Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes

Short title: **Horizontal gene transfer of fungal transporters**

Classification: **Biological Sciences; Evolution**

David S. Milner[1]*, Victoria Attah[1], Emily Cook[1], Finlay Maguire[1,2], Fiona Savory[1], Mark Morrison[1], Carolin A. Müller[3], Peter G. Foster[4], Nicholas J. Talbot[1,5], Guy Leonard[1], Thomas A. Richards[1]*

[1] Biosciences, Living Systems Institute, University of Exeter, Exeter, United Kingdom

[2] Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada

[3] The Sir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom

[4] Department of Life Sciences, Natural History Museum, London, United Kingdom

[5] Current Address: The Sainsbury Laboratory, Norwich Research Park, NR4 7UH, UK.

*Corresponding author: T.A.Richards@exeter.ac.uk and D.Milner@exeter.ac.uk

Abstract

Many microbes acquire metabolites in a 'feeding' process where complex polymers are broken down in the environment to their subunits. The subsequent uptake of soluble metabolites by a cell, sometimes called osmotrophy, is facilitated by transporter proteins. As such, the diversification of osmotrophic micro-organisms is closely tied to the diversification of transporter functions. Horizontal gene transfer (HGT) has been suggested to produce genetic variation that can lead to adaptation, allowing lineages to acquire traits and expand niche ranges. Transporter genes often encode single-gene phenotypes and tend to have low protein-protein interaction complexity and, as such, are potential candidates for HGT. Here we test the idea that HGT has underpinned the expansion of metabolic potential and substrate utilisation via transfer of transporter-encoding genes. Using phylogenomics, we identify seven cases of transporter-gene HGT between fungal phyla, and investigate compatibility, localisation, function and fitness consequences when these genes are expressed in *Saccharomyces cerevisiae*. Using this approach, we demonstrate that the transporters identified can alter how fungi utilise a range of metabolites, including peptides, polyols and sugars. We then show, for one model gene, that transporter gene acquisition by HGT can significantly alter the fitness landscape of *S. cerevisiae*. We therefore provide evidence that transporter HGT occurs between fungi, alters how fungi can acquire metabolites and can drive gain in fitness. We propose a 'transporter-gene acquisition ratchet', where transporter repertoires are continually augmented by duplication, HGT and differential loss, collectively acting to overwrite, fine-tune and diversify the complement of transporters present in a genome.

Keywords: **Lateral gene transfer | Protein connectivity | Osmotrophy/lysotrophy | Gene ratchet**

Significance Statement

Horizontal gene transfer (HGT) is the transfer of genetic information between genomes by a route other than 'vertical' inheritance. Of particular interest here are the transfers of transporter-encoding genes, which can allow an organism to utilise a new metabolite, often via the acquisition of a single foreign gene. Here we have identified a range of HGT events of transporter-encoding genes, characterised the substrate preferences for each HGT encoded protein, and demonstrated that the gain of one of these HGTs can provide yeast with a distinct competitive advantage in a given environment. This has wide implications for understanding how acquisition of single genes by HGT can drastically influence the environments fungi can colonise.

\body

Introduction

Horizontal, or lateral, gene transfer (HGT) is the movement of genetic material between reproductively isolated genomes, and operates in contrast to vertical inheritance, where genes are passed from parent to progeny. HGT events are common in prokaryotes (1, 2) and, although less common, it is becoming apparent that HGT events into eukaryotes are a significant factor (3) and contribute towards the evolution of eukaryotic microbes including oomycetes (4-8) and fungi (9-12). Previous work has demonstrated that HGT events in eukaryotes produce genetic variation that can lead to adaptation, allowing lineages to utilise alternative nutrients, colonise new environments and expand niche ranges (3, 10, 13-15). Furthermore, experimental characterisation of the functions of horizontally acquired gene products (12, 16-24) has reinforced the hypothesis that HGT can drive adaptive potential in eukaryotes.

Transporter genes are potential candidates for acquisition by HGT as they constitute single-gene phenotypes (i.e. a single gene that encodes an entire functional trait), with the proteins they

encode tending to have low protein-protein interaction complexity (12, 25). Consistent with this hypothesis, previous work has demonstrated that phenotypes can be acquired by single gene HGTs (e.g. (23, 26, 27)), and that HGTs tend to encode proteins which have a low number of protein-protein interactions (28-30). Furthermore, membrane transporter HGT events in eukaryotes have been described previously (5, 19, 31, 32), although relatively little work has been conducted to investigate the functional impact of HGT of transporter genes across fungal phyla and the subsequent consequences of gene acquisition on the fitness of recipient lineages. This analysis would, in turn, allow the examination of how differing environmental conditions can dictate whether HGT acquired genes are maintained by selection within lineages, informing our understanding of pan-genome evolution in eukaryotic microbes such as fungi.

Fungi are obligate osmotrophs and often perform a range of lysotrophic functions (33), feeding by secreting extracellular enzymes into the environment, degrading complex molecules, and importing the resulting nutrients into the cell (25). Transporter proteins are central to this process, allowing the uptake of simple monomers following extracellular enzymatic degradation. As such, the gain (e.g. by HGT and/or duplication) and loss of transporter proteins in fungi is hypothesised to be a key evolutionary process in driving niche range and influencing both social interactions and colonisation of new environments (25). We therefore sought to identify transporter gene HGT events in fungi, and examine the types of functional constraints such HGTs must overcome in order to become fixed within a recipient lineage. Here, we have identified seven instances of transporter gene HGT between fungal phyla, characterized the function of each using phenotype microarrays and/or genetic complementation, and examined the fitness effects of one model HGT event using heterologous expression in *S. cerevisiae*. These results show that single gene transfer of transporter-encoding genes can drive environment-dependant fitness traits, demonstrating that gain and loss of 'bolt-on-traits' can drive pan-genome evolution in fungi.

Results and Discussion

Identification of transporter HGTs between ascomycetes and basidiomycetes

Using phylogenomic approaches we sought to identify transporter proteins for which phylogenetic analysis identifies a tree topology consistent with recent HGT between the ascomycetes and basidiomycetes. To identify candidate HGTs, we generated a phylogeny for every protein annotated as a transporter in the *S. cerevisiae* S288c genome (34) using an automated pipeline (35). Putative HGT phylogenies were identified as tree topologies that demonstrated genes from a closely related group, the recipient, branching within a clade of distantly-related taxa, the donor group (36). Amino acid sequence alignments of candidate HGTs were then subject to additional taxon sampling checks, alignment editing and site masking, Maximum Likelihood (ML) phylogenetic reconstruction and Bayes factor alternative topology tests. We chose to focus on identifying HGTs between ascomycete and basidiomycete taxa because genome sampling for these sister groups is relatively dense and they constitute established clades (37). This process identified seven protein sequence phylogenies where the tree topology is consistent with an Ascomycota-Basidiomycota HGT. Advanced genome sampling also allowed us to contrast gene synteny relationships across taxa and investigate whether candidate HGT genes are adjacent to genes of vertical ancestry, ruling out genome project contamination as an alternative explanation for the identified phylogenetic relationship in all seven cases (File S1). In three cases, genome sampling of recently released data allowed for further phylogenetic analysis as additional putative recipient taxa were identified (File S2). In total, five of the transfers (HGT-1, -2, -3, -4, & -7) identified were detected in two or more recipient genomes (File S1 & S2), demonstrating independent sampling points of the HGT gene and providing further support that these phylogenetic relationships were not the product of genome project contamination. This additional sampling identified two further ascomycete-basidiomycete HGTs in the wider HGT-3 and HGT-6 transporter gene families (see File S2).

The seven HGTs identified were classified into a range of different transporter protein domain super-families including: Sugar_tr - Sugar transporters (Pfam00083: HGT-1 & HGT-4), LysP Amino Acid permeases (COG0833: HGT-2 & HGT-5), MFS - Major Facilitator Super family (Pfam00854: HGT-3), UhpC - sugar phosphate permease (cl27298: HGT-6) and OPT - oligopeptide transporter (Pfam03169: HGT-7). The phylogenies identified one basidiomycete-to-ascomycete HGT (HGT-5) and six ascomycete-to-basidiomycete HGTs (see Table 1, Fig. 1 for summary, and File S1/S2 for phylogenetic data).

Investigation of transporter HGT integration within a model cell network

We propose that fixation of a transporter gene acquired by HGT must overcome three functional hurdles. First, the protein must not be toxic or incompatible with the recipient genetic background or cellular environment (i.e. the protein interaction network within the host cell). Second, the HGT encoded protein must occupy the correct subcellular location to allow function and finally, the transferred gene must confer a compatible function, upon which selection can act, that is either neutral or beneficial to host fitness. Using *S. cerevisiae* as an expression chassis, we therefore investigated these functional constraints. First, we assessed growth of *S. cerevisiae* strains, each expressing an HGT transporter gene under control of the constitutive GPD (TDH3) promoter and with a C-terminal His-tag (Fig. S1). For two strains, HGT-2 and HGT-5, we were unable to detect protein expression by Western blot (Fig. S2), putatively demonstrating that expression of the two transporters is detrimental to *S. cerevisiae* under this transcriptional load. This was consistent with poor growth rate (*r*) of the two strains (Fig. 2J; Fig. S1), both of which had an *r* median value of <55% of the wild-type *S. cerevisiae* strain containing only the empty p426-GPD vector. One reason for such poor growth may be the unregulated uptake of toxic levels of substrate into the cell. The uptake of large quantities of single amino acids, for example, has been shown to be toxic to *S. cerevisiae* (38). Both HGT-2 and HGT-5 show homology to LysP amino acid permeases (Table 1), suggesting that altered amino acid

uptake may be the cause of poor growth rate when expressed in *S. cerevisiae*, and demonstrating that transferred transporters can be incompatible with foreign cellular environments, and thus represent a potential barrier to transporter HGT in certain environment, genetic and cellular contexts.

Previous work has shown that functional HGTs tend to encode proteins exhibiting a low number of protein-protein interactions (28-30). Furthermore, transporter proteins show similar levels of protein-protein interaction complexity as previously identified bacterially-derived HGTs present on the *S. cerevisiae* genome (Fig. 2A-B & (25)). Indeed, both transporter proteins and the bacterial HGT-acquired proteins show significantly lower connectivity than the remainder of the *S. cerevisiae* proteome (t-test; both $p < 0.001$; normality tested using a Kolmogorov-Smirnov test) (Fig. 2B). To examine the projected protein-protein interaction complexity of each HGT transporter within the *S. cerevisiae* proteome, we used protein-protein interaction data (39) (based on BioGRID analysis (40)) to identify the number of interactions for each native *S. cerevisiae* transporter protein which shows shared amino acid sequence identity (>20%) (Fig. 2C-I) to the HGT transporters. These data demonstrated that HGT-2 and HGT-5 putative LysP encoding proteins, which are the proteins where retardation of growth is at the highest level and where production of the transporter cannot be confirmed via Western blot analysis, also have the highest number of projected protein-protein interactions within the yeast interactome (Fig. 2D, G, J; Fig. S2). It is currently not possible to identify a significant correlation between toxicity (i.e. retardation of growth) and interaction complexity given the current sample size (Fig. 2K), but we suggest both factors are likely to be important in determining compatibility of transporters once transferred into a foreign cellular environment. The remaining five heterologous expressed transporters share comparatively reduced projected protein-protein interaction complexity (Fig. 2C, E, F, H, I, J); this pattern is consistent with the concept that HGTs are functionally viable when the protein-protein interaction complexity of the HGT protein in the recipient cell is low (28, 29).

The five transporter proteins for which we observed protein expression, HGT-1, 3, 4 and 6, were tagged with C-terminal superfolder GFP (sfGFP), and HGT-7 with N-terminal enhanced GFP (EGFP), to assess protein localization in *S. cerevisiae*. HGT-7 was tagged at the N-terminus because PSORTII[47] analysis identified a putative C-terminal PTS1/SKL motif. HGT-1, 3, 4, and 6 showed a majority localisation to the cell surface plasma membrane (Fig. 3A). The *Ustilago maydis* HGT-7 transporter showed limited plasma membrane localisation and punctate internal cellular localisation. N-terminal EGFP intracellular localization of HGT-7 was then assessed by co-localization with RFP-tagged endomembrane proteins in a library of *S. cerevisiae* strains (41). We observed evidence of co-localization with Sec13, suggesting a heterologous ER and Golgi vesicular interface localisation for HGT-7 (Fig. 3B). This may be due to incomplete trafficking as a result of the addition of the EGFP tag, or some unknown characteristic of the *U. maydis* derived gene-product, yet the *S. cerevisiae* oligopeptide transporter 2 (OPT2), which shares 25% amino acid identity with HGT-7, has previously been shown to interact with vesicles that cycle between the late Golgi and the plasma membrane (42), consistent with a heterologous vesicular function of HGT-7. These data demonstrate *in vivo* that localisation of the four plasma-membrane proteins, and possibly the vesicular protein, is consistent with their predicted function as transporters. This provides support for the hypothesis that these proteins maintain a useful cellular address when moved into heterologous systems, i.e. in a scenario analogous to that produced by HGT.

Functional characterisation of transporter HGTs

To identify possible substrates for each HGT transporter we used the OmniLog® Phenotype Microarray (PM) system to compare growth and respiration of yeast strains expressing the transporter proteins across a range of different culture conditions. The OmniLog® data demonstrate that HGT-3, -6 and -7 transport various dipeptides and HGT-4 transports hexose sugars (File S3; all substrates detailed in Table 1). To test functional predictions based on homology searches, to validate the OmniLog® data

for HGT-3 and -4, and to investigate the possible substrate for HGT-1, a range of *S. cerevisiae* deletion mutants were then used for complementation assays (Table 1). We could not identify a putative compatible mutant for HGT-6 (a putative Sugar Phosphate transporter domain containing protein – Table 1) and due to the internal membrane localisation of HGT-7 we did not carry this transporter forward for complementation analysis. Briefly, these experiments identified: *i*) HGT-1 is a glycerol transporter, a poor primary carbon source for growth in yeast, *ii*) HGT-3 mediates uptake of 19 dipeptide substrates when expressed in the *ptr2* deletion strain (*PTR2* encodes a transporter known to facilitate dipeptide uptake (43)) of which 12 substrates were also identified in the OmniLog® PM screen of strain BY4742 (File S3), *iii*) HGT-4 uptakes the hexoses mannose, glucose, galactose and the pentose xylose (Fig. 3C and D), with the latter a poor primary carbon source for growth of many *S. cerevisiae* strains (44).

The hexose/pentose transporter HGT-4 exhibits key differences in amino acid sequence from *S. cerevisiae* Gal2 (the top HGT-4 BLASTP hit in *S. cerevisiae* (42% amino acid identity)) (Fig. 3F). A similar pattern of amino acid substitutions, primarily in the loop between transmembrane (TM) domains 6 and 7, were also identified as important for defining substrate range in directed evolution experiments, where mutagenesis of Gal2 identified sequence variants that resulted in enhanced xylose uptake (45). The relatively poor level of galactose uptake facilitated by HGT-4 (Fig. 3C-D) is also in agreement with Kasahara *et al.* (46), who demonstrated that Tyr446, located within TM10 of the Gal2 protein, is essential for galactose recognition and that substitution of this residue with Phe (as observed in the same region of HGT-4; Fig. 3F) results in glucose uptake. Consistent with these amino acid characteristics, HGT-4 increases *S. cerevisiae* growth rate/respiration with xylose as the sole carbon source (Fig. 3C-D) demonstrating a dramatic phenotype difference and alteration in nutrient/environment growth dynamic with a single gene acquisition. Interestingly, the recipients of HGT-4 are Gloeophyllaceae, brown rot fungi (Fig. 1 & File S1) which play a dominant role in

communities that function to decompose wood, and xylose is the primary constituent sugar monomer found in hardwood (47). Collectively the OmniLog® and complementation data identify substrates of five heterologously expressed HGT transporter proteins (Table 1), confirming that the putative transporters function to obtain metabolites that are compatible with many fungal cellular metabolic functions (15, 48). However, both approaches are likely to underestimate the full substrate range of these transporter proteins.

Testing environmental/metabolic fitness parameters for a model HGT

Both HGT-1 and HGT-4 significantly improve how yeast can grow on 'non-favoured' carbon sources, glycerol and xylose, respectively (45, 49). This demonstrates that single HGT acquisitions when 'recapitulated' in yeast can alter growth, a proxy for fitness, across non-standard yeast culture conditions, a proxy for altered niche space. To explore this possibility, we selected the HGT-1 transporter as a model HGT acquisition to investigate environment-specific gain of fitness outcomes. We performed competitive fitness assays by culturing a GFP-labelled *S. cerevisiae* strain expressing HGT-1 with a BFP-labelled strain containing an empty p426-GPD vector or a BFP-labelled strain expressing a native glycerol transporter, *stl1* (50) (Fig. S6). Stl1 was selected, as this native protein sequence shared the highest level of amino acid identity with HGT-1 (39% protein identity with no other putative homologues showing similar levels of sequence identity Fig. 2C). The *stl1* expressing strain was investigated to determine if any environment-specific gain of fitness outcomes due to HGT-1 acquisition could be achieved by comparable transcriptional rewiring and altered protein dosage of the native glycerol transporter. Quantitative Western blot analysis demonstrated that Stl1p and HGT-1 protein expression was equivalent (Fig. S7). Strains were co-cultured in media with either glycerol or glucose as the sole carbon-source, or an equal mix (w/v) of both glycerol and glucose. For each strain, relative abundance was monitored over time using flow cytometry and a measurement at the 72-hour timepoint was used to determine the relative population dominance as a proxy for fitness.

A significant growth advantage was observed for the strain expressing HGT-1 when incubated in 1-5% glycerol with either the empty p426-GPD or p426-GPD Stl1 strain (all $p < 0.001$, generalized linear model with binomial error distribution) (Fig. 4A). This demonstrates an environment-specific gain of fitness which is not simply the product of altered transcription of a glycerol transporter gene. In the presence of 0.1% glycerol, the HGT-1 expressing strain exhibited similar 'fitness' to the empty p426-GPD vector strain, but was at a disadvantage when co-cultured with the p426-GPD Stl1 strain (all $p < 0.001$, Fig. 4A). When cells were grown in glucose or in a glycerol/glucose mix, or when the carbon source was switched every 12 hours, there was a significant disadvantage for those expressing HGT-1 relative to those containing only the empty p426-GPD vector ($p < 0.001$ at all substrate concentrations tested) (Fig. 4B-D). This is likely to be a consequence of the cost of expressing the HGT-1 protein (Fig. 2J) and points towards a requirement for environmental-specific selection for fixation of the HGT transporter under a given promoter and, therefore, specific transcriptional load. It could also, however, be the result of differences in transcription because the constitutive GPD promoter may behave differently under each nutrient condition. Nevertheless, overexpression of HGT-1 was advantageous relative to overexpression of Stl1 (under the same promoter) when cells were grown in glucose or a glycerol/glucose mix and when the carbon source was switched every 12 hours ($p < 0.001$ at all concentrations tested; Fig. 4B-D), suggesting a greater cost of Stl1 expression in non-substrate and mixed substrate environments (confirmation that glucose is not a substrate of HGT-1 or Stl1 is shown in Fig. S8). Interestingly, radiolabelled glycerol accumulation assays demonstrated a higher uptake rate of glycerol by the Stl1 transporter compared to the HGT-1 transporter (Fig. 4E), suggesting that the advantage of HGT-1 in 1-5% glycerol (Fig. 4A) is a product of the gain of a transporter with a moderate substrate uptake rate (Fig. 4E) compared to the resident xenologous transporter.

Conclusion: transporter gene repertoires are subject to a gene acquisition/loss ratchet

Collectively, these results illustrate scenarios where transporter proteins are gained, depending on a chance gene acquisition by HGT, followed by environmental selection to drive fixation of the newly acquired trait. By contrast, absence of selection results in loss of both native transporters and/or HGT acquired transporters by drift. The data presented here show that fixation of a transporter HGT is not just a property of altered substrate range, but also a consequence of acquisition of a transporter with different uptake kinetics. Collective cellular transporter function is therefore under selection for continual changes in protein function brought on by changes in the environment, including the opportunity for niche expansion and the inherent instability of environmental substrate concentrations. Consequently, selection and drift will continually drive the reconfiguration of transporter complements, causing complex patterns of phylogenetic inheritance within transporter gene families, where acquisition of transporters can continually augment and/or overwrite the resident transporter repertoire, essentially operating as a gene acquisition ratchet (Fig. 5). This ratchet renders phylogenetic tree topologies of transporter gene families highly discontinuous, as is observed for many such gene families (51-53). This ratchet and the discontinuity of transporter gene family evolution has important implications for understanding how microbial pan-genomes evolve and how microbes colonise new and variant environments, including host environments where colonisation is a precursor to disease.

Materials and Methods

Identification and phylogenetic analysis of transporter HGTs

Using the annotated yeast proteome (39) we used BLAST2GO, TMHMM and PFAM annotations to identify candidate transporter gene families. For each candidate transporter we used BLASTp searches (conducted using iterative gathering thresholds in a range between $1e^{-50}$ to $1e^{-10}$) of a custom-built database of 160 fungal genomes (File S4) to sample a diverse collection of putative transporter protein homologues. BLASTp searches were conducted separately for each individual genome in the gene database to maximise recovery of candidate protein homologues from each genome. Phylogenies were generated for each set of transporter gene families using a custom-built pipeline (35); briefly, amino acids were aligned using MAFFT (54), masked using TRIMAL v1.4 (55), and a tree was generated using FastTree2 (56) using a SH-like aLRT method for assessing topology support. Trees were then manually inspected for evidence of HGT between ascomycetes and basidiomycetes. Putative HGTs were identified when closely related groups of ascomycete transporter genes and taxa (i.e. the recipient group) branched within a diverse clade of basidiomycete genes (i.e. the donor group), or vice versa. The resulting alignments were then manually edited to remove distant paralogues (determined from the phylogenetic tree) and any highly similar sequences from the same genus/groups. These sequences were then BLASTp searched against the NCBI *nr* database and the JGI Mycocosm portal (as of May 2018) to check for missing taxon sampling (due to proteomes of taxa being released after the original analysis), the alignments were manually corrected by removing amino acid sequences with poor gene models that branched separately to the HGT donor and recipient clade and the final alignments were masked manually to generate a refined data matrix for further phylogenetic analysis.

For the final phylogenetic analysis, we used the automatic model selection feature within IQ-Tree multicore version 1.6.2 (57) to identify an appropriate substitution model using the Bayesian

293 Information Criterion (BIC) (Table S4). This was then used to calculate an ML phylogeny with 200
294 standard bootstrap replicates (not Ultra-Fast or fast bootstrap methods). All HGT topologies were
295 supported with one or more node/s of ≥90%, two nodes ≥80%, or three or more nodes ≥50% bootstrap
296 support for the recipient group branching within the donor group (i.e. an HGT tree-topology). Donor
297 groups were identified based on evidence of a local ascomycete (or basidiomycete – depending on
298 the direction of transfer) paralogue set (see File S1 & S2). This approach allowed us to exclude the
299 possibility of interpreting HGTs as differentially lost ascomycete and/or basidiomycete paralogues.
300 Using the same approach in IQ-Tree, we then constructed constrained phylogenies so that the tree-
301 search was prevented from searching tree topologies where the recipient group branched within the
302 donor group (see File S1 for illustration of each constrained donor group). Constraints were selected
303 so only the donor paralogue was constrained as monophyletic. To statistically compare alternative
304 constrained 'non-HGT' topologies with unconstrained HGT topologies, we used a Bayes factor
305 approach (58). Bayesian runs were performed for both constrained and unconstrained topology
306 searches using P4 (59), with the sequence substitution models shown in Table S4. Each tree was run
307 twice to check that the likelihood plots of the tree searches were similar, with each tree run for
308 400,000 generations. Convergence was assessed by examination of the likelihood plots (average
309 marginal likelihood difference between replicate runs was 2.23, with the largest difference being 7.9).
310 The Bayes factor B_{10} can be defined as the ratio of marginal likelihoods of alternative tree topologies
311 (hypotheses H1 and H0), with values of $\log_e B_{10}$ interpreted as follows: 1-3 positive; 3-5 strong; >5 very
312 strong (58). Using the approach proposed by Newton and Raftery (1994) to compare marginal
313 likelihood and identify Bayes factors for the unconstrained trees (H1) and the constrained trees (H0),
314 we were able to demonstrate that all HGT phylogenetic relationships were robust according to
315 established criterion for Bayes factor tests (60) (Fig. S9). Similar tests were applied for the phylogenies
316 that sampled additional recently released genome data and which allowed the identification of

317 additional putative HGT-recipients (i.e. HGT-2, -3 & -6, see File S2). As part of this analysis we used
318 constrained vs unconstrained tree topology analysis and Bayes factor tests to investigate if the
319 multiple HGT-recipient clades identified were monophyletic or paraphyletic (File S2). This process
320 identified an additional two candidate ascomycete-to-basidiomycete HGTs (HGT-3 and HGT-6).

321 **Projecting protein connectivity**

322 All *S. cerevisiae* (S288c) sequences with >20% amino acid identity according to BLASTp searches of
323 each HGT amino acid sequence were recovered. The degree of connectivity (using BioGRID data (40),
324 collated by Cotton and McInerney (39)) and protein identity for each BLASTp hit were then used to
325 compute a weighted mean of the projected protein-protein connectivity for each of the HGT gene
326 products in *S. cerevisiae*. The weighted means were calculated based on percentage protein identity
327 of the yeast amino acid sequences with shared identity to the HGT.

328 **Strains, plasmids and culture conditions**

329 *S. cerevisiae* strains listed in Table S1 were grown at 30°C, 180 rpm. Strains were typically grown in SC
330 medium (0.79% yeast nitrogen base without (w/o) amino acids (Formedium), 700 mg L^{-1} complete
331 supplement mix (Formedium), 2% (w/v) glucose and 1.8% (w/v) Agar No. 2 Bacteriological (Lab M),
332 with dropout media used where required. Strain EBY.VW4000 (61) was grown in yeast nitrogen base
333 w/o amino acids supplemented with 2% (w/v) maltose, 100 µg ml^{-1} leucine, 50 µg ml^{-1} tryptophan, 20
334 µg ml^{-1} uracil and 20 µg ml^{-1} histidine. *S. cerevisiae* strains from the RFP-tagged library (41) were grown
335 in the presence of G418 Sulfate (Invitrogen) at 200 µg ml^{-1}.

336 **Cloning and construction of yeast expression plasmids**

ORFs were synthesized *de novo* (Genewiz Inc.) and cloned into yeast expression vector p426-GPD (ATCC) with a C-terminal 6xHis tag. ORFs were also cloned into p423-GPD by digestion with SpeI and XmaI restriction enzymes (New England BioLabs) and ligation into p423-GPD (ATCC) using T4 DNA ligase (Thermo Scientific). To generate a Gateway-compatible plasmid with a C-terminal sfGFP sequence, the *attR1*/*ccdB*/*attR2* region of pAG426GPD-ccdB-EGFP was amplified using Phusion polymerase and primers attR1_BglII_F and attR2_HindIII_R (Table S2). The resulting PCR product was digested with BglII and HindIII and cloned into the BamHI/HindIII sites of the non-Gateway p426-GPD sfGFP plasmid to generate Gateway-compatible pDM002 (pAG426 GPD *ccdB* sfGFP).

To assess localization, ORFs were amplified using primers in Table S2 and cloned into pDONR221 using Gateway recombination (Life Technologies). The final construct was generated by mobilization into pDM002 (construction detailed above) to generate a C-terminal sfGFP construct fusion (HGT-1-to-6) or into pAG426GPD-EGFP-ccdB (62) to generate an N-terminal fusion (HGT-7). Plasmid p426-GPD *stl1* was constructed by amplification of the *stl1* gene from *S. cerevisiae* W303-1A genomic DNA using Phusion DNA polymerase and primers Stl1_SpeI_F and Stl1_XmaI_R (Table S2). The PCR product was then digested with XmaI and SpeI and ligated into p426-GPD. For plasmid transformation, competent cells were prepared as described by Thompson *et al*. (63), mixed with 500 ng plasmid DNA and pulsed at 1.5 kV in an Eppendorf electroporator. Cells were suspended in YPD (20 g L^{-1} peptone bacteriological (Oxoid), 10 g L^{-1} yeast extract (Oxoid), 2% (w/v) glucose) and grown at 30°C with 180 rpm shaking for 16 hours before plating on the appropriate selective growth medium.

Microplate growth assays

S. cerevisiae cells (8 replicates per strain) were grown to stationary phase at 30°C, 180 rpm, washed and suspended in SC medium (2% glucose) lacking uracil (SC-ura) and diluted to an OD_{595} of 0.1. Cells were inoculated into a 96-well flat-bottom microplate, covered with a sterile polyester film and grown

at 30°C in a Fluostar Omega Lite microplate reader (BMG Labtech Ltd.). Optical density (595 nm) was assessed at 10-minute intervals, with continuous double-orbital shaking (200 rpm) between reads. Growth curve parameters were calculated using the Growthcurver R package (64).

Western blots

Strains were grown for 16 hours at 30°C, 180 rpm shaking and 1 ml culture was pelleted and snap-frozen in liquid nitrogen. The pellet was thawed on ice and suspended in 250 µl complete cracking buffer (480 g L^{-1} urea, 50 g L^{-1} SDS, 400 mg L^{-1} Bromophenol blue, 0.1 mM EDTA, 40 mM Tris-HCl (pH 6.8)) supplemented with 2.5 µl 2-Mercaptoethanol, 8.25 µl Pepstatin A (1 mg ml^{-1}), 0.25 µl Leupeptin (1 mg ml^{-1}) and 12.5 µl phenylmethylsulfonyl fluoride (100x). Cells were disrupted by addition of glass beads, followed by heating at 70°C for 10 minutes and homogenization in a FastPrep-24 instrument (MP Biomedicals). Cell debris was pelleted by centrifugation at 16,000 x *g* at 4°C, the supernatant was removed and a second extraction was performed. The supernatant was incubated at 100°C for 2 minutes prior to protein separation on a 10% SDS-PAGE gel. Following gel electrophoresis, proteins were transferred onto a membrane in Tris-Glycine transfer buffer and the membrane was washed in TBS. The membrane was then incubated for 1 hour in blocking solution (1% Alkali-soluble casein), washed twice in TBS supplemented with 0.2% Triton X-100 and 0.05% Tween-20, then once in TBS. The membrane was incubated in a 1000-fold dilution of the His-Tag Antibody HRP-conjugate (Merck Millipore), the wash steps were repeated and the membrane was incubated in ECL2 Western Blotting substrate (Pierce) before imaging to assess presence/absence of each His-tagged transporter protein.

Quantitative Western blots

Quantitative Western blotting was performed based on the method of Schütz *et al.* (65). Briefly, strains were grown for 16 hours at 30°C, 180 rpm shaking and 4 x 10^7 cells (equivalent to an OD_{600} of

4) were pelleted, washed in water and re-suspended in 1 ml 2M lithium acetate. Cells were incubated on ice for 5 minutes, pelleted, re-suspended in 200 µL 0.4M NaOH and incubated on ice for a further 5 minutes. Cells were then centrifuged and re-suspended in 100 µL TruPAGE LDS Sample Buffer (Sigma-Aldrich) containing 50 mM DTT and incubated at room temperature for 15 minutes.

Proteins were separated on a TruPAGE 10% polyacrylamide gel (Sigma-Aldrich) alongside the Chameleon Duo ladder (LI-COR) and transferred onto Immobilon-FL membrane (Merck Millipore). Total protein staining was performed using REVERT Total Protein Stain (LI-COR) following the manufacturer's instructions.

The membrane was washed in PBS and blocked overnight in 0.1% Alkali-soluble Casein in 0.2x PBS. The membrane was then washed in PBS, incubated in Rabbit Anti-6X His tag antibody (abcam; ab9108) (1:1000) in blocking buffer + 0.1% Tween20 for 1 hour, washed 4 times in PBST, then incubated in Goat anti-Rabbit antibody conjugated to 800CW (abcam; ab216773) (1:10,000) in blocking buffer + 0.1% Tween20 for 3 hours. The membrane was washed 4 times in PBST, then once in PBS before imaging on a LI-COR Odyssey Fc imaging system.

Phenotype Microarrays

To prepare cells for OmniLog® Phenotype Microarray (PM) plates, each yeast strain was grown on SC medium (with appropriate auxotrophic selection) at 30°C for 48-72 hours. Colonies were suspended in Yeast Nutrient Supplement (NS) solution and adjusted to an OD_{600} of 0.2. The appropriate volume of cell suspension was added to inoculating fluid specific to each PM plate type, as detailed in Table S3. Initial PM analyses utilized yeast strain BY4742, with specific analyses using deletion strains described below. PM plates were incubated at 30°C for 48 hours or, for EBY.VW4000 strains, at 30°C

for 96 hours. Each assay was performed in triplicate using three independently obtained yeast transformants.

OmniLog Phenotype Microarray outputs were analysed by normalizing each individual plate against well A01 (negative control) to control for any background growth as a result of the inoculation solution and/or metabolite carry over, then analysed using OPM (66). Within OPM, an analysis of variance (aov) model was used, with multcomp algorithms (67) within the omp_mcp function used for statistical comparisons of group means. For *S. cerevisiae* EBY.VW4000 strains, which were incubated over a longer time period, data were aggregated using the 'opm-fast' method. For all other analyses, data were aggregated using the default curve parameters within the 'grofit' package. For BY4742 assays, compounds where growth reached over 75 'OmniLog units' and which rescued OmniLog dye reduction to levels not significantly different to the empty vector wildtype control were deemed to be substrates. Dipeptides containing histidine, lysine or cysteine were excluded from these analyses as such compounds have been shown to cause dye reduction even in the absence of growth (43).

Yeast complementation assays

To assess the ability of each transporter to complement yeast strains lacking specific transporter proteins, vectors were transformed into yeast strains with reported transporter gene deletions (see Table S1 for strain information). BY4742 *ptr2Δ::KanMx* was transformed with p426-GPD or p426-GPD HGT-3. *Δptr2* transformants were grown to stationary phase in SC-ura, washed twice and suspended in water, then spotted onto SC medium (2% glucose) supplemented with complete supplement mixture lacking histidine, leucine and uracil and 10 mM His-Leu dipeptide (Sigma Aldrich); or supplemented with complete supplement mixture lacking leucine and uracil and 1 mM Leu-Ala hydrate (Sigma Aldrich).

BY4742 *stl1Δ::KanMx* was transformed with p426-GPD, p426-GPD-HGT-1 or p426-GPD-*STL1*. BY4742 *stl1* transformants were grown to stationary phase and suspended in YPG lacking uracil (0.79% yeast nitrogen base w/o amino acids, 770 mg L^{-1} complete supplement mix -URA (Formedium) and 2% glycerol) at OD_{600} 0.1 and growth was observed by measuring OD_{600}.

S. cerevisiae strain EBY.VW4000 was transformed with p423-GPD or p423-GPD-HGT-4 and selected on YNB medium (yeast nitrogen base w/o amino acids supplemented with leucine, tryptophan and uracil), 2% maltose. Transformants were grown in the same medium, washed once and diluted to OD_{600} 1.0 in water, and a dilution series was spotted onto YNB medium supplemented with 2% sugar. Growth was assessed after five days incubation at 30°C. Xylose plates were assessed after 8 days incubation. The same was done for *S. cerevisiae suc2* strains (a mutant that is not able to secrete invertase to catalyse breakdown of sucrose to hexose subunits), and were spotted onto YNB medium supplemented with leucine, lysine, uracil and 2% sucrose. This was done to assess if the HGT-transporter could transport sucrose. Each image is representative of the three biological replicates performed. *S. cerevisiae* strain EBY.VW4000 was also transformed with p426-GPD-HGT-1 and p426-GPD-*STL1* to assess growth on 2% glucose (as above). Growth was assessed after five days incubation at 30°C.

Heterologous localisation of HGT transporters using spinning disc confocal microscopy

Cells were grown for 16 hours in SC-ura plus 2% glucose at 30°C, harvested, washed and re-suspended in PBS. The preferential cell wall stain Wheat Germ Agglutinin, Alexa Fluor™ 594 conjugate (Life Technologies) was added to a final concentration of 10 µg ml^{-1} and cells were observed by spinning disc confocal microscopy. Microscopy was performed using an Olympus IX81 inverted microscope and CSU-X1 Spinning Disc unit (Yokogawa, Tokyo, Japan). 488 nm and 561 nm solid-state lasers were used

in combination with an x100/1.40 oil objective. A Photometrics CoolSNAP HQ2 camera (Roper Scientific, Germany) and VisiView software package (Visitron Systems) were used for imaging.

Competitive assays

Competitive assays were performed using *S. cerevisiae* W303 strains engineered to constitutively express GFP (CAY195) or BFP (CAY173) at the endogenous URA3 locus. These fluorescent tags were chosen as the BFP fluorophore is a result of a single amino acid substitution (Y66F), and comparative growth has demonstrated that they maintain relative growth dynamics (Fig. S6). URA3 integration was followed by its replacement with the GFP/BFP cassette by modification of previously described plasmids (68). Positive transformants were selected using 5-Fluoroorotic acid (5-FOA) and strains were backcrossed three times to eliminate random mutations as a result of 5-FOA treatment. *S. cerevisiae* CAY strains, transformed with either p426-GPD or p426-GPD HGT-1/p426-GPD *STL1*, were grown overnight on selective SC-ura agar. Strains were then grown overnight in 5 ml SC-ura (30°C, 180 rpm) until late logarithmic phase, washed once, and re-suspended in water to 2.5 x 10^7 cells/ml. Cells were counted using a Beckman Coulter CytoFLEX S flow cytometer and equal cell numbers (1 x 10^6) of the reference strain and competing strain were added to 10 ml SC-ura containing 0.1-5% glycerol. At 24 hour intervals, cells were re-suspended in PBSE (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, 0.1 mM EDTA, pH 7.4) containing 1 µg ml^{-1} propidium iodide (to allow the exclusion of dead cells) and analysed by flow cytometry to calculate live cell numbers for each strain. All analyses were based on three biological replicates.

For competitive assays requiring a carbon source switch at 12-hour intervals, 1 ml of each 10 ml culture was removed, centrifuged at 17,000 x *g* for 5 minutes, and suspended in a new 10 ml culture containing the replacement carbon source. A further 1 ml of each culture was also removed, suspended in PBSE and 1 µg ml^{-1} propidium iodide (to exclude dead cells), and assayed by flow

cytometry, as above. Competitive growth assays were analysed using a generalized linear model with a binomial error distribution. The analysis was performed in R version 3.1.2 (R Core Team 2014). Cell counts for each competing strain (expressed as proportions), 72 hours after mixed-genotype populations were established, were used for the response variable. The model included a three-way interaction between substrate, concentration and type of competition (i.e. an HGT-1 expressing strain versus a p426-GPD vector only control strain or a Stl1p expressing strain).

^{14}C glycerol accumulation

S. cerevisiae strains were grown for 16 hours at 30°C shaking. Cells were then diluted, grown until early log-phase and then harvested by centrifugation at 3,000 x *g* for 5 minutes. The weight of each sample was recorded and cells were washed twice and suspended in water. 10 µl cells was added to 10 µl Tris/citrate buffer (pH 5.0) and cells were allowed to equilibrate to 30°C. [^{14}C(U)]-glycerol (Perkin Elmer) was added to a final concentration of 1 mM and cells were incubated at 30°C for 30 minutes. At intervals, cells were quenched by the addition of 1 ml ice-cold water and collected by centrifugation at 14,000 x *g* for 3 minutes. Cells were washed once, suspended in 500 µl water and added to 2.5 ml Emulsifier-Safe scintillation cocktail solution (Perkin Elmer). Radioactivity was determined in a liquid scintillation analyser (Beckman Coulter LS 6500), and all analyses were based on three biological replicates.

Acknowledgments

We would like to thank Prof. Eckhard Boles for the gift of the *S. cerevisiae* EBY.VW4000 strain and Dr. Cat Gadelha for the p426 GPD sfGFP vector. The *S. cerevisiae* BY4742 strain and reagents for transformations/Western Blots were kindly provided by Prof. Ken Haynes' lab. We also wish to thank Dr. Varun Varma for his advice on statistical analyses and Emma Chapman for her technical assistance. FM is supported by a Genome Canada fellowship and CAM is supported by Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/N016858/1. This work is supported mainly by funding from a Philip Leverhulme award from the Leverhulme Trust (PLP-2014-147) with additional support from The Gordon and Betty Moore Foundation (GBMF5514). The University of Exeter OmniLog system and associated operations were supported by a Wellcome Trust Institutional Strategic Support Award WT105618MA. TAR is supported by a Royal Society University Research Fellowship (UF130382).

References

1. Ochman H, Lawrence JG, & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299-304.
2. Popa O, Hazkani-Covo E, Landan G, Martin W, & Dagan T (2011) Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21(4):599-609.
3. Keeling PJ & Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9(8):605-618.
4. McCarthy CG & Fitzpatrick DA (2016) Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to oomycete lineages. *mSphere* 1(5).

5. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, & Talbot NJ (2006) Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr. Biol.* 16(18):1857-1864.

6. Richards TA*, et al.* (2011) Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proc. Natl. Acad. Sci. USA* 108(37):15258-15263.

7. Misner I, Blouin N, Leonard G, Richards TA, & Lane CE (2014) The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. *Genome Biol. Evol.* 7(1):120-135.

8. Belbahri L, Calmin G, Mauch F, & Andersson JO (2008) Evolution of the cutinase gene family: evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. *Gene* 408(1-2):1-8.

9. Wisecaver JH & Rokas A (2015) Fungal metabolic gene clusters-caravans traveling across genomes and environments. *Front. Microbiol.* 6:161.

10. Marcet-Houben M & Gabaldon T (2010) Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* 26(1):5-8.

11. Slot JC & Rokas A (2011) Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr. Biol.* 21(2):134-139.

12. Coelho MA, Goncalves C, Sampaio JP, & Goncalves P (2013) Extensive intra-kingdom horizontal gene transfer converging on a fungal fructose transporter gene. *PLoS Genet.* 9(6).

13. Doolittle WF (1999) Lateral genomics. *Trends Cell. Biol.* 9(12):M5-8.

14. Jain R, Rivera MC, Moore JE, & Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* 20(10):1598-1602.

15. Wisecaver JH, Slot JC, & Rokas A (2014) The evolution of fungal metabolic pathways. *PLoS Genet.* 10(12):e1004816.

16. Gojkovic Z*, et al.* (2004) Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. *Mol. Genet. Genomics* 271(4):387-393.

17. Hall C, Brachat S, & Dietrich FS (2005) Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot. Cell.* 4(6):1102-1115.

18. Hall C & Dietrich FS (2007) The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics* 177(4):2293-2307.

19. Marsit S*, et al.* (2015) Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Mol. Biol. Evol.* 32(7):1695-1707.

20. Savory FR, Milner DS, Miles DC, & Richards TA (2018) Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. *Mol. Biol. Evol.* 35(8):1887-1900.

21. Alexander WG, Wisecaver JH, Rokas A, & Hittinger CT (2016) Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. *Proc. Natl. Acad. Sci. USA* 113(15):4116-4121.

22. Goncalves C*, et al.* (2018) Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *eLife* 7.

23. Wisecaver JH, Alexander WG, King SB, Hittinger CT, & Rokas A (2016) Dynamic Evolution of Nitric Oxide Detoxifying Flavohemoglobins, a Family of Single-Protein Metabolic Modules in Bacteria and Eukaryotes. *Mol. Biol. Evol.* 33(8):1979-1987.

24. Husnik F & McCutcheon JP (2018) Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* 16(2):67-79.

25. Richards TA & Talbot NJ (2013) Horizontal gene transfer in osmotrophs: playing with public goods. *Nat. Rev. Microbiol.* 11(10):720-727.

556 26. Metcalf JA, Funkhouser-Jones LJ, Brileya K, Reysenbach AL, & Bordenstein SR (2014)
557 Antibacterial gene transfer across the tree of life. *eLife* 3.
558 27. Moran Y, Fredman D, Szczesny P, Grynberg M, & Technau U (2012) Recurrent horizontal
559 transfer of bacterial toxin genes to eukaryotes. *Mol. Biol. Evol.* 29(9):2223-2230.
560 28. Lercher MJ & Pal C (2008) Integration of horizontally transferred genes into regulatory
561 interaction networks takes many million years. *Mol. Biol. Evol.* 25(3):559-567.
562 29. Cohen O, Gophna U, & Pupko T (2011) The complexity hypothesis revisited: connectivity
563 rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.*
564 28(4):1481-1489.
565 30. Jain R, Rivera MC, & Lake JA (1999) Horizontal gene transfer among genomes: The complexity
566 hypothesis. *Proc. Natl. Acad. Sci. USA* 96(7):3801-3806.
567 31. Savory F, Leonard G, & Richards TA (2015) The role of horizontal gene transfer in the evolution
568 of the oomycetes. *PLoS Pathog.* 11(5):e1004805.
569 32. Cheeseman K*, et al.* (2014) Multiple recent horizontal transfers of a large genomic region in
570 cheese making fungi. *Nat. Commun.* 5:2876.
571 33. Richards TA & Talbot NJ (2018) Osmotrophy. *Curr. Biol.* 28(20):R1179-R1180.
572 34. Goffeau A*, et al.* (1996) Life with 6000 genes. *Science* 274(5287):546, 563-547.
573 35. Richards TA*, et al.* (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient
574 horizontal gene transfer between plants and fungi. *Plant Cell* 21(7):1897-1911.
575 36. Soanes D & Richards TA (2014) Horizontal gene transfer in eukaryotic plant pathogens. *Annu.*
576 *Rev. Phytopathol.* 52:583-614.
577 37. Spatafora JW*, et al.* (2016) A phylum-level phylogenetic classification of zygomycete fungi
578 based on genome-scale data. *Mycologia* 108(5):1028-1046.

38. Risinger AL, Cain NE, Chen EJ, & Kaiser CA (2006) Activity-dependent reversible inactivation of the general amino acid permease. *Mol. Biol. Cell.* 17(10):4411-4419.

39. Cotton JA & McInerney JO (2010) Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc. Natl. Acad. Sci. USA* 107(40):17252-17255.

40. Stark C*, et al.* (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34(Database issue):D535-539.

41. Huh WK*, et al.* (2003) Global analysis of protein localization in budding yeast. *Nature* 425(6959):686-691.

42. Yamauchi S*, et al.* (2015) Opt2 mediates the exposure of phospholipids during cellular adaptation to altered lipid asymmetry. *J. Cell Sci.* 128(1):61-69.

43. Homann OR, Cai HJ, Becker JM, & Lindquist SL (2005) Harnessing natural diversity to probe metabolic pathways. *PLoS Genet.* 1(6):715-729.

44. Toivari MH, Salusjarvi L, Ruohonen L, & Penttila M (2004) Endogenous xylose pathway in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 70(6):3681-3686.

45. Reznicek O*, et al.* (2015) Improved xylose uptake in *Saccharomyces cerevisiae* due to directed evolution of galactose permease Gal2 for sugar co-consumption. *J. Appl. Microbiol.* 119(1):99-111.

46. Kasahara M, Shimoda E, & Maeda M (1997) Amino acid residues responsible for galactose recognition in yeast Gal2 transporter. *J. Biol. Chem.* 272(27):16721-16724.

47. Bungay HR (1981) *Energy, the biomass options* (Wiley, New York) pp viii, 347 p.

48. Heavner BD, Smallbone K, Price ND, & Walker LP (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database: the Journal of Biological Databases and Curation*.

603 49. Mattenberger F, Sabater-Munoz B, Hallsworth JE, & Fares MA (2016) Glycerol stress in
604 *Saccharomyces cerevisiae*: cellular responses and evolved adaptations. *Environ. Microbiol.*
605 19(3):990-1007.
606 50. Ferreira C*, et al.* (2005) A member of the sugar transporter family, Stl1p is the glycerol/H+
607 symporter in *Saccharomyces cerevisiae*. *Mol. Biol. Cell.* 16(4):2068-2076.
608 51. McDonald SM, Plant JN, & Worden AZ (2010) The mixed lineage nature of nitrogen transport
609 and assimilation in marine eukaryotic phytoplankton: a case study of *Micromonas*. *Mol. Biol.*
610 *Evol.* 27(10):2268-2283.
611 52. McDonald TR, Dietrich FS, & Lutzoni F (2012) Multiple horizontal gene transfers of ammonium
612 transporters/ammonia permeases from prokaryotes to eukaryotes: toward a new functional
613 and evolutionary classification. *Mol. Biol. Evol.* 29(1):51-60.
614 53. Monier A*, et al.* (2012) Phosphate transporters in marine phytoplankton and their viruses:
615 cross-domain commonalities in viral-host gene exchanges. *Environ. Microbiol.* 14(1):162-176.
616 54. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7:
617 improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772-780.
618 55. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated
619 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972-1973.
620 56. Price MN, Dehal PS, & Arkin AP (2010) FastTree 2 - approximately maximum-likelihood trees
621 for large alignments. *Plos One* 5(3):e9490.
622 57. Nguyen LT, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: a fast and effective
623 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*
624 32(1):268-274.
625 58. Kass RE & Raftery AE (1995) Bayes Factors. *J. Am. Stat. Assoc.* 90(430):773-795.
626 59. Foster PG (2004) Modeling compositional heterogeneity. *Syst. Biol.* 53(3):485-495.

60. Newton MA & Raftery AE (1994) Approximate Bayesian-inference with the weighted likelihood bootstrap. *J. R. Stat. Soc. B.* 56(1):3-48.

61. Wieczorke R*, et al.* (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* 464(3):123-128.

62. Alberti S, Gitler AD, & Lindquist S (2007) A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* 24(10):913-919.

63. Thompson JR, Register E, Curotto J, Kurtz M, & Kelly R (1998) An improved protocol for the preparation of yeast cells for transformation by electroporation. *Yeast* 14(6):565-571.

64. Sprouffske K & Wagner A (2016) Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* 17:172.

65. Schutz M*, et al.* (2016) Directed evolution of G protein-coupled receptors in yeast for higher functional production in eukaryotic expression hosts. *Sci. Rep.* 6:21508.

66. Vaas LA*, et al.* (2013) *opm*: an R package for analysing OmniLog(R) phenotype microarray data. *Bioinformatics* 29(14):1823-1824.

67. Hothorn T, Bretz F, & Westfall P (2008) Simultaneous inference in general parametric models. *Biom. J.* 50(3):346-363.

68. Hittinger CT & Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449(7163):677-U671.

69. Wieczorke R*, et al.* (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* 464(3):123-128.

Table 1. Details of transporter HGT events identified and characterised using *S. cerevisiae* heterologous expression. HGT events of transporter-encoding genes; their recipient genomes and other recipient genomes in which these HGT events were identified (indicated in brackets); conserved domain predictions; *S. cerevisiae* strains used for characterisation and the transporter substrates identified using each method.

HGT	Accession no.	Recipient	Conserved Domain	*S. cerevisiae* strain	OmniLog	Complementation	Strain reference	Substrates identified	Figure
HGT-1	EJD44146.1	*Auricularia subglabra* (Basidiomycota)	pfam00083 Sugar_tr [Sugar (and other) transporter]	BY4742 Δ*stl1*	✗	✓	Euroscarf collection	Glycerol	Fig S3
				EBY.VW4000	✗	✓	(69)	*Does not transport glucose*	Fig S8
HGT-2	EJD38419.1	*Auricularia subglabra* (*Exidia* [x2]) (Basidiomycota)	COG0833 LysP [Amino acid permease]	-	-	-	-	*No expression detected*	-
HGT-3	XP_007862603.1	*Gloeophyllum trabeum* (*Heliocybe*) (Basidiomycota)	cl21472 MFS super family [The Major Facilitator Superfamily]	BY4742	✓	-	Euroscarf collection	L-glutamine; Arg-Ala; Arg-Asp; Arg-Gln; Arg-Ile; Arg-	File S3

									Leu; Arg-Met; Arg-Phe; Arg-Val; Ile-Gly; Ile-Met; Leu-Ala; Leu-Arg; Leu-Ile; Leu-Met; Leu-Ser; Met-Gln; Phe-Ser; Thr-Arg; Thr-Leu; Trp-Arg; Tyr-Ala; Tyr-Gln; Val-Asn (OmniLog)	
				BY4742 Δ*ptr2*	✓	✓	Euroscarf collection	His-Leu and Leu-Ala dipeptides (spot assays) Ala-Arg; Arg-Ala; Arg-Asp; Arg-Gln; Arg-Leu; Arg-Ser; Gly-Arg; Leu-Ala;	Fig S4 Fig S5	

								Leu-Arg; Leu-Met; Leu-Ser; Met-Arg; Met-Gln; Met-Leu; Met-Met; Trp-Arg; Trp-Ser; Tyr-Gln; Val-Asn (OmniLog)	
HGT-4	XP_007868407.1	*Gloeophyllum trabeum* (*Heliocybe*) (Basidiomycota)	pfam00083 Sugar_tr [Sugar (and other) transporter]	EBY.VW4000	✓	✓	(69)	Mannose, glucose, galactose, xylose.	Fig 3C-D
				BY4742 Δ*suc2*	-	✓	Euroscarf collection	*Does not transport sucrose*	Fig 3E
HGT-5	ODQ72208.1	*Lipomyces starkeyi* (Ascomycota)	COG0833 LysP [Amino acid permease]	-	-	-	-	*No expression detected*	-

HGT-6	XP_008042479.1	*Trametes versicolor* (Basidiomycota)	cl27298 UhpC superfamily [Sugar phosphate permease]	BY4742	✓	✗	Euroscarf collection	L-glutamine; Ala-Arg; Arg-Ala; Arg-Asp; Arg-Ile; Arg-Leu; Arg-Met; Arg-Phe; Arg-Ser; Arg-Tyr; Arg-Val; Ile-Arg; Leu-Ala; Leu-Arg; Leu-Asp; Leu-Glu; Leu-Gly; Leu-Met; Leu-Ser; Met-Gln; Phe-Ser; Trp-Arg; Tyr-Gln; Val-Asn; Gly-Asn; Ile-Asn; Leu-Asn; Phe-Asp; Phe-Glu; Phe-Val; Pro-Asn; Ser-Asn;	File S3

									Val-Ser (OmniLog)	
	HGT-7	XP_011390539.1	*Ustilago maydis* (*Sporisorium* [x2]; *Pseudozyma* [x2]; *Melanopsichium*; *Kalmanozyma*; *Moesziomyces*) (Basidiomycota)	pfam03169OPT [oligopeptide transporter protein]	BY4742	✓	✗	Euroscarf collection	Arg-Asp; Arg-Met; Leu-Ala; Phe-Ser; Thr-Leu; Tyr-Ala; Val-Asn; Gly-Asn; Ile-Asn; Phe-Asp; Phe-Glu (OmniLog)	File S3

650

651

652

Figure Legends

Fig. 1. Phylogeny illustrating seven transfers of predicted transporter-encoding genes. Phylogeny of published genomes showing proximate points of HGT origin and acquisition for the seven primary HGTs identified (File S1). For two additional HGTs identified with increased genome sampling within HGT-3 and -6 wider gene families see File S2. The species phylogeny was calculated from an alignment of 79 taxa (File S5) and 134,948 characters based on the JGI-1086 HMMs (37) (https://github.com/1KFG/Phylogenomics_HMMs) using an ultra-fast bootstrap approach in IQ-Tree v1.5.4 with LG model and 1000 ultra-fast bootstraps.

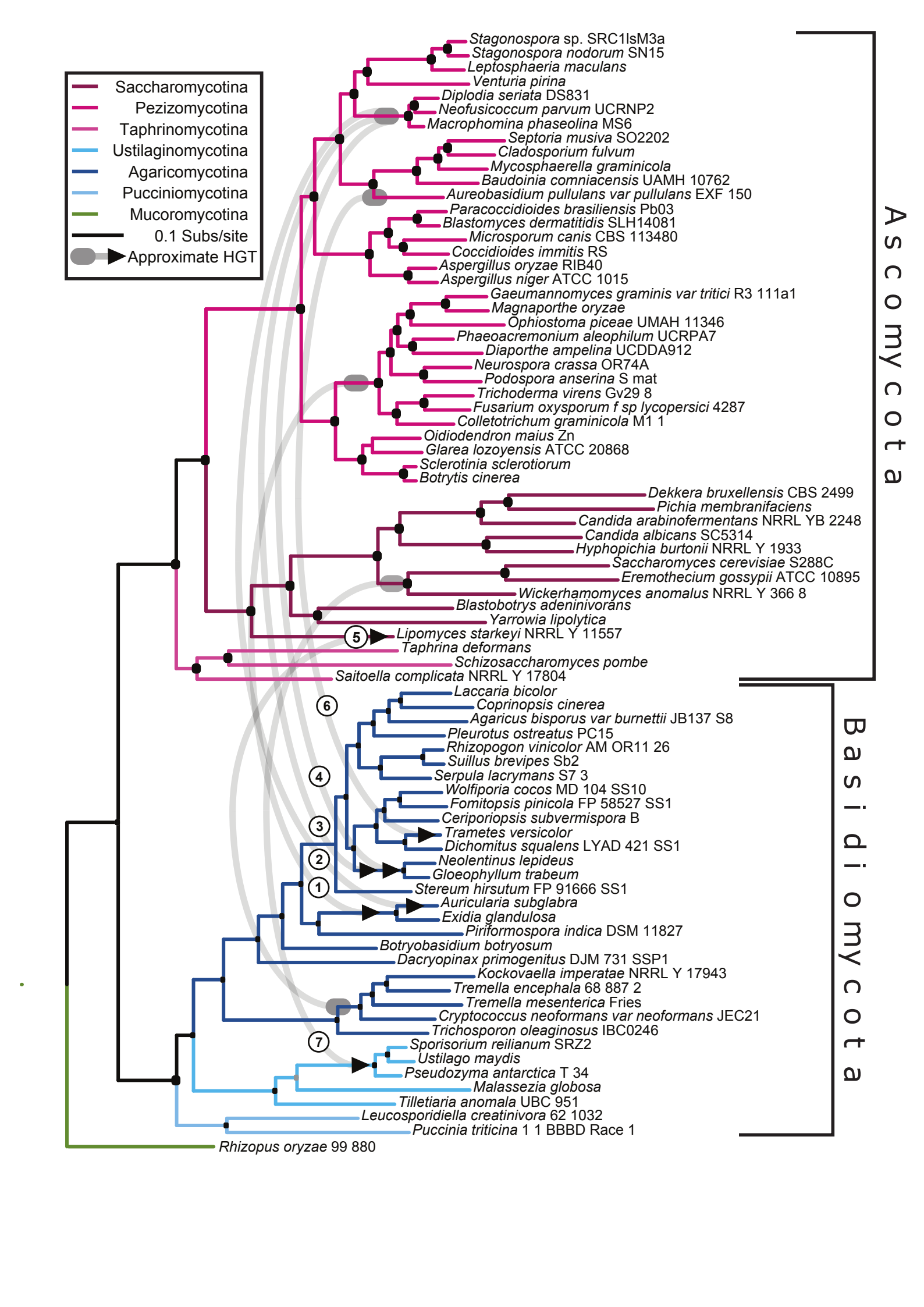

Saccharomycotina
Pezizomycotina
Taphrinomycotina
Ustilaginomycotina
Agaricomycotina
Pucciniomycotina
Mucoromycotina
0.1 Subs/site
Approximate HGT
Stagonospora sp. SRC1lsM3a
Stagonospora nodorum SN15
Leptosphaeria maculans
Venturia pirina
Diplodia seriata DS831
Neofusicoccum parvum UCRNP2
Macrophomina phaseolina MS6
Septoria musiva SO2202
Cladosporium fulvum
Mycosphaerella graminicola
Baudoinia comniacensis UAMH 10762
Aureobasidium pullulans var pullulans EXF 150
Paracoccidioides brasiliensis Pb03
Blastomyces dermatitidis SLH14081
Microsporum canis CBS 113480
Coccidioides immitis RS
Aspergillus oryzae RIB40
Aspergillus niger ATCC 1015
Gaeumannomyces graminis var tritici R3 111a1
Magnaporthe oryzae
Ophiostoma piceae UMAH 11346
Phaeoacremonium aleophilum UCRPA7
Diaporthe ampelina UCDDA912
Neurospora crassa OR74A
Podospora anserina S mat
Trichoderma virens Gv29 8
Fusarium oxysporum f sp lycopersici 4287
Colletotrichum graminicola M1 1
Oidiodendron maius Zn
Glarea lozoyensis ATCC 20868
Sclerotinia sclerotiorum
Botrytis cinerea
Dekkera bruxellensis CBS 2499
Pichia membranifaciens
Candida arabinofermentans NRRL YB 2248
Candida albicans SC5314
Hyphopichia burtonii NRRL Y 1933
Saccharomyces cerevisiae S288C
Eremothecium gossypii ATCC 10895
Wickerhamomyces anomalus NRRL Y 366 8
Blastobotrys adeninivorans
Yarrowia lipolytica
5
Lipomyces starkeyi NRRL Y 11557
Taphrina deformans
Schizosaccharomyces pombe
Saitoella complicata NRRL Y 17804
6
Laccaria bicolor
Coprinopsis cinerea
Agaricus bisporus var burnettii JB137 S8
Pleurotus ostreatus PC15
Rhizopogon vinicolor AM OR11 26
Suillus brevipes Sb2
4
Serpula lacrymans S7 3
Wolfiporia cocos MD 104 SS10
Fomitopsis pinicola FP 58527 SS1
3
Ceriporiopsis subvermispora B
Trametes versicolor
Dichomitus squalens LYAD 421 SS1
2
Neolentinus lepideus
Gloeophyllum trabeum
1
Stereum hirsutum FP 91666 SS1
Auricularia subglabra
Exidia glandulosa
Piriformospora indica DSM 11827
Botryobasidium botryosum
Dacryopinax primogenitus DJM 731 SSP1
Kockovaella imperatae NRRL Y 17943
Tremella encephala 68 887 2
Tremella mesenterica Fries
Cryptococcus neoformans var neoformans JEC21
Trichosporon oleaginosus IBC0246
7
Sporisorium reilianum SRZ2
Ustilago maydis
Pseudozyma antarctica T 34
Malassezia globosa
Tilletiaria anomala UBC 951
Leucosporidiella creatinivora 62 1032
Puccinia triticina 1 1 BBBD Race 1
Rhizopus oryzae 99 880
Ascomycota
Basidiomycota

Fig. 2. *S. cerevisiae* predicted protein connectivity distribution and investigation of projected protein-protein connectivity of HGT transporters when expressed in *Saccharomyces cerevisiae*. A. Protein connectivity by log-predicted length for *S. cerevisiae* proteome (source data collated by Cotton and McInerney (39)); yellow: transporter proteins; mauve: proteins acquired by HGT; green: all other *S. cerevisiae* proteins for which protein connectivity data is available. **B.** Distribution of protein connectivity by type, showing that transporter proteins, and proteins acquired by HGT, typically show low connectivity (*** $p < 0.001$). **C-I.** Projected protein-protein interactions for the *S. cerevisiae* homologues (>20% amino acid identity) of each HGT transporter gene identified here. HGT-acquired genes that were shown to be expressed are indicated in black, whilst those where expression was not detected are indicated in red. Orange arrows indicate *S. cerevisiae* homologues validated by complementation (C: Stl1; E: Ptr2) **J.** Box plot showing growth rates for each *S. cerevisiae* strain expressing an HGT-acquired gene, plotted against weighted mean connectivity for each group of *S. cerevisiae* homologues (blue diamonds). Circles represent outliers. **K.** Weighted linear regression of projected protein-protein interactions for the *S. cerevisiae* homologues (>20% amino acid identity) of each HGT transporter gene, vs. mean growth rate of each *S. cerevisiae* strain expressing each HGT. The regression was weighted by the squared proportional identity. Data used to generate Fig. 1A/B is available at doi: 10.6084/m9.figshare.6834770.

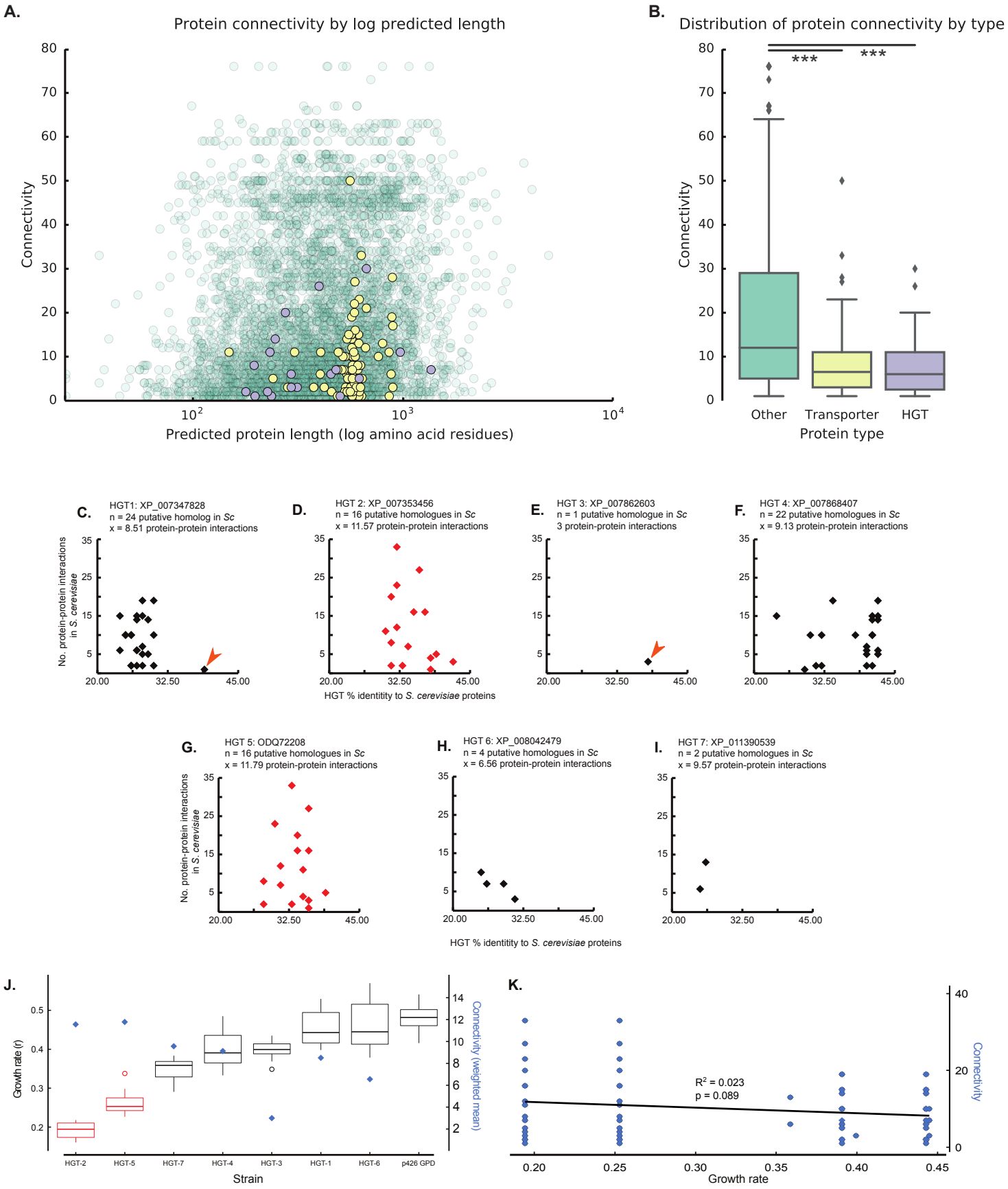

A.
Protein connectivity by log predicted length
Connectivity
Predicted protein length (log amino acid residues)
B.
Distribution of protein connectivity by type
***
***
Connectivity
Other
Transporter
HGT
Protein type
C.
HGT1: XP_007347828
n = 24 putative homolog in Sc
x = 8.51 protein-protein interactions
D.
HGT 2: XP_007353456
n = 16 putative homologues in Sc
x = 11.57 protein-protein interactions
E.
HGT 3: XP_007862603
n = 1 putative homologue in Sc
3 protein-protein interactions
F.
HGT 4: XP_007868407
n = 22 putative homologues in Sc
x = 9.13 protein-protein interactions
No. protein-protein interactions in S. cerevisiae
HGT % identity to S. cerevisiae proteins
G.
HGT 5: ODQ72208
n = 16 putative homologues in Sc
x = 11.79 protein-protein interactions
H.
HGT 6: XP_008042479
n = 4 putative homologues in Sc
x = 6.56 protein-protein interactions
I.
HGT 7: XP_011390539
n = 2 putative homologues in Sc
x = 9.57 protein-protein interactions
J.
Growth rate (r)
Connectivity (weighted mean)
HGT-2
HGT-5
HGT-7
HGT-4
HGT-3
HGT-1
HGT-6
p426 GPD
Strain
K.
R2 = 0.023
p = 0.089
Connectivity
Growth rate

Fig. 3. Localization of HGT-acquired transporter proteins and HGT-4 complementation assays and protein sequence substitution analysis. A. Localization of HGT-1, 3, 4, 6 when labelled with sfGFP. Co-localisation with the cell periphery as indicated by co-staining with WGA Alexa Fluor™ 594 conjugate, a lectin that preferentially binds to chitin. Scale bar = 3 µm. **B.** Localization of EGFP-labelled HGT-7; co-localisation was observed between EGFP-HGT-7 and Sec13 (ER-to-Golgi vesicles). Scale bar = 3 µm. **C.** Complementation of EBY.VW4000 hexose transporter null mutant with HGT-4 (+) or empty vector control (-) after 5 days growth (or 8 days for xylose) on different sugar sources. Dilution series from 10^0 to 10^{-5} (left to right) **D.** OmniLog complementation of EBY.VW4000 by HGT-4, showing restoration of growth on all substrates, with maltose as a positive control: red = p423-GPD HGT-4; blue = p423-GPD empty vector (* $p<0.05$, ** $p<0.01$, *** $p< 0.001$). **E.** Assessing for complementation of *S. cerevisiae suc2* deletion strain with HGT-4 (+) or empty vector control (-); dilution series from 10^0 to 10^{-5} (left to right). No complementation was observed, indicating that HGT-4 does not facilitate uptake of sucrose. **F.** Alignment of HGT-4 with Gal2 and Hxt7. Substitutions improving xylose uptake in Gal2 were identified by Reznicek *et al.* (45). Residues which were identical in HGT-4 to those observed in xylose-transporting Gal2 are indicted in dark red, whilst those which affected the same amino acid, but were not identical substitutions, are indicated in light red.

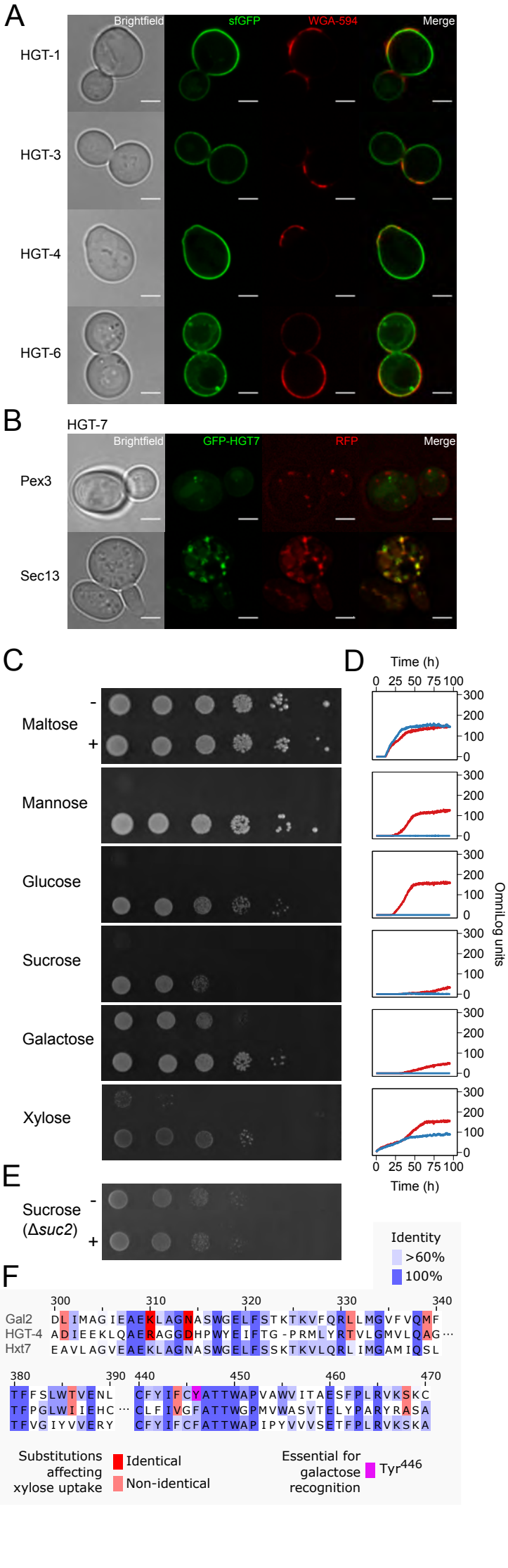
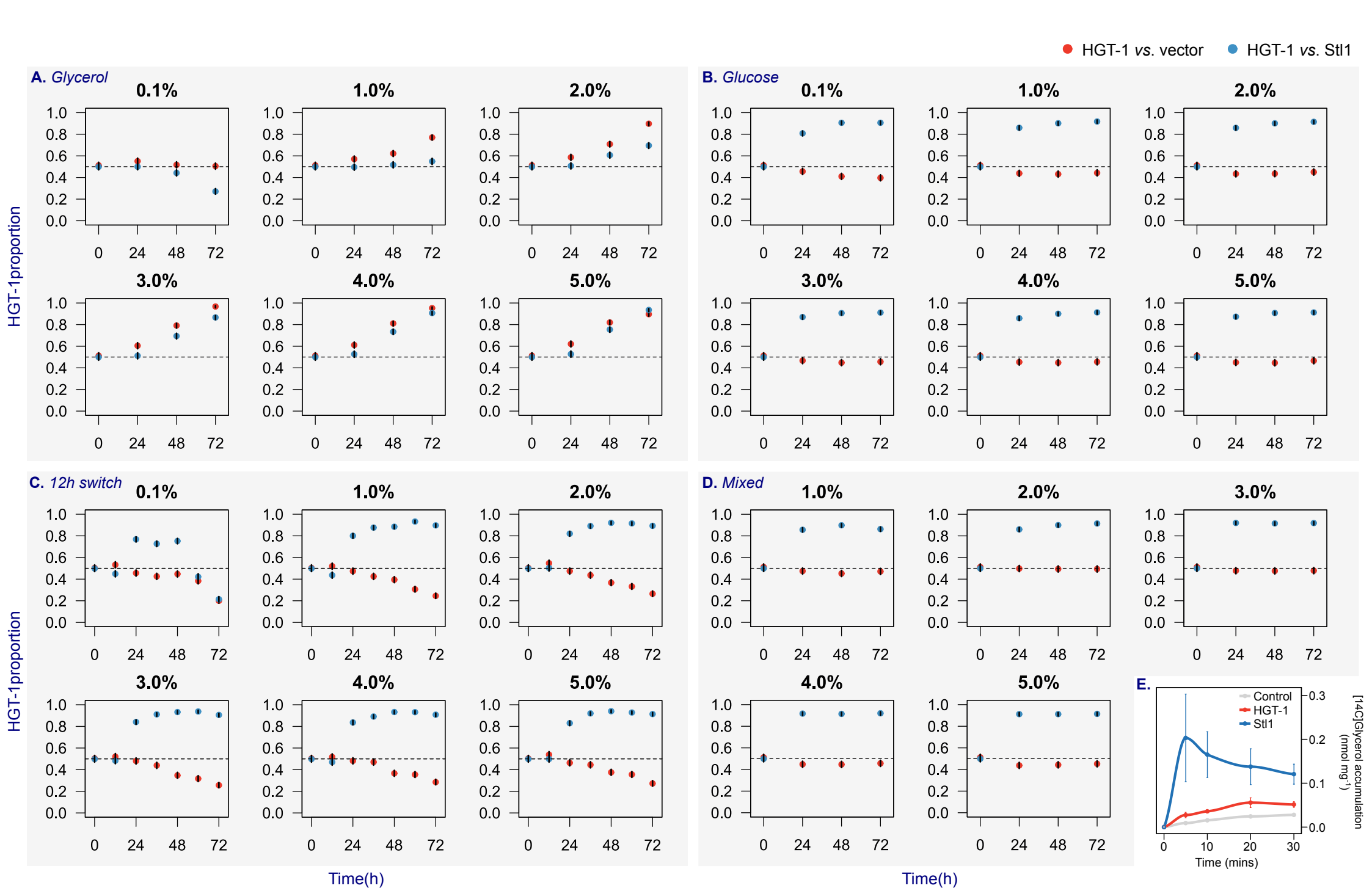
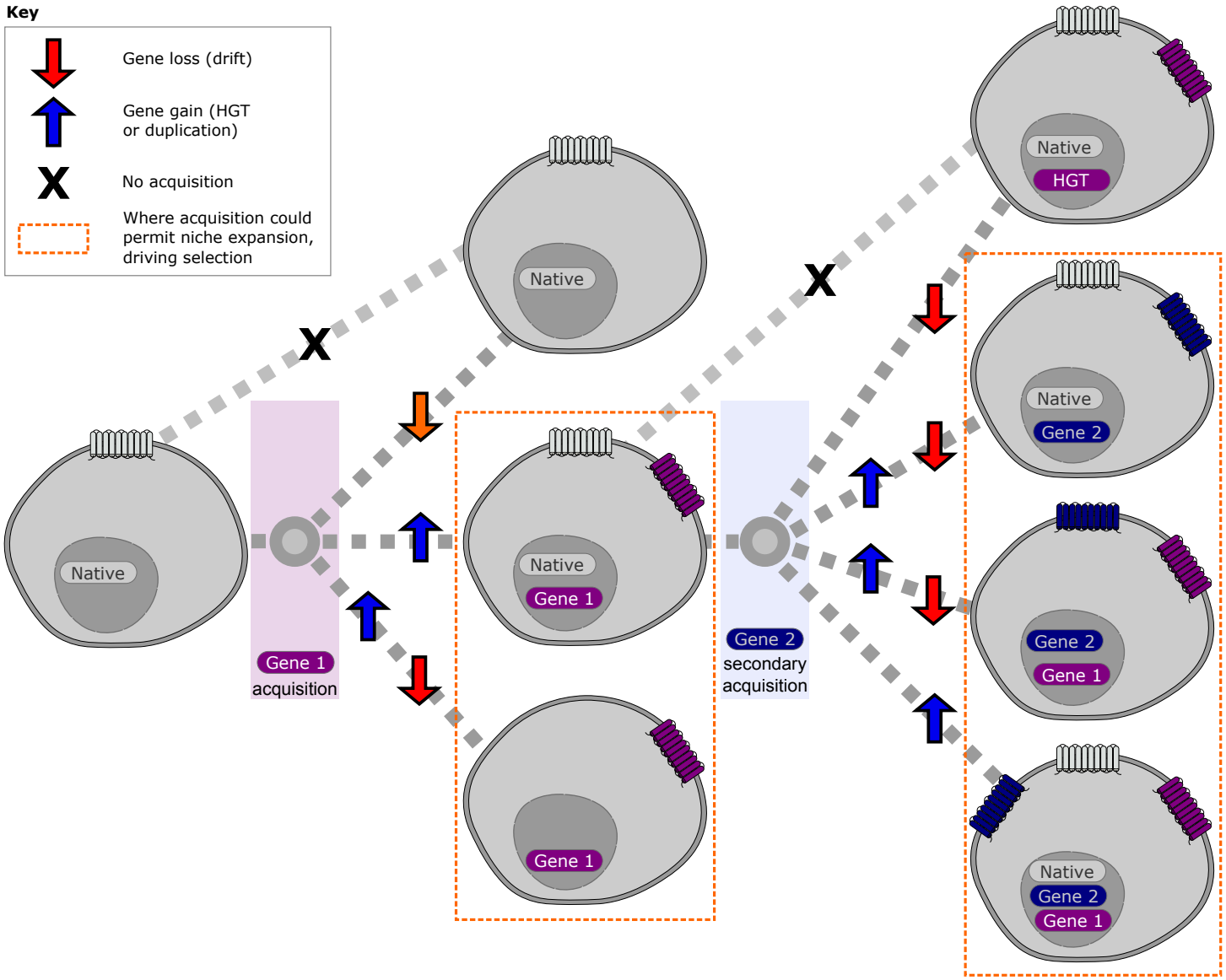

A
Brightfield
sfGFP
WGA-594
Merge
HGT-1
HGT-3
HGT-4
HGT-6
B
HGT-7
Brightfield
GFP-HGT7
RFP
Merge
Pex3
Sec13
C
Maltose
Mannose
Glucose
Sucrose
Galactose
Xylose
D
Time (h)
OmniLog units
E
Sucrose (Δsuc2)
Identity
>60%
100%
F
Gal2
HGT-4
Hxt7
Substitutions affecting xylose uptake
Identical
Non-identical
Essential for galactose recognition
Tyr446

Fig. 4. HGT-1 competitive fitness assays and glycerol uptake kinetics. Proportion of *S. cerevisiae* cells expressing HGT-1 when competed against a p426-GPD vector-only control (red) or against a *S. cerevisiae* strain expressing the native transporter, Stl1 (blue). Experiments were performed over a 72-hour time-period, where glycerol (**A**) or glucose (**B**) were provided as the sole carbon source, when glycerol and glucose were alternated at 12-hour intervals (**C**: '12h switch' between substrates), or when a combination of glycerol and glucose (**D**: 'mixed') were provided as the carbon source. Points represent the mean proportions of live cells from 3 replicates; error bars represent standard errors, obtained from a generalized linear model with binomial error distribution. Horizontal dashed lines correspond to starting proportions (0.5), when genotypes were present at equal frequencies. **E.** ^{14}C glycerol accumulation of *S. cerevisiae* with either empty vector (p426-GPD) or HGT-1, Stl1 vectors. Greater, and more rapid, accumulation was observed for the strain expressing *stl1* than for *HGT-1*.



Milner et al., Fig. 4

Fig. 5. Schematic figure showing alternative outcomes arising from HGT of genes encoding transporter proteins and demonstrating alternative scenarios that could drive a transporter gene transfer ratchet.


Key
Gene loss (drift)
Gene gain (HGT or duplication)
No acquisition
Where acquisition could permit niche expansion, driving selection
Native
Native
HGT
Native
Native
Gene 1
Native
Gene 2
Native
Gene 1
Gene 1
Gene 1
acquisition
Gene 2
secondary acquisition
Gene 2
Gene 1
Native
Gene 2
Gene 1