



**HAL**  
open science

# Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes

Shannon Sibbald, Laura Eme, John Archibald, Andrew Roger

► **To cite this version:**

Shannon Sibbald, Laura Eme, John Archibald, Andrew Roger. Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. *Trends in Parasitology*, 2020, 36 (11), pp.927-941. 10.1016/j.pt.2020.07.014 . hal-03368658

**HAL Id: hal-03368658**

**<https://hal.science/hal-03368658v1>**

Submitted on 7 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Lateral gene transfer mechanisms and pan-genomes in eukaryotes**

2

3 **Shannon J. Sibbald<sup>1,2\*</sup>, Laura Eme<sup>3</sup>, John M. Archibald<sup>1,2</sup> and Andrew J. Roger<sup>1,2\*</sup>**

4 <sup>1</sup> *Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia,*  
5 *Canada*

6 <sup>2</sup> *Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University,*  
7 *Halifax, Nova Scotia, Canada*

8 <sup>3</sup> *Ecologie Systématique Evolution, CNRS, Université Paris-Sud, Orsay, France*

9 \*Correspondence: [Andrew.Roger@dal.ca](mailto:Andrew.Roger@dal.ca) (A. J. Roger), [Shannon.sibbald@dal.ca](mailto:Shannon.sibbald@dal.ca) (S. J. Sibbald)

10

11 **Keywords:** Lateral gene transfer, horizontal gene transfer, eukaryote genomes; pan-genomes

12

13 **Abstract**

14 Lateral gene transfer (LGT) is well known as an important driver of genome evolution in bacte-  
15 ria and archaea, but its importance in eukaryote evolution has yet to be fully elucidated. There is  
16 now abundant evidence indicating that LGT has played a role in the adaptation of eukaryotes to  
17 new environments and conditions, including host-parasite interactions. However, the mecha-  
18 nisms and frequency of LGT across the tree of eukaryotes remain poorly understood. Here we  
19 review evidence for known and potential mechanisms of LGT into diverse eukaryote lineages  
20 with a particular focus on protists, and discuss trends emerging from recently reported examples.  
21 We also explore the potential role of LGT in generating ‘pan-genomes’ in diverse eukaryotic  
22 species.

23

## 24 **The role and prevalence of lateral gene transfer in genome evolution**

25 Lateral (or horizontal) gene transfer (LGT or HGT) is an important mechanism by which genetic  
26 material is inherited in a non-vertical fashion. In bacteria, DNA is transmitted horizontally by  
27 **transduction, transformation** and **conjugation** (see Glossary) (e.g., [1]). Although the frequen-  
28 cy of LGT varies greatly amongst taxa, these processes are largely responsible for gene content  
29 variation between and within prokaryotic species. Indeed, the total gene content of a given spe-  
30 cies – the **pan-genome** – can exceed the number of ‘core’ genes shared between all strains of  
31 that species by an order of magnitude or more; those accessory genes are commonly acquired by  
32 LGT (e.g., [2] and references therein). It remains unclear what proportion of the genes found in  
33 some but not all strains of a bacterial species are adaptive versus neutral or deleterious [3,4], but  
34 there is no doubt that LGT has a major role in prokaryote genome evolution.

35 By contrast, the prevalence, mechanisms and roles of LGT in eukaryote evolution are  
36 poorly understood. Putative LGTs are increasingly reported as the number of comparative ge-  
37 nomic analyses of diverse unicellular and multicellular eukaryotes rapidly grows (Box 1). It is  
38 generally accepted that LGT likely occurs less frequently in eukaryotes relative to most prokary-  
39 otes (e.g., [5]), but this does not mean that it is an unimportant mechanism in eukaryotic genome  
40 evolution [6]. Laterally acquired genes can provide eukaryotic organisms with new functions that  
41 impact their biology and evolutionary trajectory (e.g., [7]); LGT has been suggested to play a  
42 role in the adaptation of eukaryotes to ‘extreme’ conditions such as high temperature (e.g., [8]),  
43 high salt (e.g., [9]), and anaerobiosis (e.g., [10, 11]). LGT also appears to have contributed to the  
44 evolution of parasitism (e.g., [12] and references there-in, [13,14]) and in the reversal from a  
45 parasitic to a free-living lifestyle (e.g., [15]).

46           How do eukaryotes acquire genes? Numerous potential sources of foreign DNA have  
47 been suggested (Figure 1) but the evidence supporting their involvement in LGT varies. Some  
48 phenomena, like **endosymbiotic gene transfer** (EGT), are well supported by evidence. Others,  
49 such as inter-domain conjugation or natural transformation, have been documented in the lab  
50 [16] but their frequencies and relevance in nature is unknown. Regardless of the source, a ‘suc-  
51 cessful’ LGT must involve several steps including: (i) the introduction of foreign DNA (or RNA)  
52 into the recipient eukaryotic cell; (ii) translocation of the DNA to the nucleus; and (iii) incorpo-  
53 ration of the DNA into one or more chromosomes, either accidentally by host DNA repair and  
54 recombination pathways, or by an active integration mechanism. For transferred DNA to be  
55 faithfully inherited, it must be stably integrated in all sister chromosomes and, for multicellular  
56 eukaryotes, must exist in the germline. To persist over longer time periods, newly acquired DNA  
57 must be fixed in the population by neutral drift and/or positive natural selection. If the transferred  
58 genes are expressed and provide a function that is useful (e.g., a novel adaptive trait, or compen-  
59 sation for the loss of a pre-existing gene), they can be maintained by purifying selection. Other-  
60 wise, further mutations may erode these genes into pseudogenes (e.g., [17]) and they will even-  
61 tually be lost or become unrecognizable intergenic DNA. Clearly the likelihood of any given  
62 LGT event being successful is multifactorial, and a number of factors potentially influence the  
63 frequency of LGT in any given species (Box 2).

64           Here we review evidence for several means by which foreign DNA is introduced into  
65 eukaryotic cells and potential mechanisms that integrate such DNA into eukaryotic genomes. We  
66 also discuss aspects of genome biology that may influence the rate of LGT and discuss its im-  
67 portance to adaptive evolution in eukaryotes by acting as a ‘catalyst’ for the acquisition of niche-  
68 specific traits.

69

## 70 **Viruses and their hosts exchange genetic material**

71 Virus-mediated transduction has long been recognized as an integral part of the biology  
72 of prokaryotic cells. A growing body of literature supports the notion that viruses can act as vec-  
73 tors of LGT from prokaryotes to eukaryotes and between eukaryotes (reviewed in [18]). The  
74 nature of viral transmission (horizontal), replication (inside the host cell) and survival strategies  
75 (e.g., integration into the host genome) together lead to intimate associations with their host's  
76 genetic material, providing ample opportunity for gene exchange. At least some viruses can in-  
77 fect a wide range of hosts and/or change hosts (e.g., [19]) and some eukaryotes harbour multiple  
78 divergent viruses, each of which may enable gene exchange beyond species boundaries. The  
79 virus-mediated export of nuclear DNA in plants was recently documented in real time by Catoni  
80 et al. [20], where *Beta vulgaris* chromosomal DNA integrated into the viral genome of a single-  
81 stranded DNA virus during infection was encapsidated into viral particles and spread throughout  
82 the plant population via subsequent viral infections. Notably, the *B. vulgaris* DNA present in the  
83 hybrid minicircles was replicated and transcribed in the newly infected host.

84 It is well established that host genes can be integrated into a viral genome and persist,  
85 particularly in double stranded DNA (dsDNA) viruses that infect a variety of eukaryotes. In the  
86 case of nucleocytoplasmic large DNA viruses, host DNA can make up a substantial portion of  
87 the virus' genome (e.g., [21]). A recent study of the giant DNA virus of the green alga  
88 *Tetraselmis*, TetV, identified 110 protein coding genes (17% of its protein coding genes) that  
89 appeared to be acquired from eukaryotes, including many specifically of Viridiplantae origin  
90 [21]. At least four of the eukaryotic homologs identified in TetV were novel viral genes that sug-  
91 gested recent incorporation of genes from both its host (*Tetraselmis*) and other unicellular eukar-

92 yotes that likely allow manipulation of the host fermentation pathways during anaerobic condi-  
93 tions. Giant virus genomes are highly mosaic – not only do they acquire genes from their eukar-  
94 yotic host(s), they also pick up genes from bacteria and other viruses present within the host cell  
95 (e.g., [22]).

96         The fact that genes of host origin exist in viral genomes implies that host DNA is regular-  
97 ly introduced into viral genomes with varying consequences on fitness; while some genes are  
98 fixed, others presumably persist temporarily over a limited number of replication cycles. Short-  
99 lived insertions, however, could still allow the foreign genes to be transferred to a new host. This  
100 was shown experimentally by Gilbert et al. [23], who investigated the frequency of gene transfer  
101 in 21 baculovirus (AcMNPV) populations infecting two moth species. Around 5% of AcMNPVs  
102 were found to harbor at least one moth-derived DNA sequence at a given time, all of which per-  
103 sisted only temporarily. There is nevertheless evidence for a continuous flow of genetic material  
104 from the host to AcMNPV and for repeated transfers of the short-lived insertions from the viral  
105 genome into the one of the various insects they infect. Host DNA was predominately inserted  
106 into viral genomes via transposition. However, there is also evidence for a variety of  
107 microhomology-mediated and homology-independent recombination events at double stranded  
108 breaks. Whether these recombination events were mediated by host or viral factors, or some  
109 combination of the two is currently unknown, but Gilbert et al. [23] speculate that as long as  
110 there is a double stranded break, viral and host DNA have the potential to be joined.

111         A large number of viral genomes and genome fragments from a wide range of both retro-  
112 viruses and other viruses have been found in eukaryote genomes, including those of multicellular  
113 organisms (e.g., [24-26]). In order to replicate, retroviruses become endogenized, i.e., integrated  
114 in their hosts genome via an autonomous mechanism using enzymatic machinery they encode.

115 Beyond retroviruses, it has been suggested that viral DNA is inserted via accidental chromoso-  
116 mal integration at double stranded DNA (dsDNA) breaks through homology mediated recomb-  
117 nation, non-homologous end joining repair, or a hybrid mechanism that utilizes the enzymatic  
118 machinery of retrotransposons [24]. For example, inactive remnants of endogenous  
119 florendoviruses (non-retroviral viruses that lack their own integrases) found across a broad diver-  
120 sity of plant genomes are frequently inserted in transposable element (TE)-rich regions, specifi-  
121 cally at TA dinucleotide simple repeats [27]. These repeats are thought to give rise to fragile  
122 chromosome regions, making them more prone to dsDNA breaks – breaks that create potential  
123 integration sites for foreign DNA.

124 Further clear evidence for the flux of genetic material from virus to host and *vice versa*  
125 has been reported in various eukaryotes. Substantial gene exchange is particularly evident be-  
126 tween the amoebozoan *Acanthamoeba castellanii* and its large DNA viruses [28]. 267 genes of  
127 probable viral origin were identified in the genome of *A. castellanii*, many of which were tran-  
128 scriptionally active and clearly integrated in the nuclear genome. Phylogenetic analyses also de-  
129 tected a kinase gene that appears to have been transferred on multiple occasions – at least once  
130 from eukaryotes to viruses and at least twice subsequently from viruses to amoebae [28]. Recent-  
131 ly, experimental evidence suggested that two densovirus genes transferred to the nuclear genome  
132 of pea aphids from a densovirus that infects rosy apple aphids are responsible for their ability to  
133 produce winged offspring during population crowding, a phenotype with clear adaptive benefit  
134 [29]. Similarly, endogenous chromosomally-transmitted bracoviruses in parasitic wasps produce  
135 dsDNA circles encoding virulence genes that are injected into host caterpillars, integrate into  
136 their genomes and are expressed suppressing the immune response allowing the wasp larvae to  
137 survive in the lepidopteran body. These endogenous viruses appear to have been vectors for

138 transfer of C-type lectin genes from the wasps to lepidopterans that have been domesticated by  
139 the latter and allow them to resist baculovirus infection [30].

140

#### 141 **Transposable elements—agents of intra- and inter-genomic mobility**

142 Transposable elements (TEs) have an intrinsic ability to move around within a genome,  
143 which may make it easier for them to be transferred to other genomes. Numerous examples of  
144 laterally transferred TEs have been reported in animals, plants, insects and fungi (e.g., [31–34]).  
145 A large-scale survey of TEs in 195 diverse insect genomes identified 2,248 lateral transfers of  
146 TEs, constituting 2-24% of a given insect genome [33]. Reiss et al. [34] recently extended this  
147 survey to screen 460 insect genomes for patterns of TE transfer. They concluded that DNA  
148 transposons were the most frequently transferred TE class in arthropods, followed closely by  
149 Long Tandem Repeats (LTRs), and that there was a clear excess of TE transfer in species hosting  
150 baculoviruses.

151 With the exception of the studies mentioned above, large-sale surveys of TEs in eukary-  
152 otes are generally lacking, especially in protists, making it difficult to make general conclusions  
153 about the extent of inter-genomic TE mobility and its global impact on eukaryotic genome evolu-  
154 tion. Regardless, studies looking at the rate of TE transfer have found that geographic proximity  
155 as well as genetic distance between the host and recipient lineage impact the rate of LGT (e.g.,  
156 [33]). Some of the most abundant retrotransposon families in eukaryotes were suggested to have  
157 been transferred multiple times between aquatic organisms (L1) and between animals (BovB) via  
158 various parasite vectors [35]. Some viral genomes contain TEs acquired from eukaryotes (e.g.  
159 [23,36]), including the widespread invasion of miniature inverted-repeat TEs found in high copy  
160 numbers in viruses that use host machinery to transpose themselves [37]. Several TEs have been

161 shown to produce extrachromosomal circular forms during transposition [38], which if  
162 encapsidated in a viral particle may be more likely to undergo virus-mediated LGT than linear  
163 DNA [18].

164 Under stress conditions, LTR retrotransposons are activated and generate double-stranded  
165 breaks in DNA (e.g., [39]); this could potentially facilitate integration of foreign DNA at these  
166 breakpoints via normal DNA repair pathways. Furthermore, imperfect excision of a TE from a  
167 chromosome, as has been observed in bacteria [40], can result in large adjacent portions of the  
168 genome accidentally being carried along with a TE during transposition or horizontal transfer.  
169 Due to the ability of TEs to capture and transduce sequences at a high frequency, it has been  
170 suggested that horizontal TE transfer could be responsible for the direct movement of genes  
171 among eukaryotes as well [41]. Consistent with this hypothesis, transferred genes are often asso-  
172 ciated with TE-rich regions in metazoan genomes [14]. The transfer of a wheat virulence protein  
173 (*ToxA*) between three fungal wheat pathogens in two independent events was likely facilitated by  
174 a type II DNA transposon [42]. Long-read sequencing based genome assemblies for several of  
175 these fungi showed that the genomic context of the two *ToxA* LGTs included 14 kbp of both cod-  
176 ing and non-coding DNA, including conserved terminal inverted repeats bearing strong resem-  
177 blance to features of type II DNA transposons, suggesting LGT and lateral TE transfer can occur  
178 simultaneously [43]. Altogether, this strongly suggests that TEs are not only transferred laterally,  
179 but that they also can facilitate the movement of other genes between species.

180

## 181 **Endosymbiosis and gene transfer**

182 A large variety of eukaryotes harbour prokaryotic and/or eukaryotic **endosymbionts** ex-  
183 hibiting varying degrees of cellular and genomic integration. In a number of cases (e.g., mito-

184 chondria and plastids), the endosymbionts have co-evolved with the host to the point where they  
185 have become permanent, semi-autonomous organelles. In any case, the close proximity of the  
186 host nucleus and endosymbionts or endosymbiont-derived organelles provides an opportunity for  
187 the movement of genetic material [43,44]. Endosymbiont-to-host nucleus gene transfers played  
188 an important role in the evolution of mitochondria and plastids, a process specifically referred to  
189 as endosymbiotic gene transfer (EGT) [45,46]. In these instances, if the transfer of genes from  
190 the endosymbiont-derived organelle to host is followed by the retargeting of the gene product to  
191 the organelle, the copy of the gene in the organellar genome can be lost. In the fullness of time,  
192 the result of EGT is the existence of complementary gene inventories and coordinated metabo-  
193 lisms in host and endosymbiont-derived compartments [10,44]. The substrate for EGT is thought  
194 to be organelle lysis; the frequency of EGT over short evolutionary timescales appears to be pos-  
195 itively correlated with the number of organelles per cell (see [16,47] and references therein).

196         The extent to which bacteria other than the endosymbiotic progenitors of mitochondria  
197 and plastids contributed to organelle establishment remains controversial (e.g., [10, 48–50]). In  
198 the case of plastids, phylogenetic evidence suggests that LGTs from diverse bacteria – not just  
199 cyanobacteria – may have facilitate organelle evolution (e.g., [51]). Beyond canonical plastids,  
200 the contribution of LGT to the recently evolved ‘chromatophores’ in the rhizarian amoeba  
201 *Paulinella chromatophora* is clear. Sequencing of the *P. chromatophora* genome revealed 228  
202 nuclear genes of various bacterial affinities, only ~25% of which appear to come from the  
203 cyanobacterium-derived chromatophore [52]. The suite of EGTs and LGTs, along with native  
204 host nuclear genes, contribute to many essential biosynthetic pathways partially encoded by the  
205 substantially reduced chromatophore genome, and encode transporters facilitating host-  
206 chromatophore metabolite exchange [52,53]. The phagotrophic nature of the ancestors of *P.*

207 *chromatophora* likely played an important role in sourcing these various bacterial genes, driving  
208 LGT and, ultimately, chromatophore integration.

209         Various other examples of endosymbiosis-associated gene transfers have also been doc-  
210 umented. Some of these transfers are directly from endosymbionts to host, while others seem to  
211 be LGTs from ‘third party’ organisms that allow endosymbionts to lose the corresponding genes.  
212 The latter is exemplified by the mealybug *Planococcus citri*, in which the loss of enzymes for  
213 synthesizing eight amino acids in its endosymbiont *Tremblaya* were made possible by >20 LGTs  
214 from various bacteria to the *P. citri* nuclear genome [54]; a related mealybug harbors a less  
215 genomically streamlined *Tremblaya* endosymbiont that still encodes all the enzymes necessary  
216 for the synthesis of these amino acids. This endosymbiosis actually involves a tripartite relation-  
217 ship between *P. citri* and two bacteria – a betaproteobacterium (*Tremblaya*) and its  
218 gammaproteobacterial endosymbiont *Moranella* – which exhibit complementary gene losses and  
219 are themselves metabolically intertwined [54]. Peptidoglycan biosynthesis is just one pathway  
220 that appears to be a mosaic of *P. citri* encoded LGTs and endosymbiont encoded genes. Using a  
221 variety of techniques including Nanoscale Secondary Ion Mass Spectrometry (NanoSIMS),  
222 Bublitz et al. [55] recently showed that the laterally acquired genes for peptidoglycan synthesis  
223 are translated in the cytoplasm of *P. citri* and the resulting proteins are transported across multi-  
224 ple membranes to the *Moranella* endosymbiont where peptidoglycan is assembled. These results  
225 clearly show that LGT can give rise to stable, complex, mosaic biochemical pathways in  
226 endosymbioses outside the bounds of mitochondria and plastids.

227         Several species of single-celled Foraminifera harbor intracellular bacteria capable of  
228 denitrification – a process widespread in prokaryotes but not in eukaryotes, which allows for the  
229 storage and use of nitrate in energy metabolism [56]. Recently, a novel denitrification pathway

230 was identified in the genome of the foraminiferan *Globulimina* where two of the key enzymes  
231 are clearly of prokaryotic origin and likely originated from the denitrifying endosymbiotic bacte-  
232 ria [57]. In the ctenophore *Mnemopsis leidy*, various LGTs have been suggested to be a result of  
233 the association between these enigmatic invertebrates with various bacteria and protists such as  
234 amoebae [58]. Numerous studies have shown a variety of LGTs from *Wolbachia* endosymbionts  
235 into various insect and nematode hosts ([59] and references therein), ranging from smaller inser-  
236 tions to genome-sized fragments. These are just a few of many examples illustrating that  
237 endosymbionts contribute DNA to their hosts and that gene transfers from various sources during  
238 and/or prior to the establishment of a permanent endosymbiont can influence host-endosymbiont  
239 integration. The fact that EGTs have occurred in diverse lineages across the eukaryotic tree  
240 shows that there are few – if any – barriers to the uptake and integration of foreign DNA into  
241 eukaryote nuclear genomes.

242

### 243 **“You are what you eat”**

244 Many protists feed on other organisms via phagocytosis. Doolittle’s ‘you are what you  
245 eat’ hypothesis [60] posits that both prokaryotic and eukaryotic prey could act as sources of new  
246 genes simply as a by-product of digestive vacuole rupture and release of nucleic acids. In this  
247 respect, the overall mechanism of gene transfer from phagocytosed prey is similar to gene acqui-  
248 sition from endosymbionts. Even if food vacuole or symbiont rupture is a very rare occurrence,  
249 both generate free DNA that could enter the nucleus and be incorporated into the nuclear genome  
250 at a low frequency. Such sequences will accumulate mutations and may be deleted. But if the  
251 gene within the inserted DNA happens to be expressed, it could be fixed in the population by  
252 drift or, if some advantage is conferred, positive selection. Because of their overall similarity in

253 mechanism and outcome, it is difficult if not impossible to determine if gene transfer events orig-  
254 inated from transient ancestral endosymbionts or phagocytosed food. A number of potential ex-  
255 amples of the latter have nevertheless been suggested.

256         Adaptive gene transfers of bacterial origin in the bacterivorous protist *Halocafeteria*  
257 *seosinensis* allow it to produce an osmoprotectant that relieves significant osmotic stress found in  
258 its hypersaline habitat. This, in combination with other LGTs from bacteria, likely facilitated the  
259 adaptation of *H. seosinensis* to a high salt environment [9]. In the anaerobic oxymonad  
260 *Monocercomonoides*, a laterally acquired SUF-system from a bacterial donor (with genes veri-  
261 fied to be in the nucleus by fluorescent in situ hybridization) is likely linked to the ability of  
262 *Monocercomonoides* to ingest and digest bacterial prey [61]. This laterally acquired pathway  
263 allows for iron-sulfur cluster assembly to take place in the cytosol rather than *via* the typical ISC  
264 pathway in mitochondrion, ultimately making the last essential function of *Monocercomonoides*'  
265 mitochondria-related-organelle redundant and dispensable [61]. Eukaryovores are relatively rare  
266 compared to bacterivorous protists, which may contribute to apparent differences in the frequen-  
267 cy of gene transfer between eukaryotes and between eukaryotes and prokaryotes [62]. However,  
268 these differences may also reflect the difficulties associated with confidently identifying eukary-  
269 ote-to-eukaryote gene transfer events (Box 3).

270

### 271 **Parasites as LGT recipients and vectors**

272         LGT has played an important role in the adaptation of many successful parasite lineages  
273 to their host and interactions with other host-associated microbes. Cases of adaptive LGT from  
274 prokaryotes to a variety of unicellular and multicellular parasitic eukaryotes have been well doc-  
275 umented (e.g., [7,12]). For example, LGTs have been reported in diplomonad protozoa (e.g.,

276 *Spiroucleus* and *Giardia intestinalis* (e.g., [63]), *Plasmodium*, *Toxoplasma*, *Leishmania*,  
277 *Trichomonas* and *Entamoeba* [12,64] as well as parasitic nematodes (e.g., [14, 65]).

278         Several trends in the patterns of laterally transferred genes into parasitic lineages have  
279 been noted. For example, there is an excess of genes received from organisms that live in the  
280 same ‘host’ environment as the parasites (e.g., [12, 64]). Furthermore, the functions of the ac-  
281 quired genes tend to reflect the lifestyles of the different parasite types. For example, the muco-  
282 sal parasite *Trichomonas vaginalis* has acquired genes related to harvesting glycans from host  
283 epithelial tissues and secretions (Box 4). A recent analysis of the anaerobic gut mucosal parasite  
284 *Blastocystis* sp. is consistent with these observations. 2.5% of the protein-coding genes in this  
285 organism were acquired laterally from a wide range of both prokaryotes and eukaryotes, allow-  
286 ing for successful host infection, evasion of host immune systems, modulation of other gut mi-  
287 crobes (e.g., via production of indole and polyketides) and metabolic adaptation to the gut envi-  
288 ronment [13]. Several of the LGTs in *Blastocystis* appear to have come directly from animal  
289 hosts. Others appear to have been transferred multiple times between this organism and other  
290 mucosal protists (e.g., *Trichomonas*, *Giardia* and *Entamoeba*), suggesting the potential existence  
291 of a pool of adaptive genes shared between anaerobic mucosal parasites. Numerous other strong  
292 examples of LGT facilitating parasite adaptation have been reported across the eukaryotic tree.  
293 For example, Dean et al. [66] used a combination of ancestral gene reconstruction and functional  
294 assays to show that the acquisition of a single nucleotide transporter from bacteria to the com-  
295 mon ancestor of *Microsporidia* species allowed them to become energy parasites that steal host  
296 ATP, GTP and NAD<sup>+</sup> and subsequently lose the capacity to synthesize nucleotides on their own.

297         In phytoparasitic nematodes, functionally important genes encoding enzymes involved in  
298 plant cell-wall degradation and related metabolism were identified as LGTs from bacteria that

299 are plant pathogens of the same hosts, as well as from a variety of eukaryotes (including the plant  
300 pathogen *Phytophthora infestans*) [14]. Interestingly, LGTs in these nematodes are often found  
301 in TE regions of the genome, and the bacterial homologs are found on plasmids, suggesting a  
302 mechanistic role for mobile genetic elements in LGT. Another example of host-parasite mediated  
303 LGT is the acquisition of mammalian TE families by the blood-sucking insect *Rhodnius prolixus*  
304 and the pond snail *Lymnaea stagnalis*. These TEs had extremely high sequence identities to TEs  
305 in the genomes of host mammals of the insect and trematode parasites for which the snail acts as  
306 an intermediate host [67], implicating host-parasite interactions in horizontal transmission of the  
307 TEs. Based on the foregoing examples, it seems likely that parasites that infect a wide range of  
308 hosts or species that exist in general symbiotic associations could be particularly potent vectors  
309 for LGT.

310

### 311 **Conjugation and natural transformation in eukaryotes**

312 A wide variety of eukaryotes have been transformed in the lab. While artificial in nature,  
313 such experiments indicate that there is no fundamental obstruction to LGT when barriers to  
314 transfer are intentionally removed. Established eukaryotic transformation systems are frequently  
315 exploited for commercial applications such as the production of genetically modified crops. For  
316 example, *Agrobacterium tumefaciens*, which uses a type IV secretion system to inject a tumor  
317 inducing (Ti) plasmid carrying specific genes into plant cells, has been used to modify plant spe-  
318 cies to confer pest resistance [68]. Once in the host cell, the Ti plasmid utilizes the plants ma-  
319 chinery to enter the nucleus and integrate into the genome, using eukaryotic promoters to express  
320 its genes. Importantly, there is evidence that this transformation occurs naturally in plants as  
321 well. Genes from *Agrobacterium* have been found to reside and be expressed in the germ line of

322 various plant species that have not been artificially transformed (e.g., [69,70]), suggesting that  
323 these gene transfers are maintained and functional.

324         Recently a 100% efficient transformation system was developed in the red alga  
325 *Porphyridium purpueum* where the transformation vectors are maintained episomally as  
326 extrachromosomal high-copy number plasmids [71]. This shows that different bacterial origins  
327 of replication can function in a eukaryote system and result in stable, autonomous replication  
328 within the nucleus. Naturally occurring extrachromosomal plasmids of unknown origin have  
329 been found in some red algae; these plasmids contain both bacterial and viral ORFs (e.g., [72]).  
330 The apparent genetic compatibility between plasmid replication systems of bacteria and compo-  
331 nents of DNA replication machinery in the red algal nucleus [73] raises intriguing questions  
332 about algal genome evolution. The extended presence and replication of a nuclear episomal ele-  
333 ment could provide opportunities for gene transfer to the nuclear genome and may have facilitat-  
334 ed the contribution of bacterial genes to red algae (estimated to be between 1% and 9.3% of the  
335 total gene content, depending on the lineage [73,74]).

336         Interestingly, a conjugation-based transformation system capable of delivering episomal  
337 vectors from *E. coli* has been developed for the diatoms *Phaeodactylum tricornutum* and  
338 *Thalassiosira pseudonana* [75]. Although direct conjugation has not been observed in nature,  
339 this system suggests that bacterial conjugation may be one mechanism that could explain the  
340 high percentage of putative bacterial-derived genes (5-10% of genes) in diatom genomes [76].  
341 The transfer of plasmids via bacterial conjugation has been observed in yeast [77], as has the  
342 natural transformation of exogenous DNA under non-artificial conditions [78].

343

344 **Difficulties in detecting LGT in eukaryotes**

345 Eukaryotes typically have relatively large and complex genomes compared to prokary-  
346 otes, making genome sequencing, assembly and analysis difficult. The paucity of high-quality  
347 genome data from diverse eukaryotes, especially from closely related species and multiple  
348 strains of the same species, is a present barrier to the study of LGT [79]. In regions of the eukar-  
349 yotic tree in which there is denser genomic sampling of lineages, the extent of adaptive LGT is  
350 becoming clearer. For example, Ropars et al. [80] analyzed 10 high-quality genomes from  
351 *Penicillium* in the context of genome data from 416 strains of 65 fungal species. The authors con-  
352 fidently identified 104 LGTs between *Penicillium* species, including some large recently trans-  
353 ferred genomic regions spanning more than 10 kb, some of which show signatures of TE mediat-  
354 ed insertion [80]. Gene expression analysis and fitness experiments confirmed that at least some  
355 of the LGTs encode functional genes that provide metabolic adaptations to a cheese environment  
356 (e.g. lactose metabolism) or antifungal proteins used to out-compete other fungal species [80].  
357 Until a similar density of high-quality genomic sampling is available for other branches of the  
358 eukaryote tree— particularly for microbial eukaryote lineages – it is impossible to know whether  
359 the dynamic nature of fungal genomes is the exception or the rule.

360 Distinguishing true LGTs from contamination (see 81) can be difficult and mitigated by  
361 careful inspection of candidate genes (Box 3). Many (but by no means all) published studies  
362 have used appropriately stringent phylogenetic criteria for reporting LGT. Initial bioinformatic  
363 steps generally lean towards the exclusion of bacterial/symbiont sequences from assemblies,  
364 which means that the impact of LGT to a given genome can be underestimated. Additional diffi-  
365 culties arise when trying to determine the scale of how far back in evolutionary time one has to  
366 go to consider a gene as having been vertically or laterally transferred. More ancient LGTs are  
367 harder to detect as the sequence similarity of a transferred gene to the gene of the donor lineage

368 decays over time. At the same time, confidently detecting eukaryote-eukaryote LGT is generally  
369 more difficult than prokaryote-eukaryote LGT due to the more recent common ancestry shared  
370 between donor and recipient, resulting in less obvious phylogenetic incongruence. A ‘patchy’  
371 gene distribution across the eukaryote tree is often interpreted as evidence of LGT between eu-  
372 karyotic lineages. However, secondary loss of ancestral genes is well known to have occurred in  
373 parallel in many eukaryotic lineages (e.g. genes in mitochondrial genomes, reviewed in [10]) and  
374 is sometimes an equally plausible alternative explanation for the gene distribution pattern.  
375 Therefore, strongly supported single gene trees that contradict the expected species tree relation-  
376 ships are needed to support suspected LGT cases. This standard is sometimes challenging to  
377 achieve due to the limited phylogenetic signal within many single gene alignments. Identifying  
378 gene transfers in eukaryotes also requires a sufficient level of taxonomic sampling of the given  
379 gene of interest, which is often lacking. Strategic sequencing of diverse eukaryote lineages will  
380 make LGT detection easier in the future.

381

## 382 **LGT and pan-genomes in eukaryotes**

383 In prokaryotes, within-species variation in gene content is such that the core genome (i.e.,  
384 gene families found across all or most strains) represents 3-84% of the total gene families found  
385 in all sequenced strains of a given species (reviewed by [3]). The accessory genome includes  
386 genes found in a single or subset of strains (Figure 2). This pan-genome pattern is due to the  
387 combined effects of LGT and differential gene loss.

388 The extent to which a pan-genome-like pattern is observed in various eukaryote lineages  
389 – and, in particular, protists – is less clear, as inferences are limited by the paucity of genomes  
390 that cover the breadth of diversity for a given species outside of animals, plants and fungi [79].

391 Another problem is the difficulty in obtaining accurate and comprehensive gene predictions in  
392 eukaryotes, as gene prediction is much more difficult than in prokaryotes because of the pres-  
393 ence of spliceosomal introns and alternative splicing. As a result, potential pan-genome patterns  
394 must be cross-checked to be certain they are not simply artefacts of failure to predict existing  
395 genes. Nevertheless, across various model fungal species where a dozen or more genomes have  
396 been sequenced, pan-genomes have been observed with core genomes ranging from 80-90% of  
397 the species coding capacity [e.g., 82 and references therein]. Some of the accessory genes are  
398 inferred to be a result of LGT; others appear to be a consequence of gene duplication,  
399 pseudogenization and outright gene loss. In plants, the accessory genome component of the spe-  
400 cies pan-genome seems to be more extensive (up to ~47% of the pan-genome). For example,  
401 across >3,000 genomes of *Oryza sativa* L., 46.5% of the pan-genome was found to be accessory  
402 genes, most of which are related to immune and defense responses [83].

403         In the world of protists and algae, fascinating insight into the eukaryotic pan-genome has  
404 come from investigation of the coccolithophorid *Emiliania huxleyi* [84]. In analyzing the ge-  
405 nomes of 14 highly-similar strains of *E. huxleyi* (>98% 18S ribosomal DNA (rDNA) identity),  
406 Read et al. [84] found a core genome (experimentally confirmed via microarrays) comprised of  
407 only 69.5% of the total gene set. While Read et al. were unable to confidently assign LGTs from  
408 bacteria or viruses as the main source of the accessory genes, they did find that *E. huxleyi* ge-  
409 nomes contain a lot of repetitive elements, which may foster accessory gene formation and serve  
410 as preferential target sites for LGT.

411         As with prokaryotes, there are serious problems in defining the ‘pan-genome’ because it  
412 is confounded with how species are defined and delimited. For example, Worden et al. [85]  
413 found significant gene content differences between two isolates of *Micromonas* that were origi-

414 nally considered the same species on the basis of morphology and 18S rDNA similarity. The fact  
415 that these two isolates were found to share only 90% of their genes was interpreted to mean that  
416 they were different species. Yet, an alternative interpretation is that these two isolates are mem-  
417 bers of the same pan-genome-containing species. In the absence of evidence for sexual recomb-  
418 ination between individuals of a putative species – information that is lacking for the majority of  
419 protists – defining a species is difficult. Without a clear species definition for most microbial  
420 eukaryotes, it is difficult to distinguish between these two scenarios.

421

## 422 **Concluding Remarks**

423         The past few years have seen a rapid increase in the sequencing of complete genomes  
424 from diverse eukaryotes and, concordantly, abundant evidence for LGT into eukaryotic genomes.  
425 Many questions nevertheless remain unanswered (see Outstanding Questions). Indeed, for most  
426 documented cases that are not obviously EGTs, the means by which the transferred DNA was  
427 introduced into cell and incorporated into the chromosomes remain unclear. Many of the mecha-  
428 nisms that could have been involved, and their relative importance, are ‘known unknowns’. For  
429 example, endosymbiont- or ‘you-are-what-you-eat’ mediated transfers are probably much more  
430 common in eukaryotes that commonly harbor endosymbionts and phagocytose prey, respective-  
431 ly. On the other hand, virus-mediated transfer will depend on the number and types of viruses  
432 that infect particular lineages and the degree to which they can integrate into host genomes and  
433 mobilize host DNA. To address these questions the following data are needed. First, a greater  
434 density of sampling of eukaryotic genomes from each of the major eukaryotic groups should  
435 allow us to ‘catch’ more very recent transfer events that may still bear the hallmarks of their ori-  
436 gins (e.g., flanking TEs, viral genes, breakpoints of recombination events etc.). This should also

437 allow us to assess the extent to which pan-genomes are a common feature of diverse eukaryote  
438 species, assuming we can get reliable information about how such species are delimited (i.e.,  
439 evidence for recombination within species). Our understanding of eukaryote LGT will also be  
440 greatly enhanced by the development of new model systems and long-term laboratory evolution  
441 experiments that allow the process of LGT to be studied in real time. Ideally, organisms from  
442 different branches of the eukaryotic tree should be developed as model systems and the various  
443 insights they yield considered in isolation and together. Long-term progress in our understanding  
444 of the frequency and significance of LGT in eukaryotes will rely on the synthesis of information  
445 gleaned from both experimental and bioinformatic approaches.

446

#### 447 **References**

448 1 Soucy, S. et al. (2015) Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16,  
449 472–482

450

451 2 Land, M. et al. (2015) Insights from 20 years of bacterial genome sequencing. *Funct. Integr.*  
452 *Genomics* 15, 141–161

453

454 3 McInerney, J.O. et al. (2017) Why prokaryotes have pangenomes. *Nat. Microbiol.* 2, 17040

455

456 4 Vos, M. and Eyre-Walker, A. (2017) Are pangenomes adaptive or not? *Nat. Microbiol.* 2,  
457 1576–1576

458

459 5 Danchin, E.G. (2016) Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube?  
460 *BMC Biol.* 14, 101

461

462 6 Leger, M.M. et al. (2018) Demystifying eukaryote lateral gene transfer. *BioEssays* 40,  
463 1700242

464

465 7 Lukeš, J. and Husník, F. (2018) Microsporidia: a single horizontal gene transfer drives a great  
466 leap forward. *Curr. Biol.* 28, R712–R715

467

468 8 Schönknecht, G. et al. (2013) Gene transfer from bacteria and archaea facilitated evolution of

469 an extremophilic eukaryote. *Science* 339, 1207–1210  
470  
471 9 Harding, T. et al. (2017) Adaptations to high salt in a halophilic protist: differential expression  
472 and gene acquisitions through duplications and gene transfers. *Front. Microbiol.* 8, 944  
473  
474 10 Roger, A.J. et al. (2017) The origin and diversification of mitochondria. *Curr. Biol.* 27,  
475 R1177–R1192  
476  
477 11 Stairs, C.W. et al. (2018) Microbial eukaryotes have adapted to hypoxia by horizontal acquisi-  
478 tions of a gene involved in rhodoquinone biosynthesis. *eLife* 7, e34292  
479  
480 12 Hirt, R.P. et al. (2015) Lateral gene transfers and the origins of the eukaryote proteome: a  
481 view from microbial parasites. *Curr. Opin. Microbiol.* 23, 155–162  
482  
483 13 Eme, L. et al. (2017) Lateral gene transfer in the adaptation of the anaerobic parasite  
484 *Blastocystis* to the gut. *Curr. Biol.* 27, 807–820  
485  
486 14 Paganini, J. et al. (2012) Contribution of lateral gene transfers to the genome composition and  
487 parasitic ability of root-knot nematodes. *PLoS One* 7, e50875  
488  
489 15 Xu, F. et al. (2016) On the reversibility of parasitism: adaptation to a free-living lifestyle via  
490 gene acquisitions in the diplomonad *Trepomonas* sp. PC1. *BMC Biol.* 14, 62  
491  
492 16 Bock, R. (2017) Witnessing genome evolution: experimental reconstruction of endosymbiotic  
493 and horizontal gene transfer. *Annu. Rev. Genet.* 51, 1–22  
494  
495 17 Strese, Å. et al. (2014) A recently transferred cluster of bacterial genes in *Trichomonas*  
496 *vaginalis* – lateral gene transfer and the fate of acquired genes. *BMC. Evol. Biol.* 14, 119  
497  
498 18 Gilbert, C. and Cordaux, R. (2017) Viruses as vectors of horizontal transfer of genetic materi-  
499 al in eukaryotes. *Curr. Opin. Virol.* 25, 16–22  
500  
501 19 Longdon, B. et al. (2014) The evolution and genetics of virus host shifts. *PLoS Pathog.* 10,  
502 e1004395  
503  
504 20 Catoni, M. et al. (2018) Virus-mediated export of chromosomal DNA in plants. *Nat.*  
505 *Commun.* 9, 5308  
506  
507 21 Schvarcz, C.R. and Steward, G.F. (2018) A giant virus infecting green algae encodes key  
508 fermentation genes. *Virology* 518, 423–433

509  
510 22 Moreira, D. and Brochier-Armanet, C. (2008) Giant viruses, giant chimeras: the multiple evo-  
511 lutionary histories of Mimivirus genes. *BMC Evol. Biol.* 8, 12  
512  
513 23 Gilbert, C. *et al.* (2016) Continuous influx of genetic material from host to virus populations.  
514 *PLoS Gene.* 12, e1005838  
515  
516 24 Feschotte, C. and Gilbert, C. (2012) Endogenous viruses: insights into viral evolution and  
517 impact on host biology. *Nat. Rev. Genet.* 13, 283–296  
518  
519 25 Marchi, E. *et al.* (2014) Unfixed endogenous retroviral insertions in the human population. *J.*  
520 *Viro.* 88, 9529–9537  
521  
522 26 Filée, J. (2014) Multiple occurrences of giant virus core genes acquired by eukaryotic ge-  
523 nomes: The visible part of the iceberg? *Virology* 466, 53–59  
524  
525 27 Geering, A.D. *et al.* (2014) Endogenous florendoviruses are major components of plant ge-  
526 nomes and hallmarks of virus evolution. *Nat. Commun.* 5, 5269  
527  
528 28 Maumus, F. *et al.* (2016) Study of gene trafficking between *Acanthamoeba* and giant viruses  
529 suggests an undiscovered family of amoeba-infecting viruses. *Genome Biol. Evol.* 8, 3351–3363  
530  
531 29 Parker, B.J. and Brisson, J.A. (2019) A laterally transferred viral gene modifies aphid wing  
532 plasticity. *Curr. Biol.* 29, 1–6  
533  
534 30 Gasmi, L. *et al.* (2015) Recurrent domestication by Lepidoptera of genes from their parasites  
535 mediated by bracoviruses. *PLoS Genet.* 11:e1005470  
536  
537 31 Hill, T. and Betancourt, A.J. (2018) Extensive exchange of transposable elements in the *Dro-*  
538 *sophila pseudoobscura* group. *Mobile DNA* 9, 20  
539  
540 32 Baidouri, M. *et al.* (2014) Widespread and frequent horizontal transfers of transposable ele-  
541 ments in plants. *Genome Res.* 24, 831–838  
542  
543 33 Peccoud, J. *et al.* (2017) Massive horizontal transfer of transposable elements in insects. *Proc.*  
544 *Natl. Acad. Sci. U.S.A.* 114, 4721–4726  
545  
546 34 Reiss, D. *et al.* (2019) Global survey of mobile DNA horizontal transfer in arthropods reveals  
547 Lepidoptera as a prime hotspot. *PLoS Genet.* 15, e1007965  
548

549 35 Ivancevic, A.M. et al. (2018) Horizontal transfer of BovB and L1 retrotransposons in eukary-  
550 otes. *Genome Biol.* 19, 85  
551

552 36 Gilbert, C. et al. (2014) Population genomics supports baculoviruses as vectors of horizontal  
553 transfer of insect transposons. *Nat. Commun.* 5, 3348  
554

555 37 Zhang, H. et al. (2018) Unexpected invasion of miniature inverted-repeat transposable ele-  
556 ments in viral genomes. *Mobile DNA* 9, 19  
557

558 38 O’Brochta, D.A. et al. (2009) Transpositionally active episomal hAT elements. *BMC Mol.*  
559 *Biol.* 10, 1–15  
560

561 39 Hedges, D.J. and Deininger, P.L. (2007) Inviting instability: Transposable elements, double-  
562 strand breaks, and the maintenance of genome integrity. *Mutat. Res.* 616, 46–59  
563

564 40 Frost, L.S. et al. (2005) Mobile genetic elements: the agents of open source evolution. *Nat.*  
565 *Rev. Microbiol.* 3, 722–732  
566

567 41 Schaack, S. et al. (2010) Promiscuous DNA: horizontal transfer of transposable elements and  
568 why it matters for eukaryotic evolution. *Trends Ecol. Evol.* 25, 537–546  
569

570 42 McDonald, M.C. et al. (2019) Transposon-mediated horizontal transfer of the host-specific  
571 virulence protein ToxA between three fungal wheat pathogens. *mBio* 10, e01515-19  
572

573 43 Sieber, K.B. et al. (2017) Lateral gene transfer between prokaryotes and eukaryotes. *Exp. Cell*  
574 *Res.* 358, 421–426  
575

576 44 Husnik, F. and McCutcheon, J.P. (2017) Functional horizontal gene transfer from bacteria to  
577 eukaryotes. *Nat. Rev. Microbiol.* 16, 67–79  
578

579 45 Martin, W. et al. (1993) Evidence for a chimeric nature of nuclear genomes: eubacterial origin  
580 of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl. Acad. Sci. U.S.A.*  
581 90, 8692–8696  
582

583 46 Timmis, J.N. et al. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic  
584 chromosomes. *Nat. Rev. Genet.* 5, 123–135  
585

586 47 Smith, D.R. et al. (2011) Correlation between nuclear plastid DNA abundance and plastid  
587 number supports the limited transfer window hypothesis. *Genome Biol. Evol.* 3, 365–371  
588

589 48 Ku, C. et al. (2015) Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited  
590 chimerism in eukaryotes. Proc. Natl. Acad. Sci. U.S.A. 112, 10139–10146  
591  
592 49 Archibald, J.M. (2015) Genomic perspectives on the birth and spread of plastids. Proc. Natl.  
593 Acad. Sci. U.S.A. 112, 10147–10153  
594  
595 50 Pittis, A.A. and Gabaldón, T. (2016) Late acquisition of mitochondria by a host with  
596 chimaeric prokaryotic ancestry. Nature 531, 101–104  
597  
598 51 Qiu, H. et al. (2013) Assessing the bacterial contribution to the plastid proteome. Trends Plant  
599 Sci. 18, 680–687  
600  
601 52 Nowack, E.C. et al. (2016) Gene transfers from diverse bacteria compensate for reductive  
602 genome evolution in the chromatophore of *Paulinella chromatophora*. Proc. Natl. Acad. Sci.  
603 U.S.A. 113, 12214–12219  
604  
605 53 Singer, A. et al. (2017) Massive protein import into the early-evolutionary-stage photosyn-  
606 thetic organelle of the amoeba *Paulinella chromatophora*. Curr. Biol. 27, 2763-2773  
607  
608 54 Husnik, F. et al. (2013) Horizontal gene transfer from diverse bacteria to an insect genome  
609 enables a tripartite nested mealybug symbiosis. Cell 153, 1567–1578  
610  
611 55 Bublitz, D.C. et al. (2019) Peptidoglycan production by an insect-bacterial mosaic. Cell 179,  
612 703-712  
613  
614 56 Kamp, A. et al. (2015) Nitrate storage and dissimilatory nitrate reduction by eukaryotic mi-  
615 crobes. Front. Microbiol. 6, 1492  
616  
617 57 Woehle, C. et al. (2018) A novel eukaryotic denitrification pathway in foraminifera. Curr.  
618 Biol. 28, 2536-2543  
619  
620 58 Hernandez, A.M. and Ryan, J.F. (2018) Horizontally transferred genes in the ctenophore  
621 *Mnemiopsis leidyi*. PeerJ 6, e5067  
622  
623 59 Doudoumis, V. et al. (2012) Tsetse-*Wolbachia* symbiosis: comes of age and has great poten-  
624 tial for pest and disease control. J. Invertebr. Pathol. 112, S94-S103  
625  
626 60 Doolittle, F.W. (1998) You are what you eat: a gene transfer ratchet could account for bacte-  
627 rial genes in eukaryotic nuclear genomes. Trends Genet. 14, 307-311  
628

629 61 Karnkowska, A. et al. (2016) A eukaryote without a mitochondrial organelle. *Curr. Biol.* 26,  
630 1274–1284  
631  
632 62 Keeling, P.J. and Palmer, J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat.*  
633 *Rev. Genet.* 9, 605–618  
634  
635 63 Andersson, J.O. et al. (2007) A genomic survey of the fish parasite *Spironucleus salmonicida*  
636 indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukary-  
637 ote genome evolution. *BMC Genomics* 8, 51  
638  
639 64 Alsmark, C. et al. (2013) Patterns of prokaryotic lateral gene transfers affecting parasitic mi-  
640 crobial eukaryotes. *Genome Biol.* 14, R19  
641  
642 65 Mayer, W.E. et al. (2011) Horizontal gene transfer of microbial cellulases into nematode ge-  
643 nomes is associated with functional assimilation and gene turnover. *BMC Evol. Biol.* 11, 13  
644  
645 66 Dean, P. et al. (2018) Transporter gene acquisition and innovation in the evolution of  
646 *Microsporidia* intracellular parasites. *Nat. Commun.* 9, 1709  
647  
648 67 Gilbert, C. et al. (2010) A role for host–parasite interactions in the horizontal transfer of  
649 transposons across phyla. *Nature* 464, 1347–1350  
650  
651 68 Quispe-Huamanquispe, D.G. et al. (2017) Horizontal gene transfer contributes to plant evolu-  
652 tion: the case of *Agrobacterium* T-DNAs. *Front. Plant Sci.* 8, 2015  
653  
654 69 Matveeva, T.V. et al. (2012) Horizontal gene transfer from genus *Agrobacterium* to the plant  
655 *Linaria* in nature. *Mol. Plant-Microbe Interact.* 25, 1542–1551  
656  
657 70 Kyndt, T. et al. (2015) The genome of cultivated sweet potato contains *Agrobacterium* T-  
658 DNAs with expressed genes: An example of a naturally transgenic food crop. *Proc. Natl. Acad.*  
659 *Sci. U.S.A.* 112, 5844–5849  
660  
661 71 Li, Z. and Bock, R. (2018) Replication of bacterial plasmids in the nucleus of the red alga  
662 *Porphyridium purpureum*. *Nat. Commun.* 9, 3451  
663  
664 72 Lee, J. et al. (2016) Reconstructing the complex evolutionary history of mobile plasmids in  
665 red algal genomes. *Sci. Rep.* 6, 23744  
666  
667 73 Bhattacharya, D. et al. (2013) Genome of the red alga *Porphyridium purpureum*. *Nat.*  
668 *Commun.* 4, 1941

669  
670 74 Rossoni, A.W. et al. (2019) The genomes of polyextremophilic cyanidiales contain 1% hori-  
671 zontally transferred genes with diverse adaptive functions. *eLife* 8, e45017  
672  
673 75 Karas, B.J. et al. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat.*  
674 *Commun.* 6, 6925  
675  
676 76 Bowler, C. et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of dia-  
677 tom genomes. *Nature* 456, 239–244  
678  
679 77 Heinemann, J.A. and Sprague, G.F. (1989) Bacterial conjugative plasmids mobilize DNA  
680 transfer between bacteria and yeast. *Nature* 340, 205–209  
681  
682 78 Nevoigt, E. et al. (2000) Cells of the yeast *Saccharomyces cerevisiae* are transformable by  
683 DNA under non-artificial conditions. *Yeast* 16, 1107–1110  
684  
685 79 Sibbald, S.J. and Archibald, J.M. (2017) More protist genomes needed. *Nat. Ecol. Evol.* 1,  
686 145  
687  
688 80 Ropars, J. et al. (2015) Adaptive horizontal gene transfers between multiple cheese-associated  
689 fungi. *Curr. Biol.* 25, 2562–2569  
690  
691 81 Richards, T.A. and Monier, A. (2016) A tale of two tardigrades. *Proc. Natl. Acad. Sci. U.S.A.*  
692 113, 4892–4894  
693  
694 82 McCarthy, C.G. and Fitzpatrick, D.A. (2019) Pan-genome analyses of model fungal species.  
695 *Microb. Genomics* DOI: 10.1099/mgen.0.000243  
696  
697 83 Wang, W. et al. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated  
698 rice. *Nature* 557, 43–49  
699  
700 84 Read, B.A. et al. (2013) Pan genome of the phytoplankton *Emiliana* underpins its global dis-  
701 tribution. *Nature* 499, 209–213  
702  
703 85 Worden, A.Z. et al. (2009) Green evolution and dynamic adaptations revealed by genomes of  
704 the marine picoeukaryotes *Micromonas*. *Science* 324, 268–272  
705  
706 86 McDonald, S.M. et al. (2010) The mixed lineage nature of nitrogen transport and assimilation  
707 in marine eukaryotic phytoplankton: a case study of *Micromonas*. *Mol. Biol. Evol.* 27, 2268–  
708 2283

709  
710 87 Savory, F. et al. (2018) Ancestral function and diversification of a horizontally acquired  
711 oomycete carboxylic acid transporter. *Mol. Biol. Evol.* 35, 1887–1900  
712  
713 88 Milner, D.S. et al. (2019) Environment-dependent fitness gains can be driven by horizontal  
714 gene transfer of transporter-encoding genes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5613–5622  
715  
716 89 Kominek, J. et al. (2019) Eukaryotic acquisition of a bacterial operon. *Cell* 176, 1356–1366  
717  
718 90 Dunning, L.T. et al. (2019) Lateral transfers of large DNA fragments spread functional genes  
719 among grasses. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4416–4425  
720  
721 91 McVey, M. and Lee, S. (2008) MMEJ repair of double-strand breaks (director’s cut): deleted  
722 sequences and alternative endings. *Trends Genet.* 24, 529–538  
723  
724 92 Huang, J. (2013) Horizontal gene transfer in eukaryotes: The weak-link model. *BioEssays* 35,  
725 868–875  
726  
727 93 Matsuo, M. et al. (2018) Characterization of spliced leader trans-splicing in a photosynthetic  
728 rhizarian amoeba, *Paulinella micropora*, and its possible role in functional gene transfer. *PLoS*  
729 *One* 13, e0200961  
730  
731 94 Doolittle, W. et al. (2003) How big is the iceberg of which organellar genes in nuclear ge-  
732 nomes are but the tip? *Philos. Trans. R. Soc. London, Ser. B* 358, 39–58  
733  
734 95 Vembar, S. et al. (2016) Complete telomere-to-telomere de novo assembly of the *Plasmodium*  
735 *falciparum* genome through long-read (>11 kb), single molecule, real-time sequencing. *DNA*  
736 *Res.* 23, 339–351  
737  
738 96 Funkhouser-Jones, L.J. et al. (2015) *Wolbachia* co-infection in a hybrid zone: discovery of  
739 horizontal gene transfers from two *Wolbachia* supergroups into an animal genome. *PeerJ* 3,  
740 e1479  
741  
742 97 Brelsfoard, C. et al. (2014) Presence of extensive *Wolbachia* symbiont insertions discovered  
743 in the genome of its host *Glossina morsitans morsitans*. *PLoS Negl. Trop. Dis.* 8, e2728  
744  
745 98 Pinheiro, J. et al. (2018) The protozoan *Trichomonas vaginalis* targets bacteria with laterally  
746 acquired NlpC/P60 peptidoglycan hydrolases. *mBio* 9, e01784-18  
747  
748 99 Pereira-Neves, A. and Benchimol, M. (2007) Phagocytosis by *Trichomonas vaginalis*: new  
749 insights. *Biol. Cell* 99, 87–101  
750  
751 100 Liu, L. et al. (2012) The human microbiome: A hot spot of microbial horizontal gene trans-

752 fer. Genomics 100, 265–270

753

754 101 Parfrey, L. et al. (2011) Microbial eukaryotes in the human microbiome: ecology, evolution,  
755 and future directions. Front. Microbiol. 2, 153

756

757

758

759

760

761

## 762 **Acknowledgements**

763 We thank the reviewers for a number of constructive suggestions and apologize to authors whose  
764 work could not be cited due to length constraints. SS was supported by a CGS-D scholarship  
765 from the Natural Sciences and Engineering Research Council of Canada. The work was also  
766 supported by a Foundation Grant, FRN-142349, from the Canadian Institutes of Health Research  
767 awarded to AJR, a Symbiosis Grant from the Moore Foundation awarded to JMA and LE is sup-  
768 ported by funding from the European Research Council (ERC) under the European Union's  
769 Horizon 2020 research and innovation programme (grant agreement No 803151).

770

## 771 **Glossary**

### 772 **Transduction**

773 Virus-mediated transfer of genetic material. In bacteria, generalized transduction involves the  
774 integration and transfer of random pieces of bacterial DNA into the viral genome, while in spe-  
775 cialized transduction only defined regions of the donor genome can be packaged into the virus  
776 and transferred.

777

778 **Transformation**

779 Uptake of exogenous DNA by a cell. This can occur naturally or in an experimental context.

780 Both prokaryotic and eukaryotic cells are capable of being transformed.

781

782 **Conjugation**

783 Transfer of genetic material mediated by cell-cell contact; single-stranded DNA moves through a  
784 thin connection between the donor and recipient cells. Although best known to occur between  
785 bacteria, artificial conjugation between bacteria and eukaryotes such as diatoms has been carried  
786 out.

787

788 **Pan-genome**

789 The term refers to the entire set of genes in the genomes of members of a given species. The  
790 word was invented to underscore the fact that individuals or strains of a given species can vary in  
791 their gene content. In bacterial and archaeal species genome content variability can be immense  
792 such that the set of genes found in all members of a given species (the 'core') can represent a  
793 small fraction of the total pan-genome. This variability in gene content has a number of causes  
794 including gene duplications, gene losses, *de novo* gene formation and the acquisition of new  
795 genes by lateral gene transfer (LGT).

796

797 **Protists**

798 Protists are any eukaryotic organisms that are not animals, plants or fungi. Protists are an ances-  
799 tral 'grade' of eukaryotic organization and encompass the bulk of eukaryotic diversity. Although

800 most protist species are unicellular microbes, charismatic multicellular forms are well known  
801 (e.g., macroalgae like giant kelp).

802

### 803 **Endosymbiont**

804 An organism that thrives within another organism, typically without killing its host. Endosym-  
805 biosis of bacteria, archaea and microbial eukaryotes (e.g., algae) is relatively common in free-  
806 living protists, but more rare in parasites. Organelles such as mitochondria and chloroplasts  
807 evolved from ancient bacterial endosymbionts. Readers should keep in mind that (en-  
808 do)symbiosis encompasses a continuum from mutualism to parasitism, and determining where a  
809 given symbiotic system lies along this continuum can be very difficult.

810

### 811 **Endosymbiotic gene transfer (EGT)**

812 The transfer of genetic material from endosymbionts and their organellar derivatives to the host  
813 cell nuclear genome. This process involves endosymbiont / organelle lysis, migration of released  
814 DNA into the nucleus, and integration of random pieces of foreign DNA into nuclear chromo-  
815 somes by DNA repair enzymes.

816

817 **Box 1: Robust examples of LGT in eukaryotes**

818           This review focuses on a select few of the numerous reports of LGT in eukaryotes. A  
819 recent example of adaptive LGT concerns the acquisition of the gene *rquA*, which is involved in  
820 rhodoquinone (RQ) biosynthesis, in various anaerobic protistan lineages [11]. The RQ produced  
821 by RquA allows for fumarate to be used as an alternative terminal electron acceptor during mito-  
822 chondrial ATP production under anaerobic conditions, allowing these protists to adapt to low  
823 oxygen conditions. Phylogenetic analyses suggest that *rquA* was transferred on at least two sepa-  
824 rate occasions from bacteria to eukaryotes, followed by multiple eukaryote-to-eukaryote trans-  
825 fers.

826           Single gene-encoded transporters may be especially amenable to LGT as they could pro-  
827 vide a direct selective advantage to the recipient by allowing for the uptake and utilization of a  
828 particular metabolite. Numerous examples of LGTs involving transporter genes have been re-  
829 ported (e.g., [68,86-88 and references therein]). Milner et al. [88] identified seven cases of trans-  
830 porter-gene LGT between fungal phyla and demonstrated that they conferred a competitive ad-  
831 vantage in certain environments. They proposed a “transporter gene acquisition ratchet” where  
832 the range of transporters encoded in a given genome are continually being altered by various  
833 mechanisms, including LGT.

834           Another notable report of LGT is the acquisition of an entire operon involved in  
835 siderophore biosynthesis from Enterobacteriaceae into a group of closely related yeasts [89]. The  
836 co-linearly arranged genes were found to be expressed and functional (via transcriptomics, meta-  
837 bolic assays, and mass spectrometry), allowing for the sequestering and uptake of iron from the  
838 environment – a likely beneficial feature as the yeast reside in the iron-limited insect gut envi-  
839 ronment co-inhabited with numerous Enterobacteriaceae. After the initial LGT, the operon struc-

840 ture appears to have undergone a ‘eukaryotification’ process that included the addition of  
841 polyadenylation sites, various structural rearrangements, addition of some eukaryotic genes, and  
842 loss of some genes encoded by the operon in closely related yeast lineages. In principle, acquir-  
843 ing an entire operon in a single LGT would allow for complex multi-gene pathways to spread  
844 rapidly and effectively from donor to recipient.

845         Analyses of chromosome-level genomes for the grass *Alloteropsis semialata* also re-  
846 vealed recent acquisitions of chunks of foreign DNA encoding multiple genes, including genes  
847 related to disease resistance, abiotic stress tolerance and C4 photosynthesis [90]. Phylogenetic  
848 analyses including genomes from >140 other grass species led to the discovery of 59 unambigu-  
849 ous, recent LGTs into *A. semialata* from at least nine distantly related donor grasses. These genes  
850 were found to be mostly expressed and clustered in 23 transferred fragments up to 170 kbp long,  
851 indicating that large stretches of DNA containing multiple functional genes as well as non-  
852 coding regions were transferred in a single event. In some cases, TEs were also encoded on the  
853 transferred region.

854

## 855 **Box 2: Factors influencing the frequency of lateral gene transfer**

856         The frequency of gene transfer into a given species may hinge on a number of genomic  
857 and environmental factors. For example, non-homologous end joining repair could be a major  
858 mechanism of incorporation of foreign DNA [44] and thus impact the frequency of LGT in a  
859 given genome. It is generally thought that successful LGTs can occur more frequently between  
860 closely related organisms due to similarities between their genomes (codon usage, cis-acting  
861 regulatory elements, etc.); high levels of sequence identity also foster homologous recombina-  
862 tion, which can lead to the transfer of alleles from one species to another [1]. Genetic exchange,

863 however, is not completely limited to very similar sequences. As long as there are conserved  
864 flanking regions, divergent alleles can recombine by homologous recombination [1] and the  
865 presence of small short repeats may allow for micro-homology mediated recombination as well  
866 [91].

867         Dynamic and repetitive regions of the genome that are more prone to duplication are like-  
868 ly to have increased rates of gene integration. LGTs have been observed in regions enriched in  
869 DNA transposons and retrotransposons and in gene-poor regions such as telomeres or in/near  
870 TEs (discussed in [43]). Additionally, selfish genetic elements (e.g., self-splicing introns, inteins  
871 etc.) are common in multicellular eukaryotes and could promote LGT and genome rearrange-  
872 ments [1]. ‘Fragile’ genomic regions that are prone to double stranded breaks or environmentally  
873 induced DNA damage and repair could also create opportunities for foreign gene integration.  
874 LGTs may occur when viruses or parasites manipulate DNA damage/repair pathways. Certain  
875 life/developmental stages may be more likely entry points for foreign genes into the nuclei of  
876 both unicellular eukaryotes (particularly those who are not phagotrophic), and in particular, the  
877 germline of multicellular eukaryotes (see the ‘weak-link’ model [92]).

878         In the testate amoeba *Paulinella micropora*, Matsuo et al. [93] identified a common ~30  
879 bp sequence added post-transcriptionally by spliced leader (SL) trans-splicing to a diverse set of  
880 mRNAs, including those derived by LGT and EGT. Based on these observations, the authors  
881 suggested that SL trans-splicing could mitigate the potential detrimental effects of LGT in eu-  
882 karyotes by ensuring translation of native nuclear genes disrupted by the insertion of foreign  
883 genes in their 5’UTRs. This could in principle apply not only to *Paulinella* species but to other  
884 trans-splicing organisms including euglenids, dinoflagellates and some metazoans.

885

886 **Box 3: Detecting LGT with confidence**

887 Genomic features that differ between eukaryotic and prokaryotic genes can be used to  
888 differentiate between bona fide LGT and prokaryotic contaminants. One useful distinction is the  
889 presence of introns in eukaryotic genes and lack thereof in prokaryotes; if a candidate LGT of  
890 bacterial origin contains an intron it most likely resides within a eukaryote genome. Other char-  
891 acteristics such as nucleotide composition, amino acid usage and the presence of a poly-A tail on  
892 transcripts can help identify genuine LGTs. Further support can be found by examining the ge-  
893 nomes of closely related organisms; unless the transfer is very recent, the suspected LGT will  
894 likely be found in multiple genomes. For example, most of the LGTs identified in the gut para-  
895 site *Blastocystis* sp. subtype 1 were also found in genomes of seven of the other subtypes [13].

896 Phylogenetic analysis is essential for differentiating candidate LGTs from gene duplica-  
897 tions and differential loss. A patchy gene distribution, combined with a phylogenetic relationship  
898 that contradicts the species relationship, is suggestive of an LGT. If it is assumed that all patchily  
899 distributed eukaryotic genes are due to differential gene loss rather than lateral acquisitions, then  
900 there becomes a genome of Eden problem [94] – the unlikely scenario that the last eukaryotic  
901 common ancestor (LECA) had almost every gene seen in modern eukaryotes and an enormous  
902 range of metabolic capabilities. Phylogenetic reconstruction of LGT is not without its limitations,  
903 with certain transfers being more difficult to detect and interpret (such as sequential transfers,  
904 which results in more complex gene distributions [62]). The lack of resolution associated with  
905 single gene trees can make it tough to infer the validity and origin of candidate LGTs. Phyloge-  
906 netic artifacts, leading to falsely resolved relationships that may mislead interpretations of a  
907 gene's evolutionary history, are also of concern.

908           Some of the other most conclusive bioinformatic evidence for LGT can come from inves-  
909           tigations of genomic context. A candidate LGT contained within a larger genomic scaffold clear-  
910           ly of host origin is likely to be real (e.g., viral LGTs in *Acanthamoeba* that are surrounded by  
911           amoebal host genes [28]). This can be difficult to assess if a reference genome is poorly assem-  
912           bled and highly fragmented. Fortunately, long read sequencing technologies make it possible to  
913           produce more highly contiguous genome assemblies, allowing for greater insight into genomic  
914           context, particularly in repetitive regions and near telomeric repeats that typically do not assem-  
915           ble well (e.g., [95]). Classical molecular biology techniques can also be used to verify genomic  
916           integration of candidate LGTs into host chromosomes (e.g., fluorescence in situ hybridization  
917           (FISH) has been used to confirm insertions of *Wolbachia* DNA on multiple chromosomes in the  
918           grasshopper *Chorthippus parallelus* and Tsetse flies (*Glossina* spp.) [96, 97].

919

#### 920 **Box 4: *Trichomonas* and gene transfer in the microbiome**

921           *Trichomonas vaginalis* is a sexually transmitted parasite that colonizes an environment  
922           occupied by other parasites and bacteria. 149 LGTs of both prokaryotic and eukaryotic origin  
923           were identified in the *T. vaginalis* genome, including genes likely facilitating pathogenicity and  
924           to take advantage of the rich metabolic opportunities in its environment (e.g., glycan degrada-  
925           tion) [64]. Further study on LGTs in *T. vaginalis* identified and confirmed the functionality of  
926           NIPC/P60 peptidoglycan hydrolases that were acquired laterally from bacterial co-members of  
927           the microbiota [98]. Pinheiro et al. [98] showed that the bacterial NIPC/P60 hydrolases in the  
928           genome of *T. vaginalis* were capable of degrading bacterial cell envelopes and that their expres-  
929           sion was significantly upregulated in the presence of bacteria. The transfer of NIPC/P60 hydro-  
930           lases in *T. vaginalis* appears to be critical in controlling bacterial abundance in its environment

931 and the establishment of a successful infection. This example shows that co-inhabitants of a giv-  
932 en microbiome can act as a source of genes for one another, particularly in cases of phagocytic  
933 species like *T. vaginalis* [99].

934         The human microbiome has been referred to as “a hotspot of microbial horizontal gene  
935 transfer” [100]. While studies on LGT in the human microbiome has almost entirely focused on  
936 the extensive prokaryote-prokaryote transfers (e.g., 13,514 high confidence LGTs between 308  
937 human microbiome bacteria [100]), diverse eukaryotes with lifestyles ranging from parasitic to  
938 beneficial also reside in and are important components of the human microbiome [101]. The  
939 abundance of a variety of species and phages in close proximity presumably creates a favorable  
940 environment for LGT for microbial eukaryotes; by increasing the frequency of donor DNA in the  
941 environment you increase the exposure of a species to a given DNA source [13,44,64].

942 **Figure 1. Possible sources of foreign DNA contributing to LGT in eukaryotes.** Eukaryotic  
943 cells are exposed to many potential sources of exogenous DNA including from: i) lysed  
944 endosymbiont-derived organelles (mitochondria and chloroplasts), prokaryotic or eukaryotic  
945 endosymbionts, ii) viral infection, iii) bacterial conjugation (e.g., *Agrobacterium tumefaciens*)  
946 and iv) lysed phagosomes containing food organisms.

947

948 **Figure 2. The pan-genome concept.** The pan-genome consists of all genes found in all strains  
949 of a given species and is depicted at the top of the figure as the union of all ellipses (each ellipse  
950 is the set of genes present in that strain) in the Venn diagram. The pan-genome can be further  
951 broken down into the ‘core’ genome (genes found in all strains shown in dark teal) and the ‘ac-  
952 cessory’ genome (shown colours ranging from green to blue) which are genes found in a single  
953 strain (i.e. strain specific) or a subset of strains (i.e. strain select).