 2a, b, Supplementary Table 1, Supplementary Table 3). Additionally, signalling and stress response proteins are frequently acquired and likely also contribute to regulating host physiology, gene expression, immune responses, and viral processing. The functions of viral-derived genes in eukaryotes are less obvious and have fewer functional associations, but are strongly enriched for proteins functioning in glycosylation and, to a lesser extent,

nuclear proteins (Fig. 2a, c, Supplementary Table 1, Supplementary Table 3). Bidirectionally transferred genes are also enriched in metabolic processes, protein modification, and stress response proteins, which represent a subset of functions most often acquired by viruses (Fig. 2d, Supplementary Table 1, Supplementary Table 3). These data show that eukaryote-to-virus and virus-to-eukaryote HGTs both involve functional tendencies which are not equivalent, but reflect the different adaptive contexts of viruses and eukaryotes.

To understand how these genes are used in viral and eukaryotic systems, we first examined the subcellular targets of eukaryote-derived viral proteins to understand where the proteins may operate in host cells. Cellular localizations were predicted using a neural network-based approach (DeepLoc) [33], revealing that most eukaryote-to-virus HGTs likely function in the cytoplasm (*n* = 909), nucleus (*n* = 482), mitochondrion (*n* = 284), and extracellular space (*n* = 214) (Fig. 3a, Supplementary Table 1). However, relative to all eukaryotic protein families, viral-acquisitions were enriched in cytoplasmic, endoplasmic reticulum (ER), extracellular, and peroxisomal proteins, the last of which suggests functions involving lipid catabolism and oxidation (Fig 3b). Moreover, predicted localizations were generally equivalent between donor and recipient proteins, with variation likely resulting from prediction inconsistencies and viral sequence divergence (Fig 3c, 71% consistent), indicating that genes acquired by viruses tend to function in their original subcellular contexts.

To corroborate the predicted localizations and better understand the impact of these genes on cellular compartments, we conducted localization-based functional enrichments revealing additional cellular processes targeted during infection. Cytoplasmic proteins were largely involved in translation, metabolism, proteolysis, and signaling, whereas nuclear proteins mainly functioned in DNA processing, chromatin organization, cell cycle regulation, and protein modification (Fig 3d, e, Supplementary Table 1, Supplementary Table 4). Endoplasmic reticulum proteins were predominantly associated with lipid metabolism and membrane remodeling (Fig. 3f, Supplementary Table 4). Proteins such as sphingolipid synthesis enzymes contribute to the localization bias, since many function in the ER, were frequently transferred (Supplementary Table 1), and are known to be used by diverse viruses for cellular regulation [16,34,35]. Additionally, ER remodeling is important for generating membrane-enclosed viral factories [36]. Extracellular proteins acquired by viruses were enriched for functions including carbohydrate metabolism, protein maturation, and proteolysis, implying a tendency for cell-surface modulation (Fig. 3g, Supplementary Table 4). These results highlight the key cellular systems targeted by eukaryote-derived genes during infection. However, these processes are also known to be manipulated by viruses that lack eukaryotic genes (e.g., many non-NCLDV viruses), which instead often rely on small, functionally cryptic effectors. This suggests that cellular manipulation strategies are ubiquitous, but that the mode through which modification is accomplished may depend on viral coding capacity (e.g., reduced coding limitations in the NCLDV could permit the use of more and larger eukaryotic genes).

Lastly, to gain insights into the role viral genes play in eukaryotic systems, we inspected the distributions and functions of viral-derived glycosyltransferases, which were strongly enriched in virus-to-eukaryote HGTs (Fig. 2c). We identified 63 instances of eukaryotes acquiring viral glycosyltransferases, of which 13

mapped to ancestral nodes, implying functional relevance under long term selection (Supplementary Table 5). Plotting transfer events and annotations over a eukaryotic phylogeny revealed the functional diversity and recurrent acquisitions of these enzymes across eukaryotic lineages (Fig. 4a, Extended Data Fig. 4). These HGTs were often correlated with morphological and structural synapomorphies including algal cell wall elaboration (e.g., lipopolysaccharide (LPS) and cellulose synthesis enzymes) [37], long-chain polyamine-containing scale formation in haptophytes (spermidine synthase) [38], cellular aggregation in the opisthokonts and dictyostelid slime molds (hyaluronan synthase and GlcNAc transferase), and mitochondrial divergence in the kinetoplastids (fucosyltransferase), a group primarily comprised of animal parasites such as trypanosomes (Fig 4a). Experimental data supported a number of these correlations, including the unusual identification of LPS in the cell walls of *Chlorella* [39], the importance of hyaluronan in vertebrate tissues [40], and the role of the dictyostelid N-acetylglucosamine transferase, Gnt2, in calcium-independent cellular aggregation [41,42], demonstrating that virally sourced genes are co-opted during the evolution of cellular traits (Fig. 4a). We further examined two glycosyltransferase acquisitions in kinetoplastids, hypothesizing that, given the correlation between the HGT acquisitions and the origin of the highly derived kinetoplastid mitochondria (called kinetoplasts), they should function in that compartment. Phylogenetic analyses revealed that both genes were derived from the NCLDV, highlighted the prokaryotic origin of the fucosyltransferase (COG000231), and confirmed that both genes were conserved throughout kinetoplastids (Fig. 4b, c). Moreover, both proteins localized to the kinetoplast in *Trypanosoma brucei* (identifiable as a non-nuclear DNA-stained foci) both when tagged with mNeonGreen (Fig. 4d) and by organellar proteomics (Fig. 4e). A recent report also suggests an essential role for the fucosyltransferase in kinetoplast function in *T. brucei* [43], altogether indicating that these viral-derived glycosyltransferases were co-opted for use in the kinetoplast at the same time as it underwent massive evolutionary change. These data, along with the tendency for viruses to modify cell surfaces, suggest that viral-derived genes may have played various roles in the evolution of cellular morphology across the eukaryotic tree of life, possibly affecting the diversification of eukaryotic forms.

Horizontal gene transfer between viruses and eukaryotes has been observed and assumed to impact genome evolution in both participants, but until now we lacked the systematic characterization of these gene exchanges necessary to generalize their mode and functional significance in both viral and eukaryotic contexts. As with all computational surveys, our dataset is limited by specificity and sensitivity, but nonetheless it provides an extensive resource from which phylogenetic patterns can be observed and their genomic and functional importance may be predicted. From a viral perspective, the apparent ubiquity of host-manipulation strategies suggests that the cellular processes outlined above may represent targets for the development of broad-spectrum, host-targeting, antiviral therapeutics. Indeed, many important emerging human pathogens, such as Ebola virus, Zika virus, and coronaviruses, depend on the manipulation of the same cellular processes outlined above, such as autophagy, proteolysis, ER modification, and sphingolipid metabolism [35,44–46]. Functional investigations of eukaryote-derived viral genes, particularly using heterologous expression [7], may also provide insights into how viruses manipulate these cellular pathways while circumventing the need for tractable host-virus model systems. From a eukaryotic perspective, our analyses suggest that viruses can not only mediate intra-eukaryotic

gene exchange but that the evolution of cellular morphology and structure has been influenced by viral genes, particularly glycosyltransferases. These have recurrently impacted transitions as fundamental as the evolution of tissues or divergent mitochondria, reminiscent of how retroviral fusogens have repeatedly driven placental evolution in mammals and lizards [22]. Our survey also identifies protein candidates for which experimental characterizations would help reveal the full impact of these genes on cellular systems and their role in driving the evolution of eukaryotic complexity.

Materials And Methods

Dataset assembly

To systematically identify instances of viral-eukaryotic gene exchange, groups of homologous eukaryotic, viral, and prokaryotic proteins were clustered into protein families and phylogenetic analyses were performed (Extended Data Fig. 1a). The eukaryotic dataset was generated from 196 genome-predicted eukaryotic proteomes, primarily from UniProt (release 2018_11, see **Data and Material Availability**) [50], that represented diverse species from all available major eukaryotic lineages, were individually clustered at 99% percent-identity with CD-Hit v4.8.1 [51], and combined. The eukaryotic dataset was further supplemented with five high quality transcriptomes to fill taxonomic gaps in lineages with poor genomic sampling (four dinoflagellates and a cercozoan) (Extended Data Fig. 1a, b) [52,53]. Viral proteins predicted from viral genomes were obtained from UniProt and filtered to exclude those derived from human immunodeficiency virus-1, which were over-represented, and additional viral proteins were acquired from low-contamination nucleocytoplasmic large DNA virus (NCLDV) metagenomes from diverse environments (Extended Data Fig. 1a) [10]. Viral taxonomic annotations were assigned to metagenomes based on previously conducted phylogenomic analyses [10].

Eukaryotic and viral proteins were then clustered into protein families using a similarity-based approach and the Markov clustering (MCL) algorithm (inflation = 2) after comparing sequences to one another using Diamond v2.0.2 BLASTp (sensitive mode, e-value < 10^{-5}, query coverage > 50%) [54–56]. Protein families containing both viral and eukaryotic representatives were retained, aligned with MAFFT v.7.397 [57], and used to generate profile hidden Markov models (HMMs) which were used to search 9,035 prokaryotic proteomes from UniProt with HMMER v.3.2.1 (e-value < 10^{-5}, incE < 10^{-5}, domE < 10^{-5}) [58] (Extended Data Fig. 1b). Due to the large number of prokaryotic sequences, the resulting hits were reduced by taking the most significant hit (based on e-value) per genus or per strain, to a maximum of 150 sequences (on average 39% of the total hits). Sequences assigned to viral-eukaryotic protein families were then combined with the prokaryotic proteins and re-clustered, as above (Extended Data Fig. 1a).

Phylogenetic analysis

Phylogenetic trees were generated from clustered protein families to infer the evolutionary relationships between viral and eukaryotic homologues. Protein families were filtered to retain only those with viruses and eukaryotes, aligned with MAFFT, trimmed using a gap-threshold of 30% in trimAl v1.2, and sequences

with less than 50 amino acid positions were removed [59]. Maximum likelihood phylogenies were conducted in IQ-Tree v1.6 using the LG+F+R5 substitution model, and statistical support was calculated using SH-aLRT (Shimodaira-Hasegawa approximate likelihood ratio test, *n* = 1,000), which was chosen due to its speed, insensitivity to model violations and taxon sampling, and its comparable conservativeness to standard bootstrapping [48,60,61]. Phylogenies for large protein families with over 1,500 sequences (*n* = 103) were generated using the fast search mode in IQ-Tree. Phylogenetic rooting was done using minimal ancestral deviation (MAD), which is a rooting method that is more robust to heterotachy then midpoint rooting [62].

For individual phylogenies of particular interest, such as those shown in Fig. 4, Extended Data Fig. 3, and Extended Data Fig. 4, analyses were repeated as above but after alignment with the more accurate L-INS-i algorithm in MAFFT and limited curation (e.g., the removal of long-branching taxa as defined below, see **Horizontal gene transfer detection**). Additionally, substitution models were selected using ModelFinder in IQ-Tree [48,49] and phylogenies were visualized and annotated using iTOL [63]. Notably, the topologies of these trees were consistent with their initial iterations and ModelFinder consistently selected the LG substitution model similar to that used in the other phylogenies, corroborating the use of the aforementioned methods (see Extended Data Fig. 4).

Horizontal gene transfer detection

After generating phylogenetic trees for each protein family, we developed an automated pipeline using the python package, ETE 3 [64], to identify HGT-indicative topologies. Specifically, we aimed to identify eukaryotic species nested within viral clades (viral-to-eukaryote HGT) or viral taxa within eukaryotic clades (eukaryote-to-virus HGT) (Extended Data Fig. 1a). To this end, phylogenies were initially processed to account for statistical support and directionality (i.e., rooting), and to assign taxonomic annotations. Firstly, phylogenetic nodes with SH-aLRT values below 0.8 were collapsed, a threshold with a similar false positive rate to standard bootstrapping that balances specificity and sensitivity [60]. Collapsed phylogenies were then rooted using MAD (96.4%) [62] or midpoint rooting and taxa were annotated as eukaryotic, viral, or prokaryotic based on National Centre for Biotechnology Information (NCBI) taxonomy (Extended Data Fig. 2a).

Following tree processing and annotation, but before identifying HGT events, the phylogenies were analyzed to assess rooting ambiguity. In particular, we checked whether viral and eukaryotic sequences could be separated into two monophyletic groups using alternative root placements. In this case rooting becomes unclear unless the phylogeny is strongly biased toward viral or eukaryotic species representation (e.g., it is unlikely that a gene conserved throughout a eukaryotic supergroup was derived from a single virus). To evaluate this, if a phylogeny could be split into two discrete taxonomic clades, the ratio of eukaryotic to viral species was determined. If the ratio was heavily skewed towards eukaryotes or viruses (eukaryote:viral species ratio > 49 or < 0.15, reflecting the top and bottom 20% of all protein families), the tree was rooted normally. Otherwise, the topology would be classified as an HGT with unknown directionality. Lastly, single prokaryotic taxa and HGTs between prokaryotes and viruses or

eukaryotes (identified as below) were removed for simplicity but did not increase the false positive rate amongst viral-eukaryotic HGTs.

After processing, phylogenies were screened for HGT topologies. To achieve this, viral and eukaryotic clades were identified and the taxonomy of their sister group (i.e., the most closely related phylogenetic group) and 'cousin' group (i.e., the second most closely related phylogenetic group) were determined. A eukaryote-to-virus HGT topology was defined as a viral clade with a eukaryotic sister and cousin whereas a virus-to-eukaryote HGT required a eukaryotic clade with a viral sister and cousin (Extended Data Fig. 2a). Initially, viral and eukaryotic clades were identified and the taxonomy of their sister and cousin groups were assessed. To classify the taxonomy of these groups, the number of viral, eukaryotic, and prokaryotic sequences in each group was counted. Sister and cousin groups were then classified as viral, eukaryotic, or prokaryotic if the taxonomies were consistent across the members of the group. If the taxonomies of a group were mixed (e.g., if both viral and eukaryotic sequences were present), but viral or eukaryotic taxa dominated at least 80% of the sequences, the group was described as 'probably' viral or eukaryotic, or else the group received an ambiguous designation. In the event of a polytomy, multiple sister and cousin groups could be present. To account for this, the taxonomy of the polytomy-wide group would be summarized by determining the taxonomy of each group within the polytomy (as above). If all candidate sisters or cousins within the polytomy were classified consistently, the group would be identified as viral, eukaryotic, or prokaryotic. Likewise, if a majority (>66%) of the groups were consistently classified, the sister or cousin would be denoted as 'probably' viral or eukaryotic, otherwise it would be labeled as ambiguous. After classifying both sister and cousin groups, if the topology was consistent with one of the aforementioned scenarios, an HGT event would be noted. Each phylogeny was screened for eukaryote-to-virus and virus-to-eukaryote HGTs three times iteratively, given that once a viral or eukaryotic clade had been classified as an HGT, it would be interpreted as eukaryotic or viral, respectively, in subsequent iterations. Finally, after three cycles of HGT identification, if there were remaining viral and eukaryotic clades sister to one another with ambiguous or prokaryotic cousins, they were labeled as HGTs with unknown transfer directionality.

Once an HGT was identified characteristics including the recipient, donor, phylogenetic statistics, and topology notes were recorded (Supplementary Table 1). Recipient and donor taxa were assessed by determining the last common ancestor of the recipient and donor (sister), respectively, based on NCBI taxonomy. Moreover, node statistical support values were recorded along with the branch length of the recipient. If the branch length of the recipient or donor represented an extreme outlier (defined as the median branch length plus three times the interquartile range), the HGT was highlighted as a potential long branch attraction (LBA) artifact. Additionally, if the donor only had a 'probable' taxonomic classification, ambiguity would also be noted. In both of these cases, HGTs were labeled as weakly supported and excluded in downstream analyses. Lastly, the approximate origin of viral-derived eukaryotic genes was determined by moving up through phylogenetic nodes from the donor clade towards the root until a cellular lineage was reached, if possible (Extended Data Fig. 3a).

Contamination scoring

After identifying the HGTs, individual transfer events were assessed for possible alternative sources of phylogenetic incongruence, specifically contamination. This is important given that eukaryotic and viral genes can be artifactually present in viral and eukaryotic genomes, respectively, which may give the impression that HGT has occurred. To address this, only complete viral genomes and metagenomes with low contamination scores were included in the analysis [10] and individual viral-derived eukaryotic genes were assessed based on a series of criteria and a contamination scoring scheme (Extended Data Fig. 2b-e). Contamination was assessed based on two main attributes: 1) the presence of related taxa in the HGT, and 2) the characteristics of the genomic contig upon which the gene was encoded. Firstly, the taxonomic composition of the HGT-recipients was assessed based on the assumption that the same contamination is unlikely to occur in multiple independently sampled genomic datasets, particularly if the species from which they are derived are closely related. Therefore, points were given if the HGT-recipients included multiple members of the same (+3) or different (+1) phyla as the species encoding the gene of interest (Extended Data Fig. 2b).

Secondly, the characteristics of the genomic contigs encoding each viral-derived gene were inspected based upon the notion that they should share attributes with the host genome, such as consistent GC-content, reasonable contig size, and that the gene should be flanked by eukaryotic regions. To this end, contigs were identified by mapping proteins to the genome using tBLASTn (e-value < 10^{-5}) and points were given if the contig was within one standard deviation of the median genomic GC-content (+1) and if the contig was a reasonable size (greater than half of the L50) (+1). Lastly, the genomic context was inspected by extracting DNA regions (5 kbp) upstream (-2.5 kbp) and downstream (+2.5 kbp) of each gene. Extracted regions were then taxonomically classified by comparing them to the SWISS-PROT database (release 2019_11) using BLASTx [65] and assessing the taxonomy of the resulting hits. A maximum of 20 hits (e-value < 10^{-3}) were evaluated and normalized scores for viral, eukaryotic, and prokaryotic classifications were calculated to account for database bias. These scores were calculated for each taxonomic group as:

$$score_t = -\log_{10}(t_{prop}) \times S_t \times n_t$$

where t_{prop} is the proportion of a taxonomic group in the database, and n_t and S_t are the number and median bit-score of hits to that taxonomic group, respectively. The taxonomy was assigned to the group with the maximum score and points were given for eukaryotic classifications (+1 per region) and subtracted for prokaryotic ones (-1 per region), which are more indicative of contamination. Viral and uncertain classifications were neutral (+0) to account for the possibility of large viral insertions.

After assigning contamination scores to each putative eukaryotic sequence involved in a virus-to-eukaryote HGT or an HGT with unknown directionality, sequences with scores less than two were excluded and HGTs and recipient taxonomies were reassessed (Extended Data Fig. 2c, d). Due to the strict criteria applied during filtering and HGT identification, false positive rates should be low.

Functional analyses

To examine HGT function, eukaryotic and viral proteins were annotated with eggNOG, Pfam and PANTHER (Protein analysis through evolutionary relationships) using a combination of eggNOG-Mapper v2 and InterProScan v.5.48 with the default parameters [66–70]. For clarity, the resulting gene ontology (GO) terms were simplified by mapping the terms to the yeast GO-slim subset using Map2Slim (see https://github.com/owlcollab/owltools/wiki/Map2Slim). Protein families were given functional annotations based on a majority rule (Supplementary Table 1) and labeled with gene ontology (GO) terms if a given term was assigned to at least 20% of annotated proteins within a family. To conduct GO-enrichment analyses, protein families exhibiting HGT were compared against a eukaryotic background comprising all eukaryotic protein families containing either a virus or at least ten eukaryotic species. The frequencies of individual GO-terms in the HGT families were compared to the eukaryotic background using permutation tests which involved randomly sampling equally sized sets of annotated protein families without replacement ($n = 10^7$), with the null hypothesis being that GO-terms associated with the HGTs reflect a random sampling of eukaryotic protein families, as has been done previously [71]. Significantly enriched GO terms ($p < 0.01$) were summarized and visualized using REVIGO [47].

To investigate the predicted subcellular localizations of eukaryote-derived viral genes, all eukaryotic proteins were annotated using DeepLoc v1.0 and the BLOSSUM62 matrix [33]. Localization predictions with likelihoods less than 0.5 were re-classified as unknown and cellular targets were assigned to individual eukaryote-to-virus HGTs based on the majority localization of the donor (i.e., eukaryotic) sequences. Enrichments were assessed by comparing the frequency of individual localizations in the HGTs to an equally sized random sampling of annotated eukaryotic proteins ($p < 0.05$, $n = 10^6$). The null hypothesis was that viruses randomly acquire eukaryotic genes irrespective of their predicted subcellular localizations.

Data and code availability

All data, including proteomes, protein families, annotations, alignments, phylogenies, and Python scripts for phylogenetic analyses, contamination scoring, and functional enrichments are available from Dryad (https://datadryad.org/stash/share/jT_8Q2Yh3197gDLiAFh4JBiTs0-WbKYg_DYD-3Zqml4) (Reviewer accessible link).

Declarations

Acknowledgements

We thank Richard Wheeler for providing fluorescent micrographs of *Trypanosoma brucei*, as part of TrypTagDB. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2014-03994) and from the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9201) to P.J.K. N.A.T.I. was supported by an NSERC Canadian Graduate Scholarship and a Junior Research Fellowship from Merton College, Oxford. A.A.P. was supported by

European Molecular Biology Organization (EMBO) long-term fellowship. T.A.R. was supported by a Royal Society University Research Fellowship (UF130382).

Author Contributions

Conceptualization, N.A.T.I. and A.A.P.; Funding acquisition, P.J.K. and T.A.R.; Investigation N.A.T.I. and A.A.P.; Resources, P.J.K. and T.A.R.; Supervision, P.J.K. and T.A.R.; Writing N.A.T.I with input from all authors.

Competing Interests

The authors declare no competing interests.

Materials and Correspondence

Material requests and correspondence should be addressed to N.A.T.I.

References

1. Chen, J. *et al.* Genome hypermobility by lateral transduction. *Science* **362**, 207–212 (2018).
2. Koonin, E. V. & Krupovic, M. The depths of virus exaptation. *Curr. Opin. Virol.* **31**, 1–8 (2018).
3. Frank, J. A. & Feschotte, C. Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* **25**, 81–89 (2017).
4. Touchon, M., Moura de Sousa, J. A. & Rocha, E. P. Embracing the enemy: The diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr. Opin. Microbiol.* **38**, 66–73 (2017).
5. Zimmerman, A. E. *et al.* Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **18**, (2019).
6. Filée, J., Pouget, N. & Chandler, M. Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. *BMC Evol. Biol.* **8**, 1–13 (2008).
7. Monier, A. *et al.* Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. *Proc. Natl. Acad. Sci.* **114**, E7489–E7498 (2017).
8. Monier, A. *et al.* Phosphate transporters in marine phytoplankton and their viruses: Cross-domain commonalities in viral-host gene exchanges. *Environ. Microbiol.* **14**, 162–176 (2012).
9. Monier, A. *et al.* Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Res.* 1441–1449 (2009) doi:10.1101/gr.091686.109.
10. Schulz, F. *et al.* Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436 (2020).
11. Aswad, A. & Katzourakis, A. Cell-derived viral genes evolve under stronger purifying selection in rhadinoviruses. *J. Virol.* **92**, e00539-18 (2018).

12. Schulz, F. *et al.* Giant viruses with an expanded complement of translation system components. *Science* **356**, 82–85 (2017).
13. Guglielmini, J., Woo, A. C., Krupovic, M., Forterre, P. & Gaia, M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19585–19592 (2019).
14. Enav, H., Mandel-Gutfreund, Y. & Béjà, O. Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* **2**, 1–11 (2014).
15. Rozenberg, A. *et al.* Lateral gene transfer of anion-conducting channel rhodopsins between green algae and giant viruses. *Curr. Biol.* 4910-4920.e5 (2020) doi:10.1016/j.cub.2020.09.056.
16. Vardi, A. *et al.* Host-virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19327–19332 (2012).
17. Gornik, S. G. *et al.* Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Curr. Biol.* **22**, 2303–2312 (2012).
18. Medina, E. M., Turner, J. J., Gordân, R., Skotheim, J. M. & Buchler, N. E. Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi. *Elife* **5**, e09492 (2016).
19. Pastuzyn, E. D. *et al.* The neuronal gene Arc encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. *Cell* **172**, 275–288 (2018).
20. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2002).
21. Fédry, J. *et al.* The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* **168**, 904–915 (2017).
22. Cornelis, G. *et al.* An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10991–E11000 (2017).
23. Irwin, N. A. T. *et al.* Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nat. Commun.* **9**, 1535 (2018).
24. Forterre, P. & Prangishvili, D. The major role of viruses in cellular evolution: Facts and hypotheses. *Curr. Opin. Virol.* **3**, 558–565 (2013).
25. Richards, T. A., Hirt, R. P., Williams, B. A. P. & Embley, T. M. Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**, 17–32 (2003).
26. Hayward, A., Cornwallis, C. K. & Jern, P. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 464–469 (2015).
27. Cock, J. M. *et al.* The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).
28. Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A. & Aylward, F. O. Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* **588**, (2020).
29. Leonard, G. *et al.* Comparative genomic analysis of the 'pseudofungus' *Hyphochytrium catenoides*. *Open Biol.* **8**, (2018).

30. Maumus, F. & Blanc, G. Study of gene trafficking between acanthamoeba and giant viruses suggests an undiscovered family of amoeba-infecting viruses. *Genome Biol. Evol.* **8**, 3351–3363 (2016).
31. Blanc, G., Gallot-Lavallée, L. & Maumus, F. Provirophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5318–E5326 (2015).
32. Davidson, A. R. A common trick for transferring bacterial DNA. *Science* **362**, 152–153 (2018).
33. Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: Prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395 (2017).
34. Pagarete, A., Allen, M. J., Wilson, W. H., Kimmance, S. A. & De Vargas, C. Host-virus shift of the sphingolipid pathway along an *Emiliania huxleyi* bloom: Survival of the fattest. *Environ. Microbiol.* **11**, 2840–2848 (2009).
35. Schneider-Schaulies, J. & Schneider-Schaulies, S. Sphingolipids in viral infection. *Biol. Chem.* **396**, 585–595 (2015).
36. Fernández De Castro, I., Tenorio, R. & Risco, C. Virus assembly factories in a lipid world. *Curr. Opin. Virol.* **18**, 20–26 (2016).
37. Michel, G., Tonon, T., Scornet, D., Cock, J. M. & Kloareg, B. Central and storage carbon metabolism of the brown alga Ectocarpus siliculosus: Insights into the origin and evolution of storage carbohydrates in eukaryotes. *New Phytol.* **188**, 67–81 (2010).
38. Durak, G. M. *et al.* A role for diatom-like silicon transporters in calcifying coccolithophores. *Nat. Commun.* **7**, 10543 (2016).
39. Armstrong, P. B., Armstrong, M. T., Pardy, R. L., Child, A. & Wainwright, N. Immunohistochemical demonstration of a lipopolysaccharide in the cell wall of a eukaryote, the green alga, *Chlorella*. *Biol. Bull.* **203**, 203–204 (2002).
40. Laurent, T. C. & Fraser, J. R. E. Hyaluronan. *FASEB* **6**, 2397–2404 (1992).
41. Loomis, W. F., Wheeler, S. A., Springer, W. R. & Barondes, S. H. Adhesion mutants of *Dictyostelium discoideum* lacking the saccharide determinant recognized by two adhesion-blocking monoclonal antibodies. *Dev. Biol.* **109**, 111–117 (1985).
42. Chisholm, R. L. *et al.* dictyBase, the model organism database for *Dictyostelium discoideum*. *Nucleic Acids Res.* **34**, 423–427 (2006).
43. Bandini, G. *et al.* An essential GDP-Fuc: β-D-Gal α-1,2-fucosyltransferase is located in the mitochondrion of *Trypanosoma brucei*. *bioRxiv* 726117 (2019) doi:10.1101/726117.
44. Fung, T. S. & Liu, D. X. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.* **5**, 296 (2014).
45. Raaben, M. *et al.* The ubiquitin-proteasome system plays an important role during various stages of the coronavirus infection cycle. *J. Virol.* **84**, 7869–7879 (2010).
46. Leier, H. C. *et al.* A global lipid map defines a network essential for Zika virus replication. *Nat. Commun.* **11**, 3652 (2020).

47. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
48. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
49. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
50. Uniprot Consortium. UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
51. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
52. Keeling, P. J. *et al.* The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
53. Nowack, E. C. M. *et al.* Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12214–12219 (2016).
54. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
55. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–10 (1990).
56. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
58. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121 (2013).
59. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
60. Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60**, 685–699 (2011).
61. Shimodaira, H. & Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol* **16**, 1114–1116 (1999).
62. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. *Nat. Ecol. Evol.* **1**, 0193 (2017).
63. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, 256–259 (2019).

64. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
65. Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
66. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
67. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
68. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
69. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
70. Huerta-Cepas, J. *et al.* eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
71. Irwin, N. A. T. *et al.* The function and evolution of motile DNA replication systems in ciliates. *Curr. Biol.* **31**, 66-76.e6 (2021).

Figures

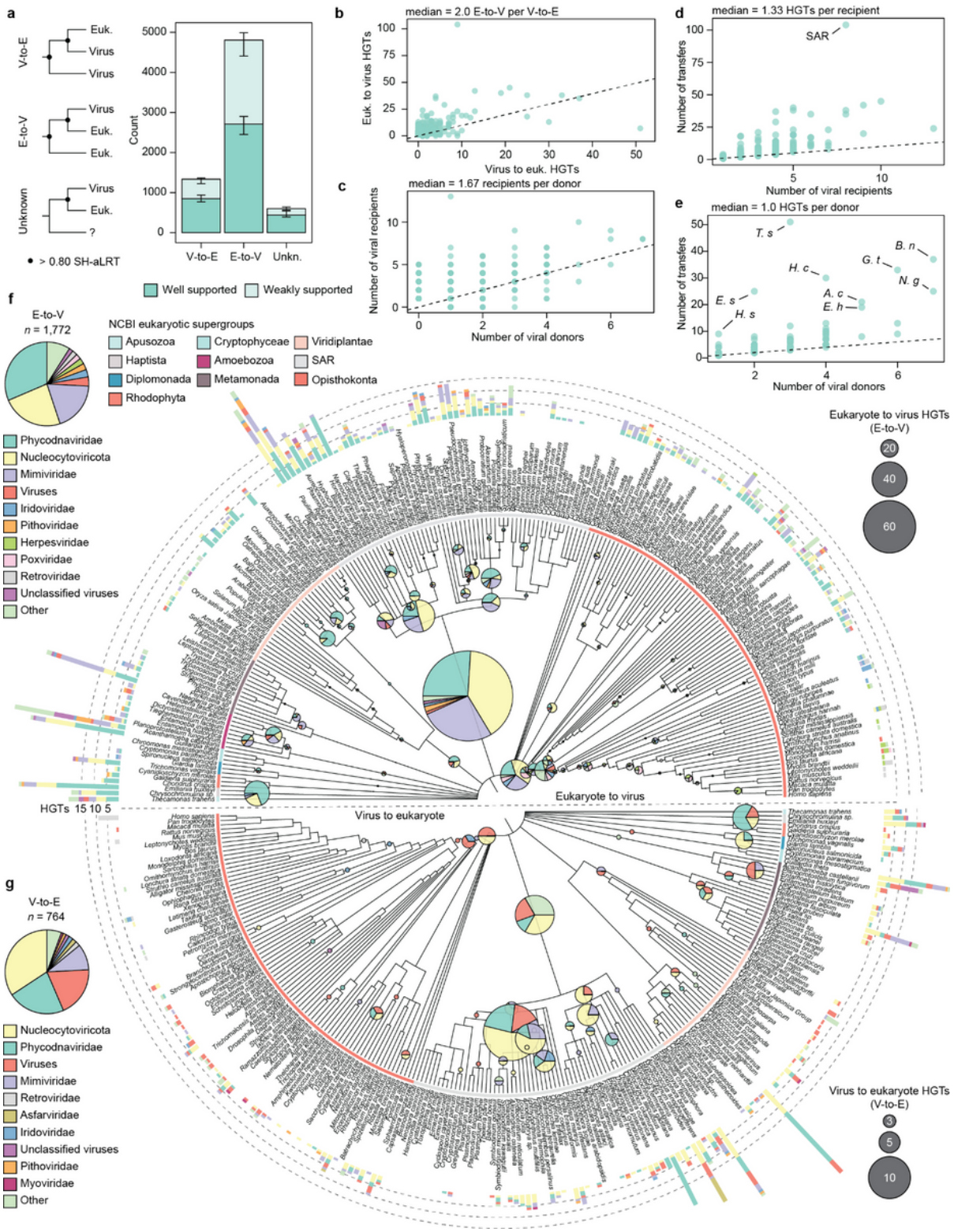

Figure 1

The mode and taxonomic distribution of viral-eukaryotic gene exchange. a. Transfers from eukaryotes to viruses (E-to-V), viruses to eukaryotes (V-to-E), and with unknown directionality (Unkn.). Weakly-supported transfers had long branching participants or ambiguous donors (see Materials and Methods). Error bars represent 95% confidence intervals from 1,000 bootstrap pseudoreplicates (random sampling of phylogenies with replacement). b-e. Scatter plots comparing gene exchange statistics. Points represent

eukaryotes (both species and higher-level classifications) and dashed lines represent lines of equality. f, g. Gene transfers from E-to-V (f) and V-to-E (g) across a eukaryotic phylogeny. Bar charts represent HGTs present in a given genome, whereas pie charts present inferred ancestral HGTs. Bar height and pie diameter reflect transfer frequency and colours reflect viral taxonomy. Viral taxa were mapped to their respective families, phyla, or genera. Taxonomic information and phylogenies are based on NCBI (National Center for Biotechnology Information) taxonomy. Transfers assigned to the last eukaryotic common ancestor are excluded but listed in Supplementary Table 1. Abbreviations: SAR, Stramenopila-Alveolata-Rhizaria; H. s., Homo sapiens; E. s., Ectocarpus siliculosus; T. s., Tetrabaena socialis; H. c., Hyphochytrium catenoides; A. c., Acanthamoeba castellanii; E. h., Emiliania huxleyi; G. t., Guillardia theta; B. n., Bigelowiella natans; N. g., Naegleria gruberi.

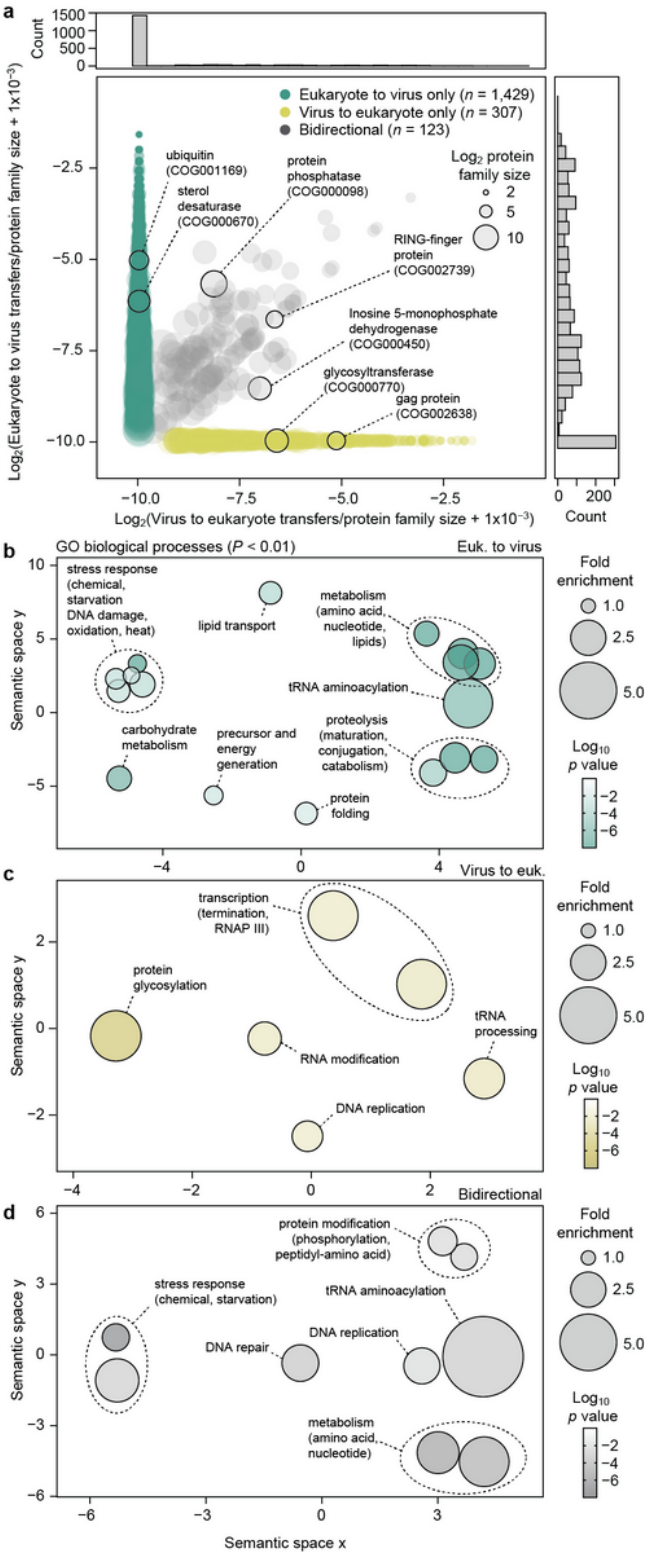

Figure 2

Gene function is related to transfer direction. a. A scatter plot relating the frequency of transfer events to protein families, normalized for family size (number of sequences). Functional annotations for exemplary families are highlighted. b-d. Scatter plots displaying enriched gene ontology (GO) biological process terms from protein families participating in unidirectional (b, c) and bidirectional transfer (d) relative to all eukaryotic protein families. Labeling has been summarized for clarity, but complete terms are available in

Supplementary Table 3. Semantic similarity was determined using REVIGO 47 and statistical significance was assessed using permutation tests (p < 0.01, n = 107).

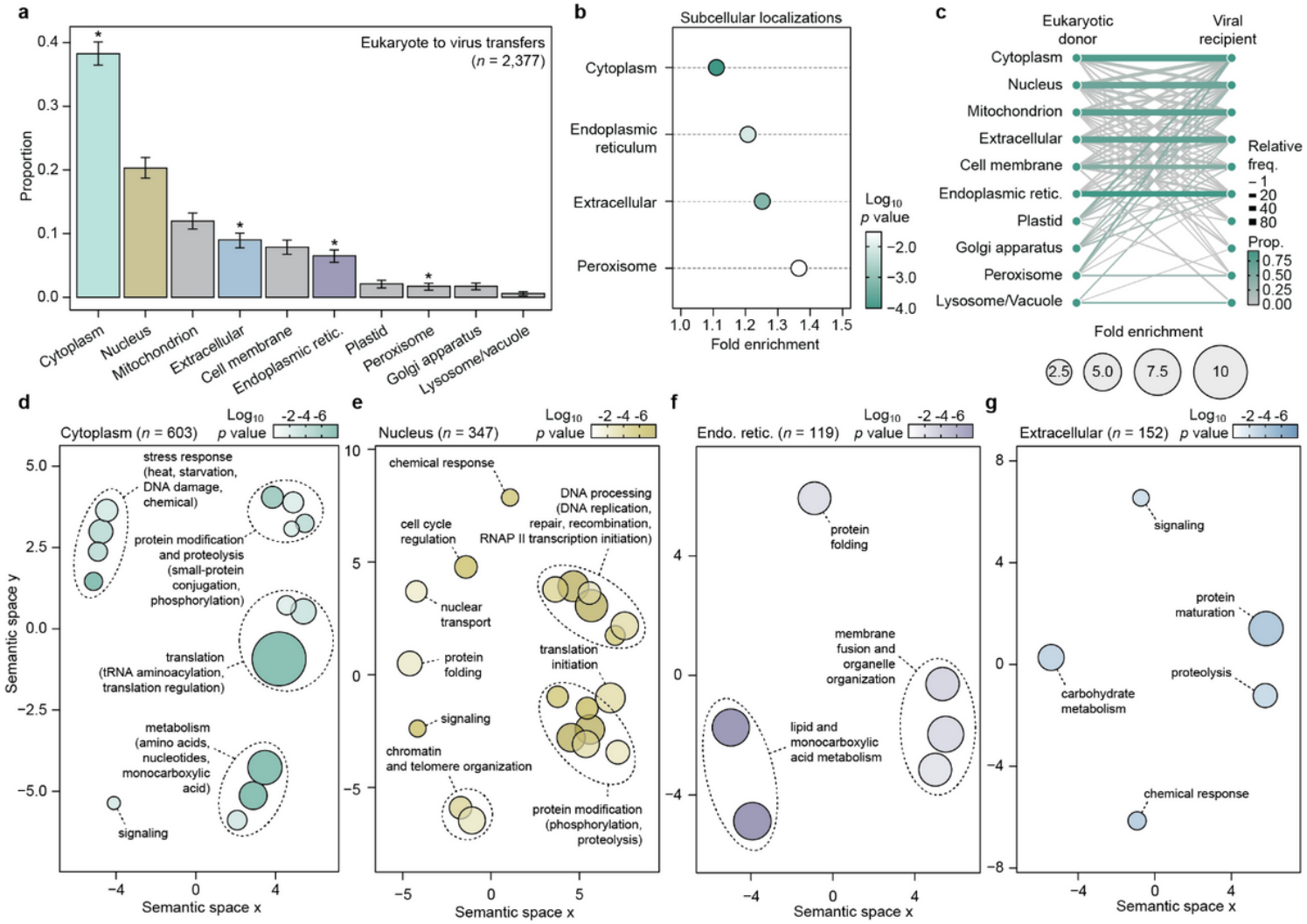

Figure 3

Predicted subcellular localizations and functions of eukaryote-derived viral genes. a. Proportions of subcellular localizations for eukaryote-to-virus HGTs based on the predicted targeting of eukaryotic donor sequences. Asterisks denote statistically significant enrichments (p < 0.05, see b). Error bars represent 95% confidence intervals determined from 1,000 bootstrap pseudoreplicates. b. The enrichment of subcellular compartments relative to total eukaryotic proteomes. Significance was assessed using permutation tests (n = 106). c. A comparison between the predicted localization of eukaryotic donors and their viral recipients. The relative frequencies and proportions are indicated by edge thickness and colour, respectively. d-g. Scatter plots displaying enriched GO biological process terms for families with a given donor localization relative to all eukaryotic protein families for localizations to (from left to right, colour coded as in a) the cytoplasm, nucleus, endoplasmic reticulum, and extracellular space. Labeling has been summarized for clarity but complete terms are available in Supplementary Table 4. Semantic similarity was determined using REVIGO and statistical significance was assessed using permutation tests (p < 0.01, n = 107).

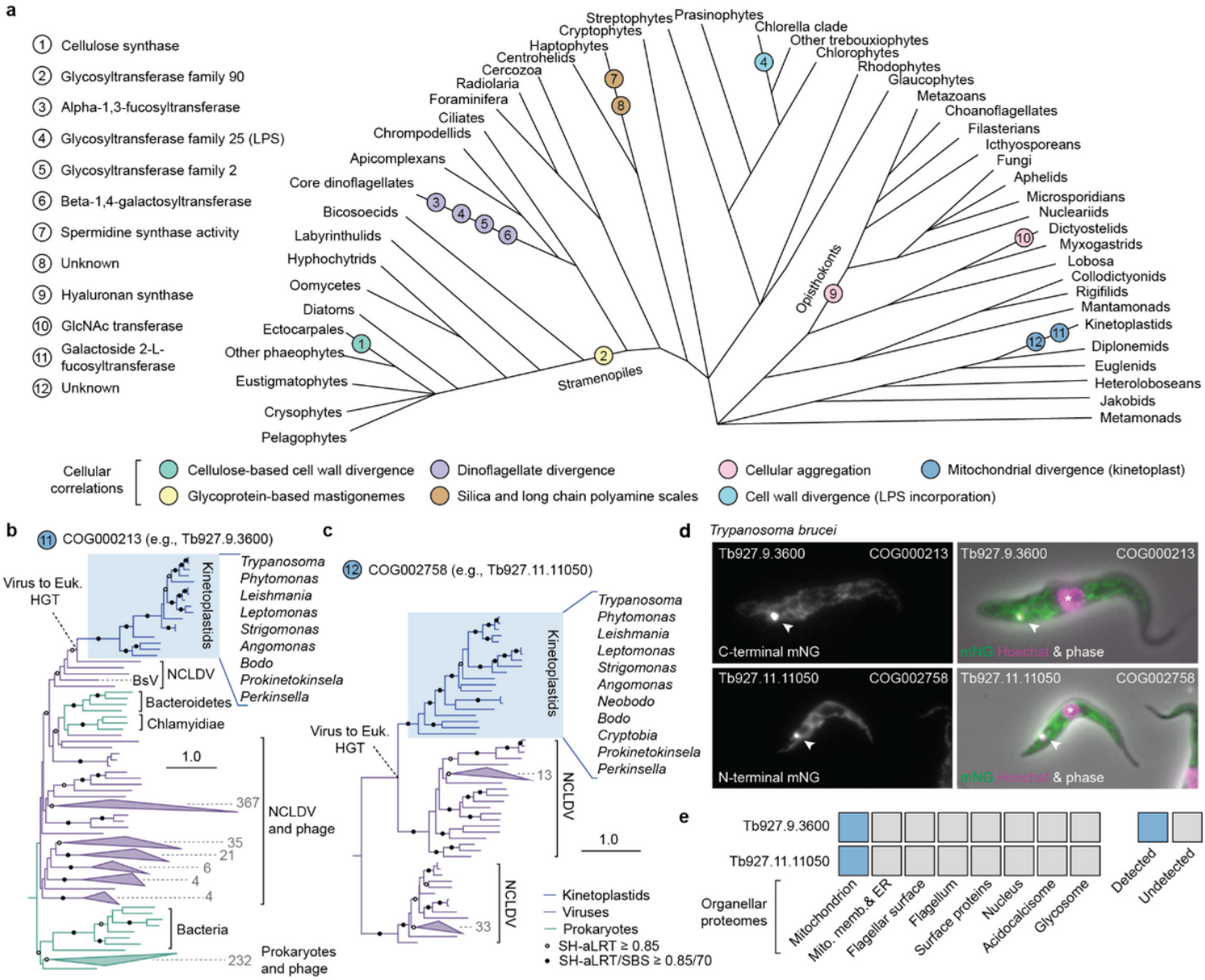

Figure 4

Recurrent acquisition of viral glycosyltransferases across the eukaryotic tree of life. a. A schematic eukaryotic phylogeny with glycosyltransferase acquisitions plotted at ancestral nodes. Protein families and annotations are denoted with numbers and phenotypic correlations are noted with colours. b, c. The phylogenetic origin of kinetoplastid glycosyltransferases from NCLDV by HGT. Maximum likelihood phylogenies were generated in IQ-Tree using the LG+F+R10 (b) and LG+F+R6 (c) substitution models as selected using ModelFinder and statistical support was assessed using SH-aLRT and standard bootstraps (n = 1,000) 48,49. Kinetoplastid genera and the gene identifiers for Trypanosoma brucei orthologues are noted. d. Viral-derived glycosyltransferases function in highly-derived kinetoplastid mitochondria (kinetoplasts). TrypTagDB fluorescent micrographs depicting the localization of both glycosyltransferases after labeling with mNeonGreen (left panel) and Hoechst DNA stain (right panel). White arrowheads denote the kinetoplast. e. Organellar proteomic data displaying the presence or absence of both glycosyltransferases. Proteomic data was assessed using TriTrypDB.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable5.xlsx
- ExtendedData.docx