Genome doubling enabled the expansion of yeast vesicle traffic pathways

Ramya Purkanti & Mukund Thattai

Vesicle budding and fusion in eukaryotes depend on a suite of protein types, such as Arfs, Rabs, coats and SNAREs. Distinct paralogs of these proteins act at distinct intracellular locations, suggesting a link between gene duplication and the expansion of vesicle traffic pathways. Genome doubling, a common source of paralogous genes in fungi, provides an ideal setting in which to explore this link. Here we trace the fates of paralog doublets derived from the 100-Ma-old hybridization event that gave rise to the whole genome duplication clade of budding yeast. We find that paralog doublets involved in specific vesicle traffic functions and pathways are convergently retained across the entire clade. Vesicle coats and adaptors involved in secretory and early-endocytic pathways are retained as doublets, at rates several-fold higher than expected by chance. Proteins involved in later endocytic steps and intra-Golgi traffic, including the entire set of multi-subunit and coiled-coil tethers, have reverted to singletons. These patterns demonstrate that selection has acted to expand and diversify the yeast vesicle traffic apparatus, across species and time.
Results

Ancestral paralog doublets in the yeast whole genome duplication clade. Species belonging to the yeast WGD clade are descended from an interspecies hybrid that underwent a genome doubling event13 (Figs. 1A, 4). The resulting cell initially contained two distinct copies of each gene (one from each parental species) which we term paralog doublets. We obtained high-confidence paralog assignments in 12 members of the yeast WGD clade from the Yeast Genome Order Browser17 (ygob.ucd.ie; “Methods”). This dataset uses conserved gene order (synteny) to identify paralogs derived via the founding hybridization event of the WGD clade, distinguishing these from other homologous copies of genes within each genome which may have arisen via earlier or later duplication events. For the remainder of our analysis we focus on the 4866 genes that we define as the ancestral paralog doublet set (“Methods”). Operationally, these are genes whose orthologs are found in the WGD clade as well as in the ZT and KLE clades that represent the closest living relatives of the two species involved in the original hybridization13. This definition has good specificity and sensitivity: with high confidence, the ancestral doublet set corresponds to genes present in two paralogous copies in the hybrid ancestor of the WGD clade. Over time, one or both copies may be lost18. Among ancestral doublets, 84% are present in at least one copy in every present-day WGD species, but only ∼10% are retained as doublets in any given species19.
within each sub-clade (Fig. 1C, bottom). In a contingent scenario, the pattern arises purely due to shared ancestry, between contingent and convergent deletions as follows (Fig. 1D). We pick a sub-clade, and only look at paralogs we find (Fig. 1D) that they are highly distinct (\( p = 2.4E^{-15} \), Fisher’s exact test; Fig. 1B). The observed doublet retention correlation could be contingent on history, the result of very early losses prior to the divergence of the sub-clades (Fig. 1C, top); or it could be convergent, the result of multiple later losses within each sub-clade (Fig. 1C, bottom). In a contingent scenario, the pattern arises purely due to shared ancestry, and is not connected to gene function. In a convergent scenario, the pattern arises because homologous genes have similar loss rates across different lineages. In this case, we expect a correlation in doublet loss between the sub-clades, conditioned on the doublet still being present in their last common ancestor. We can distinguish between contingent and convergent deletions as follows (Fig. 1D). We pick a sub-clade, and only look at paralogs present as doublets in at least one member of that sub-clade; this enforces the condition that paralogs must be present as doublets in the last common ancestor of both sub-clades. For each paralog, we count the number of doublet species in this set play important roles in single copy, but that having paralog doublets is not typically advantageous\(^{20,21}\).

**Selection drives convergent retention of doublets across species.** Paralog doublets may undergo neo-functionalization (where one copy takes on a non-ancestral function) or sub-functionalization (where the ancestral function is split between the two copies). Once either of these events occurs, selection would favour doublet retention\(^{22–25}\). We can detect signals of such selection by comparing evolutionary trajectories across multiple species. Since we are interested in long-term evolutionary patterns, we compared doublet occurrence in the two most distantly-related sub-clades of the WGD clade, which we refer to as the L and R sub-clades (Fig. 1A,B; “Methods”). If a paralog is present as a doublet in any member of a sub-clade, we can infer it was present as a doublet at the root of that sub-clade (Fig. 1C, bold red lines). If a paralog is present as a singleton in every member of a sub-clade, we cannot draw any conclusions: one copy may have been deleted early, prior to the divergence of the sub-clade; or deleted later, independently in every member of the sub-clade (Fig. 1C, light pink lines). Genes in the ancestral doublet set can be separated into four groups based on how they are retained in present-day species (Fig. 1B): those present as doublets in both sub-clades (L+R), those present as doublets in only one or the other sub-clade (L, R) or those present as singletons everywhere. Within the full gene set, doublet presence or absence is strongly correlated between the sub-clades \( p = 2.4E^{-15} \), Fisher’s exact test; Fig. 1B, left).

The observed doublet retention correlation could be contingent on history, the result of very early losses prior to the divergence of the sub-clades (Fig. 1C, top); or it could be convergent, the result of multiple later losses within each sub-clade (Fig. 1C, bottom). In a contingent scenario, the pattern arises purely due to shared ancestry, and is not connected to gene function. In a convergent scenario, the pattern arises because homologous genes have similar loss rates across different lineages. In this case, we expect a correlation in doublet loss between the sub-clades, conditioned on the doublet still being present in their last common ancestor. We can distinguish between contingent and convergent deletions as follows (Fig. 1D). We pick a sub-clade, and only look at paralogs present as doublets in at least one member of that sub-clade; this enforces the condition that paralogs must be present as doublets in the last common ancestor of both sub-clades. For each paralog, we count the number of doublet species in this sub-clade. Finally, we split these paralogs into two groups, depending on whether or not they are present as doublets in some member of the other sub-clade. If the doublet enrichment pattern is purely contingent, doublet species counts should be statistically indistinguishable between these two groups. Instead we find (Fig. 1D) that they are highly distinct \( p = 8.3E^{-16} \) for L sub-clade counts, \( p = 4.2E^{-6} \) for R sub-clade counts, Kolmogorov–Smirnov test). This rules out pure contingency, and implies convergent selection: some types of genes are more likely to be retained as doublets, and others more likely to revert to singletons, with losses occurring independently across species (Fig. 1C, bottom).

**Vesicle traffic genes have the highest fold enrichment among doublets.** Our analysis is consistent with previous work showing that doublet retention is correlated with function\(^{19–25,27}\). To explore this connection in an unbiased manner, we examined the GO categories enriched among the 887 \( S.\) cerevisiae genes belonging to the L+R doublet set (“Methods”; Table 1; note that some L+R doublets are singletons in \( S.\) cerevisiae). Among the top GO categories ranked by statistical significance, ‘endocytosis’ has by far the highest fold enrichment \( 3.14 \times 10^{-16} \), (with the super-category of ‘vesicle mediated transport’ also featuring on the list). Remarkably, this fold enrichment is even higher than that of the ribosomal genes typically presented as an extreme example of paralog retention\(^{28}\). The same result holds true across the entire WGD clade: even accounting for the genome-wide correlation of paralog doublets across species, vesicle traffic genes are significantly over-represented among the L+R set \( 0.18 \) compared to \( 0.11 \); \( p = 5.2E^{-6} \), Fisher’s exact test; Fig. 1B).

To further explore the role of function in doublet retention, we grouped genes into classes and modules. We define a module as a set of genes whose protein products act collectively to carry out specific vesicle traffic...
functions at specific cellular locations\(^1\). We manually assigned these genes to seven functional classes (Fig. 2, left) based on annotations from the *Saccharomyces* Genome Database\(^2\) ("Methods"). We further sub-divided the Coat/Adaptor class into seven pathway-specific modules.

Out of 360 ancestral vesicle traffic doublets, we assigned 204 to classes and modules. Many of the remaining 156 genes play regulatory roles. Within each class or module, we asked how many paralogs were retained as doublets in both sub-clades (L+R doublets), and compared this to the expected number given the ~18% (66/360) rate of L+R doublets among all vesicle traffic genes. Among functional classes, coat/adaptor genes and lipid control genes were enriched for L+R doublets; and tethers and ESCRT genes had no L+R doublets. Among coat/adaptor modules, ER to Golgi traffic genes and PM to EE/TGN traffic genes were enriched for L+R doublets; and intra-Golgi traffic genes had no L+R doublets. Only four of these cases were statistically significant (Fisher’s exact test, Benjamini–Hochberg correction, FDR = 0.05; "Methods"): ER to Golgi traffic (4.9x enrichment, raw \(p = 6.6 \times 10^{-6}\)); coats/adaptors (2.5x enrichment, raw \(p = 1.9 \times 10^{-4}\)); lipid control genes (3.1x enrichment, raw \(p = 4.7 \times 10^{-3}\)); and tethers (0.0x enrichment, raw \(p = 7.9 \times 10^{-3}\)).

**Doublets are retained in secretory and early-endocytic pathways.** We next considered the impact of paralogous modules in the context of the global yeast vesicle traffic system (Fig. 3A). Traffic pathways can be...
broadly classified into secretory components, from the ER via the Golgi and trans-Golgi network (TGN) to the plasma membrane (PM); and endocytic components, from the plasma membrane via early endosomes (EE) and late or pre-vacuolar endosomes (LE/PVE) to the vacuole. In *S. cerevisiae* the EE and TGN compartments appear to significantly overlap, serving as transit points during both secretion and endocytosis. Given these and other ambiguities about the sites of action of vesicle traffic proteins, it is difficult to formulate and statistically test hypotheses about whether paralogs involved in specific pathways are more likely to be retained as doublets. Nevertheless, the following patterns are suggestive of general principles.

Every step of secretion from the ER to the plasma membrane involves paralogous L+R doublets. At the ER to Golgi step, multiple components of the COPII coat and its adaptors are L+R doublets. Within the Golgi, the master regulator ARF1/ARF2 is an L+R doublet, along with many Arf modulators involved in anterograde traffic such as GEA1/GEA2. At the TGN, cargo adaptors, including exomer which regulates TGN to PM export and GGA which regulates traffic to the PVE, are L+R doublets, along with

---

**Figure 3.** Landscape of vesicle traffic evolution in the WGD clade. (A) We show the site of action of proteins within the yeast vesicle traffic network. Filled boxes are proteins or complexes corresponding to genes present as L+R doublets. Dotted boxes show selected complexes for which the majority of proteins are present as singletons: the COPII coat and AP adaptors, tethers, and ESCRT. (B,C) We separate the 360 ancestral vesicle traffic doublets into two groups: those that are present as doublets in at least one species (dark blue) and those that are not doublets in any species (light blue). For each group, the curves show the cumulative distribution of two quantities. P-value calculations are described in the main text and “Methods”. (B) Inferred sequence identity between ZT and KLE orthologs (in *Z. rouxii* and *K. lactis*) corresponding to each ancestral doublet. (C) Number of physical interaction partners of the protein corresponding to each ancestral doublet.
components of the clathrin coat. At the plasma membrane, the v/t-SNARE complex comprising SNC1/SNC2 and SEC9, which drives fusion of secretory vesicles to the PM, are L+R doublets. L+R doublets are also involved in early endocytic steps: early and intermediate clathrin coat proteins PAL1/PAL2 and ENT1/ENT2 are L+R doublets, along with components of the endocytic regulator arrestin. Two components of retromer, an adaptor which regulates cargo flow from the LE to the TGN, are L+R doublets.

In contrast to the above cases, several modules have completely reverted to singletons. These include: the COPI coat and all components of the AP adaptor complex; the entire set of coiled-coil and multi-subunit tethers; and the ESCRT complex. With the exception of the exocyst complex and the AP-2 adaptor complex, which both act at the plasma membrane, the remaining singletons are all involved in retrograde Golgi traffic and late endocytic steps. The TRAPPi tether participates in ER-to-Golgi transport. The COPI and AP-1 coats, along with tethers GARP, COG, TRAPPII and DSL1, facilitate intra-Golgi cycling and Golgi-to-ER transport. The tethers CORVET and HOPS are involved in late endosomal and vacuolar dynamics, along with the ESCRT complex which remodels late endosomal membranes.

Retained doublets have lower evolutionary rates and fewer protein interactions. Cross interactions between paralogous modules are common in newly formed yeast hybrids, even when parental species have diverged over 500 million years. Tightly-interacting modules may be subject to dominant negative effects due to mutations in their paralogous partners, suggesting doublets involving highly interacting proteins are more likely to revert to singletons. However, it is also known that highly interacting proteins have lower evolutionary rates, and in turn, lower evolutionary rates are correlated with lower rates of gene loss. We sought to understand which of these two effects dominates.

As a proxy for the evolutionary rate of each ancestral doublet, we used the nucleotide sequence identity between the corresponding orthologs in present-day members of the ZT and KLE clades (Methods; this avoids the confounding effect of evolutionary rate variation between singletons and doublets in WGD clade species). We imputed a physical interaction network among the proteins encoded by ancestral doublets, using present-day interaction data for the corresponding proteins in S. cerevisiae (Methods). We separated all 360 ancestral vesicle traffic doublets into two groups: those present as pure singletons across all present-day species, and those present as doublets in at least one present-day species. We then compared the distributions of evolutionary rates and protein interaction degrees between these two groups (Fig. 3B,C). We find that doublet retention is strongly associated with lower evolutionary rates (higher sequence identity; $p = 3.3 \times 10^{-6}$, Kolmogorov–Smirnov test; Fig. 3B). In contrast, doublet retention is only weakly associated with fewer protein interactions ($p = 2.6 \times 10^{-3}$, Kolmogorov–Smirnov test; Fig. 3C). This is consistent with prior observations: cross-interactions after gene duplication are weakly correlated with fitness, due to compensatory mechanisms such as expression attenuation; in contrast, low evolutionary rates are strongly correlated with low gene loss because functionally important genes tend to be under purifying selection. Taken together, these data reinforce our finding that doublet retention is driven by selection for specific function.
Discussion

Genome doubling is a recurring theme in eukaryotic evolution. These events provide many opportunities for selection to act, and can reveal evolutionary pressures that are invisible under normal circumstances. In this study we have taken advantage of an ancient genome doubling event to rigorously demonstrate signatures of such selection on the yeast vesicle traffic system.

Interspecies hybridization is a common route to genome doubling among fungi. Interspecies diploids are typically sterile, since mismatches between homologous chromosomes stall meiosis; genome doubling spontaneously restores fertility in such hybrids, by creating an allotetraploid cell with two identical copies of each chromosome. The alleles of the original hybrid diploid become paralog doublets of the allotetraploid. Paralogs are always at risk of being lost due to gene conversion, which occurs when homologous template sequences are used to repair double-strand breaks. Newly-formed hybrid allotetraploid genomes typically contain both homeologs (pairs derived from both parents) and ohnologs (pairs tracing to a single parent, due to pre-WGD gene conversion). Gene conversion rapidly erases variations between ohnologs but spares the more diverged homeologs, since double-strand break repair in allotetraploids uses alleles as templates and copies of multiple CORVET components VPS16 and VPS33. Paralogs may have distinct interaction partners, different cellular locations, and may operate at highly distinct sub-cellular locations. This suggests that the architecture of vesicle traffic in present-day eukaryotes is tightly constrained, and that the genome doubling route we have explored is distinct from more ancient duplication routes. It is likely the major vesicle traffic gene families were generated during an earlier, more dynamic and less constrained phase of eukaryotic evolution.

Methods

Ortholog assignments in pre-WGD and post-WGD species.

We downloaded synteny-based ortholog assignments from the Yeast Genome Order Browser. This dataset covers 20 species: 12 within the yeast WGD clade, which we split into two sub-clades for further analysis (L sub-clade: S. cerevisiae, S. mikatae, S. kudriavzevii, S. uvarum, C. glabrata, K. africana, K. lactis, N. castellii, N. dairenensis; R sub-clade: T. blattae, T. phaffii, V. polyspora); and 8 pre-WGD species comprising the ZT clade (Z. rouxii, T. delbrueckii) and the KLE clade (K. lactis, E. gossypii, C. cibulariae, L. kluveri, L. thermotolerans, L. waltii). The WGD clade is descended from an interspecies hybridization between two species whose closest living relatives are inferred to belong to the ZT clade and the KLE clade, respectively. A total of 14101 orthologs are present across all 20 species in the dataset. A subset of 11059 orthologs are found in every member of the ZT and KLE clades, and were therefore likely to be inherited as doublets in the WGD ancestor.

Defining the ancestral paralog doublet set.

We are interested in orthologs that were present as paralog doublets immediately following the original interspecies hybridization. By definition, one copy of each such gene is inherited from each parent. However, we do not know the true genetic complement of the parental species, only that of their closest living relatives. Operationally, we define the set of ancestral doublets as the set of 4866 genes found across the ZT, KLE and WGD clades. This definition has good specificity (3891 genes in this set are present in every member of the ZT and KLE clades, and were therefore likely to be inherited as doublets in the WGD ancestor) and good sensitivity (only 48 out of the 1374 genes present as doublets in any present-day WGD species are missing from this set). 4075 out of 4866 genes are present in at least one copy in every present-day WGD species. The full list of ancestral doublets is provided in Supplementary Table S1.
Classifying doublets as homeologs and ohnologs. The time of duplication of paralog doublets has been estimated using phylogenetic methods, as described in Ref. 13. We obtained supporting data associated with this study. The duplication event is assigned to a branch of a species tree spanning the KLE, ZT and WGD clades, as defined in Fig. 1 of Ref. 13. Each doublet is associated with two inferred duplication branches, based on the phylogenetic trees seeded by each doublet member. In the event that the two branches do not match, we retained the branch with higher support. Those with support below 0.95 were not considered. In this way, we could assign the duplication branch for 60% (377/620) of S. cerevisiae paralog doublets (Supplementary Table S1). Homeologs correspond to branches < n4 (duplicated prior to WGD) and ohnologs correspond to branches ≥ n5 (duplicated at or after WGD). 63% (239/377) of all doublets and 66% (31/47) of vesicle traffic doublets are homeologs. Duplication branches are listed in Supplementary Table S1.

Annotation of vesicle traffic genes. We assigned genes to functional categories based on annotations of their S. cerevisiae homologs. 426 S. cerevisiae genes are annotated with the Gene Ontology term GO:0016192 ‘vesicle-mediated transport’74 (implemented via PANTHER Version 16.0; pantherdb.org). To these we added 17 S. cerevisiae homologs. 426 S. cerevisiae of their genes whose paralogs were already part of the set. This resulted in 443 genes (323 singletons and 60 doublets in S. cerevisiae) of which 360 are present in the ancestral paralog doublet set (300 singletons and 60 doublets in S. cerevisiae). We used annotations from the Saccharomyces Genome Database29 (yeastgenome.org) to assign 236 out of 443 genes (204 out of 360 ancestral vesicle traffic doublets) to seven functional classes: Coat/Adaptor; Lipid control; Arf/Rab; GAP/GEF/GDI; SNARE; Tether; and ESCRT. We sub-divided the Coat/Adaptor class into seven pathway-specific modules74 for ‘GO Biological Process Complete’ functional categories, with the yeast genome as the background. We sorted hits by the False Discovery Rate (Benjamini–Hochberg procedure, overall FDR < 0.05). The top 12 hits are shown in Table 1, the top 20 hits and full details shown in Supplementary Table S2. We performed other enrichment analyses (Figs. 1B, 2) using the two-tailed Fisher’s exact test on 2 × 2 contingency tables. When testing for enrichment among the 14 vesicle traffic gene classes/modules (Fig. 2), we additionally applied the Benjamini–Hochberg correction for multiple hypothesis testing with a false discovery rate α = 0.05, to determine the significance threshold. For comparing between distributions (Figs. 1D, 3B,C) we used the Kolmogorov–Smirnov (KS) test. Note that the KS test reports a conservative p-value when applied to discrete distributions. Data and p-values for each test are provided in Supplementary Table S3.

Enrichment analysis and statistical tests. We identified 887 genes in S. cerevisiae that corresponded to L+R doublets. On this gene list we carried out an overrepresentation test using PANTHER (PANTHER Version 17.0; pantherdb.org)34 for ‘GO Biological Process Complete’ functional categories, with the yeast genome as the background. We sorted hits by the False Discovery Rate (Benjamini–Hochberg procedure, overall FDR < 0.05). The top 12 hits shown in Table 1, the top 20 hits and full details shown in Supplementary Table S2. We performed other enrichment analyses (Figs. 1B, 2) using the two-tailed Fisher’s exact test on 2 × 2 contingency tables. When testing for enrichment among the 14 vesicle traffic gene classes/modules (Fig. 2), we additionally applied the Benjamini–Hochberg correction for multiple hypothesis testing with a false discovery rate α = 0.05, to determine the significance threshold. For comparing between distributions (Figs. 1D, 3B,C) we used the Kolmogorov–Smirnov (KS) test. Note that the KS test reports a conservative p-value when applied to discrete distributions. Data and p-values for each test are provided in Supplementary Table S3.

Protein interaction analysis. The 360 ancestral vesicle traffic genes correspond to 420 S. cerevisiae genes (300 singletons and 60 doublets). For these genes we obtained the protein–protein interaction network from the STRING database40 (string-db.org), filtering for the physical sub-network at medium confidence, with experiments and databases as interaction data sources. For genes present as doublets, we assumed an interaction between a pair of ancestral genes if there was an interaction between any of their paralogs, as would be expected based on a sub-functionalization scenario. Note that this systematically increases the inferred interaction degree of doublets. Even so, we find that doublets overall have fewer interactions than singletons (Fig. 3C). Interaction data are provided in Supplementary Table S1.

Evolutionary rate analysis. To estimate the evolutionary rate of ancestral paralog doublets, we examined the sequence identity between the corresponding orthologs in Z. rouxii and K. lactis (representative species for the ZT and KLE clade). We used the YGOB database (Version 7; ygob.ucd.ie) to assign orthologs. Of the 4866 ancestral doublets, 4585 had orthologs in Z. rouxii and K. lactis (mean ± SEM). Sequence identity values are provided in Supplementary Table S1.

Data availability All data generated or analysed during this study are included in this published article and its Supplementary Information files.

Received: 31 January 2022; Accepted: 23 June 2022
Published online: 02 July 2022

References
1. Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. Cell 116, 153–166. https://doi.org/10.1016/s0092-8674(03)01079-1 (2004).
2. Ramadas, R. & Thattai, M. New organelles by gene duplication in a biophysical model of eukaryote endomembrane evolution. Biophys. J. 104, 2533–2563. https://doi.org/10.1016/j.bpj.2013.03.066 (2013).
3. Mali, P. D., Barlow, L. D., Rachubinski, R. A. & Dacks, J. B. Evolutionary mechanisms for establishing eukaryotic cellular complexity. Trends Cell Biol. 24, 435–442. https://doi.org/10.1016/j.tcb.2014.02.003 (2014).
43. Ortiz-Merino, R. A. et al. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol.* 15, e2002128. https://doi.org/10.1371/journal.pbio.2002128 (2017).

44. Fisher, K. J., Buskirk, S. W., Vignonna, R. C., Marad, A. A. & Lang, G. I. Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLoS Genet.* 14, e1007396. https://doi.org/10.1371/journal.pgen.1007396 (2018).

45. Libkkind, D. et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci.* 108, 14539–14544. https://doi.org/10.1073/pnas.1109430108 (2011).

46. Okano, M. et al. Next-generation sequencing analysis of lager-brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res.*, dsv037. https://doi.org/10.1093/dnares/dsv037 (2016).

47. Albertin, W. & Marullo, P. Polyolipoyd in fungi: Evolution after whole-genome duplication. *Proc. R. Soc. B Biol. Sci.* 279, 2497–2509. https://doi.org/10.1098/rspb.2012.0343 (2012).

48. Morales, L. & Dujon, B. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiol. Mol. Biol. Rev.* 76, 721–739. https://doi.org/10.1128/mmbre.00022-12 (2012).

49. Gabaldón, T. Hybridization and the origin of new yeast lineages. *FEMS Yeast Res.* 20, foa040. https://doi.org/10.1093/femsyr/foa040 (2020).

50. Chen, J.-M., Cooper, D. N., Chuzhanova, N., Férec, C. & Patrinos, G. P. Gene conversion: Mechanisms, evolution and human disease. *Nat. Rev. Genet.* 8, 762–775. https://doi.org/10.1038/nrg2193 (2007).

51. Louis, V. L. et al. *Pichia sorbitophila*, an interspecies yeast hybrid, reveals early steps of genome resolution after polyplidization. *G3 Genomes Genomes Genet.* 2, 299–311. https://doi.org/10.1534/g3.111.000745 (2012).

52. Heil, C. S. S. et al. Loss of heterozygosity drives adaptation in hybrid yeast. *Mol. Biol. Evol.* 34, 1596–1612. https://doi.org/10.1093/molbev/msx098 (2017).

53. Conant, G. C. & Wolfe, K. H. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Syst. Biol.* 3, 129. https://doi.org/10.1038/msb4100170 (2007).

54. Sato, K. & Nakano, A. Emp47p and its close homolog emp46p have a tyrosine-containing endoplasmic reticulum exit signal and function in glycoprotein secretion in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 13, 2518–2532. https://doi.org/10.1090/mc02-010027 (2002).

55. Pakzad, J. E. & Fromme, J. C. Structural basis for membrane binding and remodeling by the exomer secretory vesicle cargo adaptor. *Dev. Cell* 30, 610–624. https://doi.org/10.1016/j.devcel.2014.07.004 (2016).

56. Aguilar, R. C. et al. Epsin n-terminal homology domains perform an essential function regulating cd42 through binding cd42 GTPase-activating proteins. *Proc. Natl. Acad. Sci.* 103, 4116–4121. https://doi.org/10.1073/pnas.0510513103 (2006).

57. Conant, G. C. & Wolfe, K. H. Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Genet.* 8, e1002760. https://doi.org/10.1371/journal.pgen.1002760 (2012).

58. Ahn, S. et al. Tetrahymena thermophila: An ancient model eukaryote for motivation and discovery. *Curr. Biol.* 20, 11521–11528. https://doi.org/10.1016/j.cub.2010.11.010 (2010).

59. Ishov, V. et al. Complementary a-arrestin-ubiquitin ligase complexes control nutrient transporter endocytosis in response to amino acids. *eLife* 9, e58246. https://doi.org/10.7554/elife.58246 (2020).

60. Thompson, D. A. et al. Evolutionary principles of modular gene regulation in yeasts. *eLife* 2, e00603. https://doi.org/10.7554/elife.00603 (2013).

61. Heineke, B. M. & El-Samad, H. Paralogs in the PKA regulon traveled different evolutionary routes to divergent expression in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 30, 3005–3028. https://doi.org/10.1090/mbc.e19-02-0201 (2020).

62. Gustafson, K. M. & Fromme, J. C. Regulation of arf activation occurs via distinct mechanisms at early and late Golgi compartments. *Mol. Biol. Cell* 28, 3660–3671. https://doi.org/10.1091/mbc.e17-06-0370 (2017).

63. O’Donnell, A. F., Apffel, A., Gardner, R. G. & Cyert, M. S. Alpha-arrestins aly1 and aly2 regulate intracellular trafficking in response to nutrient signaling. *Mol. Biol. Cell* 21, 3552–3566. https://doi.org/10.1091/mbc.e10-07-0636 (2010).

64. Moorby, B. T., Sharma, A., Roettiger, D. R., Wilson, T. E. & Lemmon, S. K. Identification of suppressor of clathrin deficiency-1 (SCD1) and its connection to clathrin-mediated endocytosis in *Saccharomyces cerevisiae*. *G3 Genomes Genomes Genet.* 9, 867–877. https://doi.org/10.1534/g3.118.200782 (2019).

65. Casler, J. C. & Glick, B. S. A microscopy-based kinetic analysis of yeast vacuolar protein sorting. *eLife* 9, e56844. https://doi.org/10.7554/elife.56844 (2020).

66. Bech, C. T., Cool, L., Phillips, J. & Rine, J. Overlapping functions of the yeast oxysterol-binding protein homologues. *Genetics* 157, 1117–1140 (2001).

67. Sparrvoll, D., Zoltner, M., Cheng, C.-Y., Field, M. C. & Turkewitz, A. Diversification of CORVET tethers facilitates transmembrane complexity in *Zyazthromonas thermophilus*. *J. Cell Sci.* 133, jcs238659. https://doi.org/10.1242/jcs.238659 (2020).

68. Sperl, A. Membrane tethering complexes in the endosomal system. *Front. Cell Dev. Biol.* 4. https://doi.org/10.3389/fcell.2016.00035 (2016).

69. Mi, H. et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49, D394–D403. https://doi.org/10.1093/nar/gkaa116 (2020).

Acknowledgements

We thank the participants of the 2015 KITP program on Evolutionary Cell Biology for discussions which nucleated this project. We thank Andrew Murray for prompting us to consider the cell-biological context of genome doubling. We thank Jitu Mayor, Sunil Laxman, Benjamin Glick and Christian Landry for advice and feedback. We give special thanks to our families: without their patience and support through a year filled with challenges, this project could never have been completed. This work was supported by the Simons Foundation (287975 to MT).

Author contributions

M.T. and R.P. designed the study. R.P. performed the bioinformatic analysis, M.T. performed the statistical analysis. M.T. wrote the paper.
Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-15419-9.

Correspondence and requests for materials should be addressed to M.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022