

Starship giant transposable elements cluster by host taxonomy using *k*-mer-based phylogenetics

Rowena Hill ^{1,*} Daniel Smith ^{2,4} Gail Canning ³ Michelle Grey ¹ Kim E. Hammond-Kosack ³ Mark McMullan ^{1,*}

¹Earlham Institute, Norwich, Norfolk NR4 7UZ, UK

²Intelligent Data Ecosystems, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

³Protecting Crops and the Environment, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

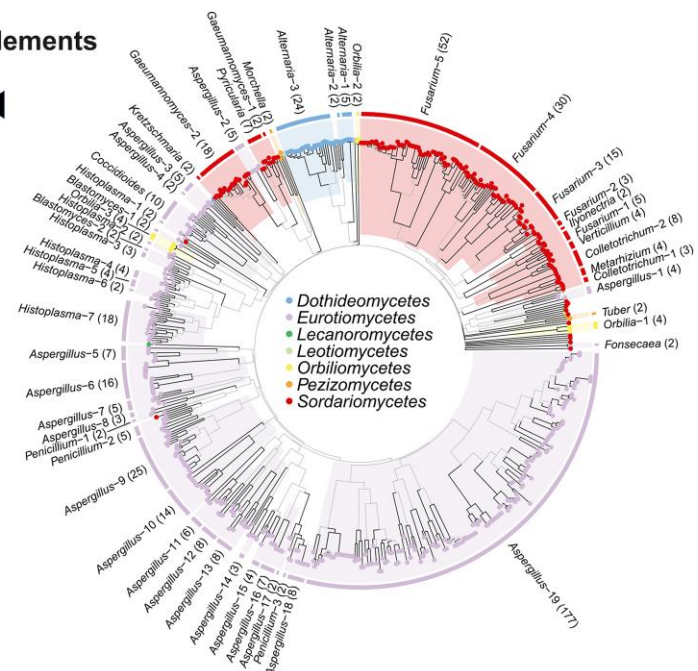
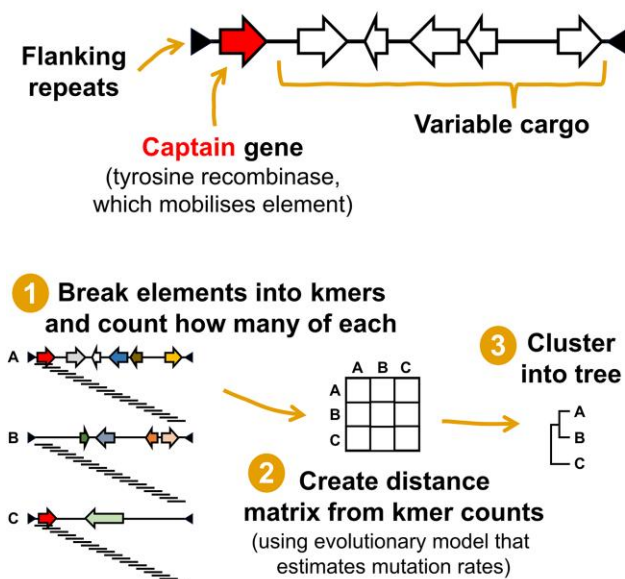
⁴Present address: Department of Computational and Systems Biology, John Innes Centre, Norwich, Norfolk, NR4 7UH, UK

*Corresponding authors: Rowena Hill, Earlham Institute, Norwich, Norfolk NR4 7UZ, UK. Email: Rowena.Hill@earlham.ac.uk; Mark McMullan, Earlham Institute, Norwich, Norfolk NR4 7UZ, UK. Email: Mark.McMullan@earlham.ac.uk

Starships are a recently established superfamily of giant cargo-mobilizing transposable elements in the fungal subphylum *Pezizomycotina* (phylum *Ascomycota*). To date, *Starship* elements have been identified up to ~700 kbp in length and carry hundreds of accessory genes, which can confer both beneficial and deleterious traits to the host genome. Classification of *Starship* elements is centered on the tyrosine recombinase gene that mobilizes the element, termed the captain. We contribute a new perspective to *Starship* relatedness by using an alignment-free *k*-mer-based phylogenetic tree-building method, which can infer relationships between elements in their entirety, including both active and degraded elements and irrespective of high variability in element length and cargo content. In doing so we found that relationships between entire *Starships* differed from those inferred from captain genes and revealed patterns of element relatedness corresponding to host taxonomy. Using *Starships* from root/soil-dwelling *Gaeumannomyces* species as a case study, we found that *k*-mer-based relationships correspond with the similarity of cargo gene content. Our results provide insights into the prevalence of *Starship*-mediated horizontal transfer events. This novel application of a *k*-mer-based phylogenetics approach overcomes the issue of how to represent and compare highly variable *Starship* elements as a whole, and in effect shifts the perspective from a captain to a cargo-centered concept of *Starship* identity.

Graphical Abstract

Starship giant cargo-mobilising transposable elements



Keywords: cargo-mobilizing elements; *Gaeumannomyces*; *Ascomycota*; *Pezizomycotina*; Fungi

Received on 10 March 2025; accepted on 06 April 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Transposable elements (TEs), or transposons, are stretches of DNA, typically between 100 and 10,000 bp in length, which can independently move and replicate within the genome (Biémont 2010; Wells and Feschotte 2020). Thanks to advances in long-read sequencing, highly contiguous genome assemblies have revealed the existence of TEs hundreds of kilobases in length (Arkhipova and Yushenova 2019). Some of these large TEs have been shown to harbor both genes necessary for their mobilization as well as miscellaneous accessory genes, and are accordingly referred to as cargo-mobilizing elements (CMEs; Gluck-Thaler and Vogan 2024). Recently, giant CMEs have been found in various species in the fungal subphylum *Pezizomycotina* (phylum *Ascomycota*; McDonald et al. 2019; Vogan et al. 2021; Urquhart et al. 2022), and have since been determined to belong to a newly established TE “superfamily” (sensu Wicker et al. (2007)) or “subclass” (sensu Wells and Feschotte (2020)) known as “Starships” (Gluck-Thaler et al. 2022). To date, *Starship* CMEs have been found to range in length from 15 Kbp (Gluck-Thaler et al. 2024) to ~700 kbp (Urquhart et al. 2024).

Starship mobilization is mediated by a leading 5' located gene containing the DUF3435 domain (protein family accession PF11917), termed the “captain,” which encodes a tyrosine recombinase that initiates movement of the TE into a new genomic location via a “cut-and-paste” mechanism (Urquhart et al. 2023). This is similar to the hypothesized mobilization process of the “Crypton” class II DNA transposon superfamily (Wells and Feschotte 2020), which was incidentally also first discovered in fungi (Goodwin et al. 2003), although this TE superfamily has since been found in other eukaryotes (Kojima and Jurka 2011). Tyrosine recombinase domains in *Starship* captain genes and *Cryptons* are very distantly related (Gluck-Thaler et al. 2022) and, unlike *Cryptons*, *Starship* elements are sometimes flanked by tandem inverted repeats (TIRs) in addition to direct repeats (DRs), and can contain a highly variable and often sizeable cargo of accessory genes (Gluck-Thaler and Vogan 2024). *Starship* cargos can harbor genes that are beneficial to the fungus, for example those associated with plant virulence (McDonald et al. 2019), metal tolerance (Urquhart et al. 2022), and climate adaptation (Tralamazza et al. 2024). However, as selfish genetic elements, *Starships* may also mobilize cargo that is neutral or even detrimental to the overall fitness of the host genome (Vogan et al. 2021).

Classification within the *Starship* CME superfamily is focused on the captain gene, using both phylogenetic relationships between captain genes to define “family” and ortholog clustering of captain genes to define “navis” (i.e. a ship) (Gluck-Thaler and Vogan 2024). Both the captain family and the flanking DRs are thought to influence the genomic site that an element is inserted into, with *Starships* of certain captain families preferentially inserting into, for instance, other TEs or 5S rDNA (Urquhart et al. 2023; Gluck-Thaler and Vogan 2024). DUF3435-containing tyrosine recombinase genes are more usually found “solo,” rather than within a cargo-carrying element, i.e., as a captain; however, it is not clear to what extent this is due to the failure to detect the boundaries of an element or because pseudogenization of the tyrosine recombinase gene has occurred (Gluck-Thaler and Vogan 2024). *Starship* captain genes do not form a single monophyletic cluster in the DUF3435 tyrosine recombinase gene tree and are instead scattered across the phylogeny amongst other apparently “solo” DUF3435-containing tyrosine recombinase genes (Gluck-Thaler et al. 2022; Hill et al. 2025). Due to their highly divergent nature, tyrosine recombinase gene sequences are also

difficult to align, introducing uncertainty into conventional alignment-based phylogenetic analyses. It is not currently possible to determine whether these relationships described by captains are preserved or representative of the *Starships* as a whole, considering that elements are highly variable in terms of cargo and overall length. This also limits phylogenetic assessment of the prevalence of (or boundaries to) horizontal exchange across the *Pezizomycotina*. In an effort to represent distinction in cargo content, Gluck-Thaler and Vogan (2024) introduced the additional definition of “haplotype,” based on clustering of *k*-mer similarity scores. Here, we have taken this approach 1 step further and used a *k*-mer -based phylogenetic tree-building method to contribute a new perspective to *Starship* relatedness. In doing so, we have revealed previously obscured patterns of *Starship* relatedness corresponding to host taxonomy.

To determine whether the relatedness revealed by the *k*-mer trees conformed with similarity in cargo gene content, we explored the cargos of *Starships* previously identified from genomes within the genus *Gaeumannomyces* (Hill et al. 2025). This genus comprises soil-dwelling fungi that are also both pathogenic and nonpathogenic wheat and wild grass root associates (Palma-Guerrero et al. 2021; Chancellor et al. 2024). These elements provided an ideal case study as they vary greatly in overall size and number of cargo genes within their host taxonomy clusters. The genomes were also all generated in parallel using the same long-read sequencing technology and a cross-referent annotation pipeline (Hill et al. 2025). Given the impact of assembly and annotation quality on *Starship* recovery (Gluck-Thaler and Vogan 2024), these *Gaeumannomyces* elements therefore represent a consistent dataset that is impacted to a lesser extent by the technology used to produce them.

Materials and methods

K-mer -based phylogenetic analysis

To compare phylogenetic reconstruction of whole elements vs captain genes, we used a curated set of 39 *Starships* from Gluck-Thaler et al. (2022) and Gluck-Thaler and Vogan (2024) alongside 14 *Gaeumannomyces* *Starships* predicted using the tool starfish v1.0.0 (Gluck-Thaler and Vogan 2024) in our previous study (Hill et al. 2025). Only *Gaeumannomyces* *Starships* with predicted flanking repeats were used. We used entire element sequences as input for the *k*-mer -based method Mashtree v1.4.6 (Katz et al. 2019) with 1,000 bootstrap replicates and the -min-depth 0 parameter to discard very unique *k*-mers, recommended to improve accuracy. We used the corresponding captain genes as input for a maximum likelihood (ML) tree, first aligning gene sequences using MAFFT v7.271 (Katoh and Standley 2013), trimming using trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009), and finally building the ML tree using RAXML-NG v1.1.0 (Kozlov et al. 2019) with bootstrapping until convergence, which occurred after 150 bootstrap replicates. We visualized concordance between the 2 phylogenies via a tanglegram, produced in R v4.3.1 (R Core Team 2023) using the packages ape v5.7-1 (Paradis and Schliep 2019), phytools v2.1-1 (Revell 2024) and ggtree v3.9.1 (Yu et al. 2017). We calculated the normalized Robinson-Foulds (RF) distance between the element and captain phylogenies using the RF.dist function from the phangorn v2.7.0 package (Schliep et al. 2017).

We then used a larger dataset of *Starships* predicted using the tool starfish v1.0.0 by Gluck-Thaler and Vogan (2024) to assess whether patterns in the curated *k*-mer tree would persist with broader sampling. Comparisons were made using the entire dataset including elements without predicted flanking repeats

(597 elements + 20 *Gaeumannomyces* elements = 617 total) against a filtered dataset of only elements with predicted flanking repeats (343 elements + 14 *Gaeumannomyces* elements = 357 total) to explore the impact of uncertain element boundaries on the topology. For both cases, entire element sequences were again run with Mashtree, but with 100 bootstrap replicates and the default –min-depth parameter to accommodate for the much larger dataset. Previously determined *Starship* family classifications, based on captain phylogenetic relationships (Gluck-Thaler and Vogan 2024), were mapped to element *k*-mer tree tips to visualize the distribution of families across clades using the additional R packages ggtreeExtra v1.10.0 (Xu et al. 2021) and glottoTrees v0.1.10 (Round 2021).

Mashtree estimates similarity between *k*-mer sketches using the Mash distance, which models mutation rates under a simple Poisson process of random site mutation (Ondov et al. 2016). To compare this with an alternative evolutionary model we used sourmash v4.8.14 (Irber et al. 2024) to calculate a distance matrix with the –estimate-ani parameter. Like the Mash distance, average nucleotide identity (ANI) as implemented in sourmash is computed from the Jaccard index, but unlike Mash it does not make the assumption that all *k*-mers are mutated independently, which can result in Mash overestimating mutation rates (Rahman Hera et al. 2023). The *k*-mer sketching algorithm within sourmash, FracMinHash, may also outperform Mash's MinHash algorithm when used on very different set sizes (Rahman Hera et al. 2023). We should caveat that ANI was developed for use with prokaryote data and has not, to our knowledge, been validated with eukaryote data, although this may predominantly be due to scalability issues when working with larger eukaryote genomes. We used the ape nj command in R to generate a neighbor-joining tree from the sourmash ANI distance matrix, which is conceptually the same tree-building approach that is integrated into Mashtree.

Exploration of cargo gene content in *Gaeumannomyces* elements

We used the aforementioned larger dataset of 20 *Starships* predicted from 7 *Gaeumannomyces* genomes to assess whether similarities in cargo gene content corresponded with the patterns of relatedness described by the *k*-mer trees. We characterized orthologous genes predicted in our previous study (Hill et al. 2025) as being core, accessory, or specific within the set of 20 elements, and their sharedness was visualized using the R package ComplexUpset v1.3.3 (Krassowski 2022). After normalizing cargo orthogroup presence-absence values with the base R scale function, we produced a Euclidean distance matrix using the R dist function and performed hierarchical clustering with the hclust function using the “complete” agglomeration method. We then compared the topology produced by hierarchical clustering with phylogenetic relationships from the larger *k*-mer-based tree using a tanglegram and calculated the normalized RF distance, as described above. We also determined the location of cargo orthogroups—i.e. whether orthologous genes were only found inside elements or also found in the wider genome.

We searched for specific genes or domains previously reported to be prevalent in *Starships* or with assigned functional roles of particular note (Gluck-Thaler et al. 2022) using BLAST v2.10 (Camacho et al. 2009) and also PFAM domain assignment from the functional annotation (Hill et al. 2025). Namely: DUF3723, ferric reductase (FRE), patatin-like phosphatase (PLP), ToxA effector, spore killing (Spok) genes, and associated domains. We additionally made BLAST searches against the Pathogen–Host Interactions Database v4.17 (PHI-base; (Urban et al. 2025) downloaded on August 1, 2024, and considered a positive match when at least

50% of genes in an orthogroup had the same hit. We assessed whether Gene Ontology (GO) terms were enriched amongst cargo genes using the R package topGO v2.52.0 (Alexa and Rahnenfuhrer 2022) with Fisher's exact test and the weight01 algorithm.

In addition to previously mentioned packages, data analysis and visualization were performed using the following R packages: cowplot v1.1.3 (Wilke 2024), ggforce v0.4.2 (Pedersen 2024), gggenomes v1.0.0 (Hackl et al. 2024), ggnewscale v0.4.10 (Campitelli 2024), ggpubr v0.6.0 (Kassambara 2023), ggrepel v0.9.5 (Slowikowski 2024), matrixStats v1.3.0 (Bengtsson 2024), patchwork v1.2.0 (Pederson 2024), scales v1.3.0 (Wickham and Seidel 2023), tgutit v0.1.15 (Chomsky and Lifshitz 2023), and tidyverse v2.0.0 (Wickham et al. 2019).

Results and discussion

A *k*-mer-based approach for *Starship* phylogenetics recovers signal corresponding to host taxonomy

We used a *k*-mer-based approach for phylogenetic analysis of *Starships* to produce a phylogenetic tree of 53 entire *Starship* element sequences from Gluck-Thaler et al. (2022) and Hill et al. (2025), encompassing 17 host genera across 6 classes in the Pezizomycotina. We found elements to broadly cluster by genus, even when differing greatly in length (Fig. 1a). This contrasted with the captain gene tree (Supplementary Fig. 1) and element and captain trees were frequently discordant (RF distance 0.73 = 73% differing bipartitions; Fig. 1b), i.e. *Starships* that were more closely related according to their *k*-mer profiles could have very divergent captain genes. There were some exceptions to element/captain discordance; for instance, similar relationships in both captain and element trees were observed for the *Alternaria* clade (Fig. 1b). *Alternaria* captains were also closely related to some *Macrophomina* captains, in reflection of expected host species relationships in the *Dothideomycetes*; however, dothideomycete captains were not monophyletic as *Macrophomina* captains were also dispersed across other clades in the captain tree (Supplementary Fig. 1). Overall, 6/10 host genera with more than 1 genome represented were monophyletic in the element tree vs 2/10 in the captain tree. Also note the placement of *Mpha_Derelict*—a previously “unclassifiable” deactivated *Starship* missing the captain gene—alongside other elements from *Macrophomina* species (Fig. 1a). Two striking disruptions of this host clustering were caused by the elements Bdot_Voyager and Pvar_Chrysaor, the latter of which has been recently asserted to be horizontally transferred between various eurotiomycete species (Urquhart et al. 2024).

To determine if these observations of clustering by host taxonomy extended more broadly across the Pezizomycotina, we used the same *k*-mer-based phylogenetics method on a larger dataset of 597 elements systematically predicted using the tool starfish by Gluck-Thaler et al. (2024) alongside 20 *Gaeumannomyces* elements (Hill et al. 2025). This again recovered widescale clustering by host taxonomy, with the additional clear formation of clades broadly corresponding to host class level (Fig. 1c; Supplementary Fig. 2). We also performed a more conservative analysis to minimize the risk of including *k*-mers from the background genome, where we filtered the larger dataset to include only elements with predicted flanking DRs (343 elements + 14 *Gaeumannomyces* elements), which broadly reflected the results from the unfiltered dataset (Supplementary Fig. 3). As may be expected from the observed element/captain tree discordance in Fig. 1b, family classifications based on captains were scattered across the larger starfish-predicted element *k*-mer trees

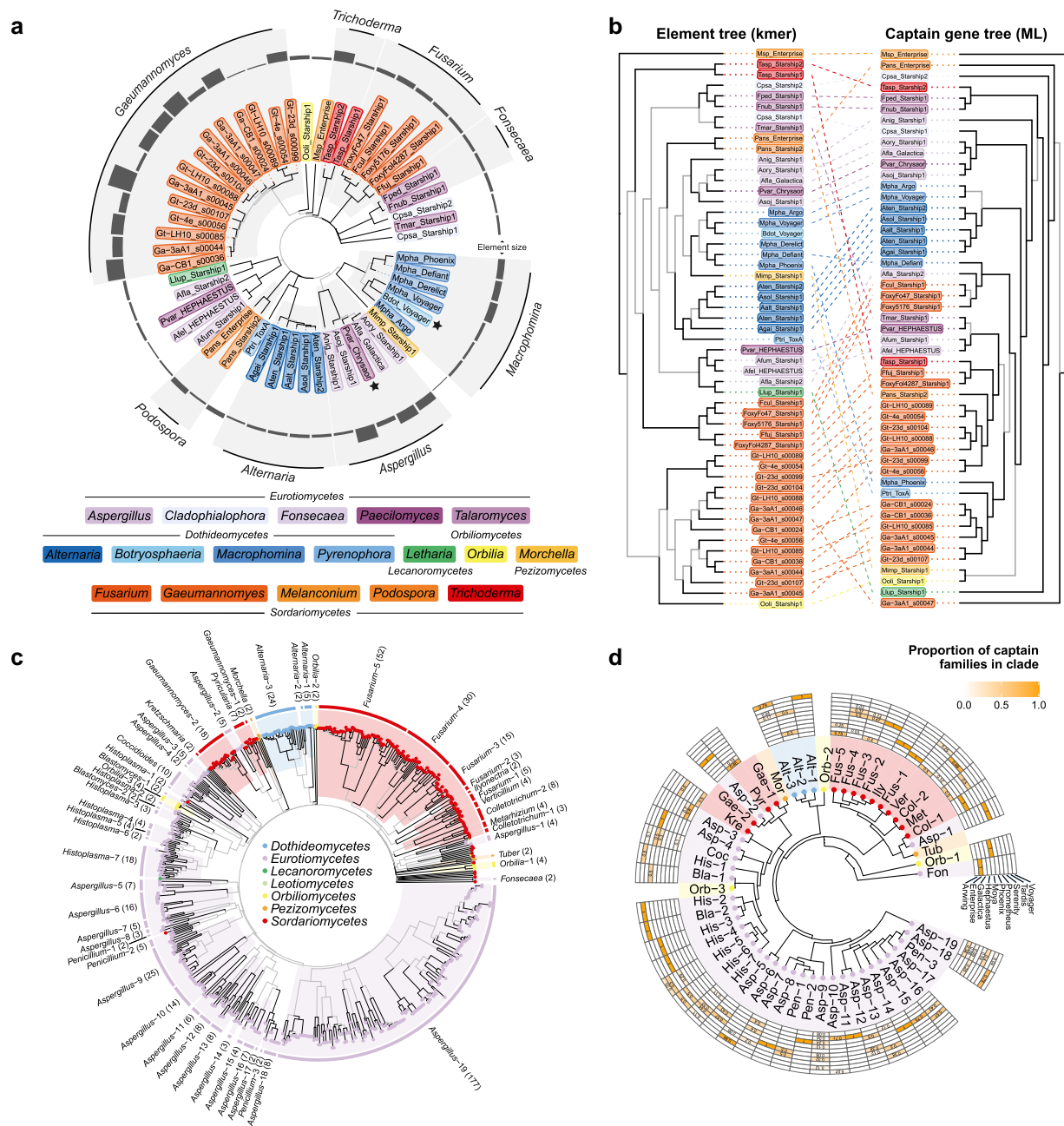


Fig. 1. K-mer-based phylogenetic analyses of Starship elements. a) An unrooted k-mer-based phylogenetic tree of 53 Starships—39 curated elements from 33 Pezizomycotina species (Gluck-Thaler et al. 2022; Gluck-Thaler and Vogan 2024) and 14 predicted by starfish from *Gaeumannomyces* species (Hill et al. 2025). Gray branches indicate bootstrap support < 70. Tip points are colored by genus and the outer ring indicates total element length. Black stars beside tips highlight elements from another genus in an otherwise monophyletic clade. b) A tanglegram comparing the topology of the k-mer-based element tree in a) and a maximum likelihood gene tree of the corresponding captain genes (see Supplementary Fig. 1 for the unrooted captain tree). Both trees are arbitrarily rooted with the *Msp_Enterprise* element. Gray branches indicate bootstrap support < 70. c) An unrooted k-mer-based phylogenetic tree of 617 Starships predicted with starfish (Gluck-Thaler and Vogan 2024; Hill et al. 2025), with gray branches indicating bootstrap support < 70. Genus-level monophyletic clades are highlighted and labeled, with the number of elements in each clade shown in brackets. Clades and tips are colored by host taxonomic class. See Supplementary Fig. 2 for element tip labels and captain-based family classifications. d) A summary of the k-mer-based tree in c) with genus-level monophyletic clades collapsed. The outer grid summarises Starship family classifications based on captain genes for the elements in each clade, with a darker grid cell colour indicating a higher proportion of the elements within the clade belonging to that family. Clades with no grid cells did not have any classified captain data.

(Fig. 1d, Supplementary Figs. 2 and 3). The degree of element/captain phylogenetic discordance is important because phylogenetic relationships of captains have been the predominant factor in element classification (Gluck-Thaler and Vogan 2024).

Phylogenetic discordance in comparison to species relationships is frequently used as evidence for horizontal gene transfer

(HGT) (Ravenhall et al. 2015); however, there are a number of alternative biological and/or analytical factors that can also result in a similar pattern (Steenwyk et al. 2023). Trans-species polymorphisms, where polymorphism originates before speciation and is preserved, potentially by balancing selection, can result in genes being more similar between species than within. Trans-species

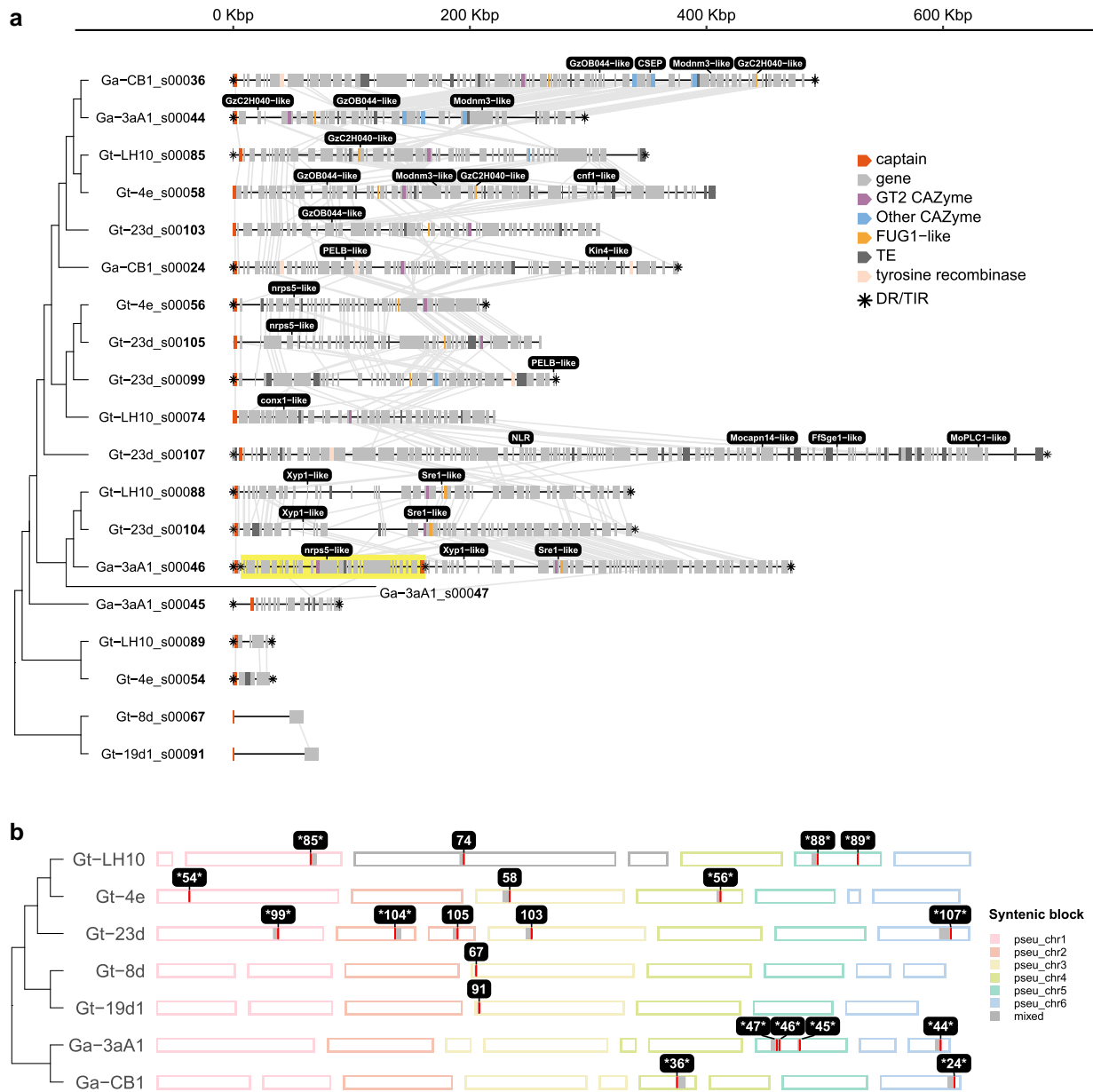


Fig. 2. Summary of the *Gaemannomyces* Starships predicted using starfish by Hill et al. (2025). a) A schematic of the 20 Starships ordered by phylogenetic relationships taken from Supplementary Fig. 2. Synteny between orthologous genes in neighboring elements is indicated by gray lines. A nested element (Ga-3aA1_s00047) is highlighted. Common genes are colored with known functions and the presence of flanking DRs or TIRs are indicated with an asterisk. Genes of note are labeled in black boxes. b) Ideograms showing the position of the 20 Starships across pseudochromosomes, adapted from Hill et al. (2025). ID numbers correspond to the bolded numbers for each element in a). Elements with flanking DRs are indicated with asterisks either side of the element ID number.

polymorphisms have been reported in fungal genes associated with vegetative incompatibility (Milgroom et al. 2018; Auxier et al. 2024), and such genes have been found multiple times in Starships (Fig. 2a; Gluck-Thaler et al. 2022, 2024; Urquhart et al. 2024). Even without natural selection, neutral processes such as incomplete lineage sorting, recombination and gene conversion, and gene duplication and loss can elevate levels of discordance (Bjornson et al. 2024). The latter is a particularly aggravating factor for misidentifying HGT as it can result in paralogues being mistaken as orthologues (Smith and Hahn 2021).

Another suite of commonly used methods to detect HGT are “surrogate” phylogenetics methods, which do not build a tree

but still assess evolutionary distances, e.g. using sequence similarity (Ravenhall et al. 2015); however, the results of surrogate methods can still be confounded by the phenomena described above. A sequence similarity approach also comes with the caveat that the best BLAST hit is not necessarily the closest related gene (Koski and Golding 2001) and requires subjective decisions about acceptable similarity thresholds. Distinguishing the cause(s) of phylogenetic discordance can be especially difficult for closely related taxa (Steenwyk et al. 2023), which is relevant here as elements from different host species were scattered amongst each other within genus-level clades in all k-mer-based tree analyses (Fig. 1a, Supplementary Fig. 2). Due to semipermeable species

boundaries in fungi, interspecific hybridization within the genus level has been detected multiple times (Steenkamp et al. 2018; Steensels et al. 2021). In such cases, *Starships* could be inherited during sexual reproduction between 2 different species and subsequent backcrossing could leave the element as an introgression, which may be mistaken as having been horizontally transferred. For all the reasons outlined above, general frequency of HGT events may have been overestimated in fungi (Kurland et al. 2003; Dupont and Cox 2017). The *k*-mer -based phylogenetics approach described here may be useful in certain contexts as 1 piece of evidence toward identifying (or dismissing) HGT, but the confounding factors described above would need to be assessed to have confidence that HGT has occurred (e.g. Fijarczyk and Babik 2015; Knowles et al. 2018). A number of the above factors contributing to discordant relationships are likely to have a greater impact for more closely related species, and it may be important to focus attention on apparent HGT events across greater evolutionary distances, which are presumed to be rarer, at least in prokaryotes (Popa et al. 2017; Burch et al. 2023; Dmitrijeva et al. 2024).

In the larger *k*-mer -based tree there were many within genus subclades of elements with captains of the same family, but also cases where minimally diverged sister elements had different captains. For example: *aspcr12_s00912* and *aspcr1_s00891* from different host genomes within the *Aspergillus*-9 clade had Phoenix and Prometheus captains, respectively; and *aspnig6_s01954* and *aspnig6_s01955* from the same host genome within the clade *Aspergillus*-19 had Hephaestus and Phoenix captains, respectively (Supplementary Fig. 2). It should be noted that there is some uncertainty as to the boundaries of these elements, as in these cases elements did not have predicted flanking repeats. A similar observation was made by Gluck-Thaler and Vogan (2024) for *Starship* pairs with near-identical cargo “haplotypes” but different captain-derived families. Together with the fact that captain genes are phylogenetically indistinguishable from “lone” tyrosine recombinase genes harboring the DUF3435 domain (Gluck-Thaler et al. 2022; Hill et al. 2025), this prompts the question as to whether *Starships* can swap the captain for a different tyrosine recombinase gene, which would render the “captain” status as somewhat transient. A previous study has already reported that *Starship* elements can lose their captain gene to become “degraded” or “derelict” (Gluck-Thaler et al. 2022), and in another study a mechanism has been suggested wherein different elements partake in cargo swapping (Urquhart et al. 2024). A similar mechanism where the captain, as opposed to the cargo, is swapped to acquire a captain gene from a different family could be a strategy to diversify insertions of virtually identical elements into different target sites. Comparing the *k*-mer profiles of regions surrounding CMEs could incidentally be another fruitful avenue for understanding target site preference, as many *Starships* have been found to insert into other TEs and AT-rich regions but without clear patterns in, for instance, TE superfamily or domain (Gluck-Thaler and Vogan 2024).

Aside from the major clade in the larger starfish *k*-mer tree overrepresented with elements from eurotiomycete hosts, other eurotiomycete elements appeared scattered amongst other clades, although there were lower support values for deeper tree nodes (Fig. 1c). It is notable that eurotiomycete elements dominate the starfish dataset—of all the genomes explored by Gluck-Thaler and Vogan (2024), *Eurotiomycetes* was the class with the highest proportion of genomes returning a *Starship* (36%; Supplementary Fig. 4). This was closely followed by the *Orbiliomycetes* (28%), despite 16 times fewer orbiliomycete genomes having been surveyed compared with the *Eurotiomycetes*, and orbiliomycete element clades were similarly widespread

across the *k*-mer tree (Fig. 1c). As one of the earliest diverging classes within the *Pezizomycotina* subphylum, the *Orbiliomycetes* are distantly related to *Eurotiomycetes* (Li et al. 2021), and they do not share ecological distributions more so than other taxonomic classes, so the underlying biological explanation is unclear. The far larger *Eurotiomycetes* class comprises diverse lifestyles including: rock-inhabiting fungi and other extremophiles; plant and animal pathogens; lichenized and lichen-associated fungi; ectomycorrhizal fungi; ant mutualists; and saprotrophs (Geiser et al. 2015). The *Orbiliomycetes* are primarily thought to be saprotrophs but include some soil-dwelling carnivorous fungi that trap invertebrates (Pfister 2015). Variation in the rate of *Starship* recovery in the genomes of different taxonomic classes could be a result of inconsistencies in assembly quality or bias within the starfish tool to recover certain elements from certain classes. However, these results do suggest that there may be a relationship between the tendency for a taxonomic class to have *Starship* elements and greater diversity of element clades.

While we consider this to be a promising application for *k*-mer -based phylogenetics, we must note that such methods were typically developed for whole-genome data. We are not aware of *k*-mer -based phylogenetic methods having been tested on sequences such as fungal CMEs. However, given that such methods are considered well-suited to viral genomes due to their high levels of mutation, gene duplication, and rearrangement (Zielezinski et al. 2017), CMEs would appear to be a similarly appropriate use case. Other than circumventing issues with alignment, *k*-mer -based methods also have the advantage of being more computationally efficient than alignment-based phylogenetic methods, which could reduce the carbon footprint of analyses (Grealey et al. 2022). There are many different approaches and tools for alignment-free sequence comparison which would warrant further testing in the context of CME phylogenetics (Luczak et al. 2019; Zielezinski et al. 2019). For instance, ANI is frequently used as a distance metric for prokaryote genomes and, as implemented in sourmash, has the benefit of a more realistic evolutionary model of mutation than that used by Mash (Rahman Hera et al. 2023), but whether it is appropriate for eukaryote data has yet to be validated. Nonetheless, we found that trees generated from ANI distance matrices produced using sourmash were broadly consistent with our Mashtree results (Supplementary Figs. 5 and 6) and supported our conclusion that *Starships* predominantly cluster according to host taxonomy. We were unable to produce a *k*-mer tree for captain genes using Mashtree, presumably due to the much smaller sequence length of a single gene. This meant we were not able to directly compare whole element and captain trees using the same *k*-mer -based method. However, at the genome-scale, previous comparisons of alignment and *k*-mer methods suggest reasonable topological congruence (VanWalleendael and Alvarez 2022; Lo et al. 2022; Van Etten et al. 2023), or no greater incongruence than might be expected from using different alignment-based methods (e.g. Shen et al. 2021). This also demonstrates the capacity for *k*-mer -based methods to reconstruct evolutionary history and, when they incorporate models of evolution, be deemed “phylogenetic”.

There are some limitations to alignment-free phylogenetics methods. Unlike conventional alignment-based phylogenetic trees, alignment-free trees do not produce branch lengths with a scale corresponding to geological time, and so one cannot extrapolate the date of divergences. Alignment-free methods also struggle with the reconstruction of deep nodes (Fan et al. 2015), which is evident from the *k*-mer trees we present here, although that issue is inherent to all phylogenetics methods (Lanier and Knowles 2015). This may limit the ability of these methods to

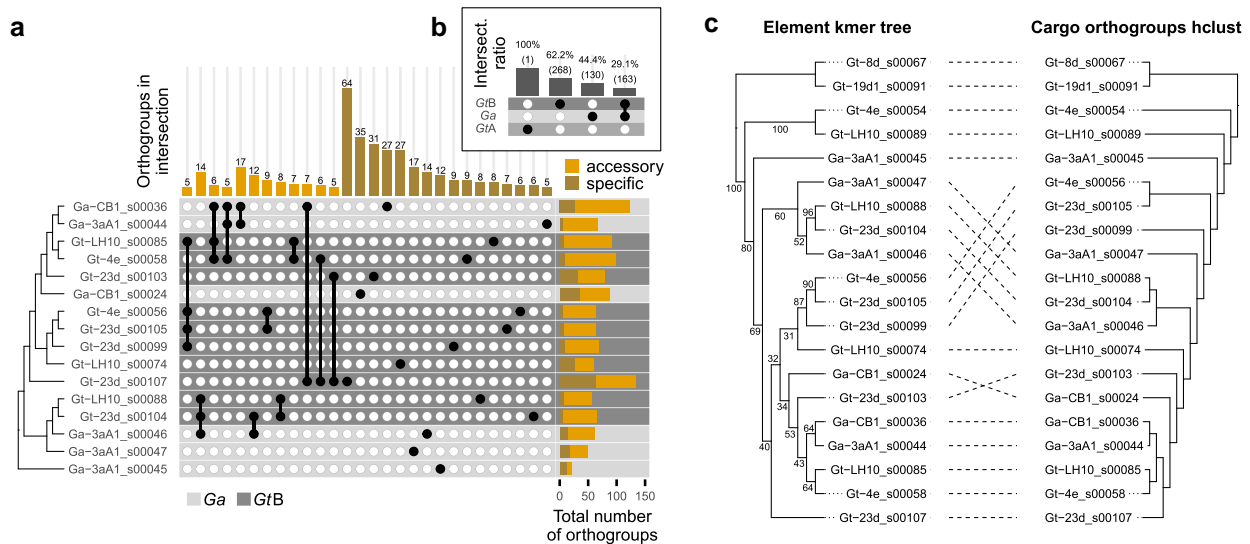


Fig. 3. Comparison of cargo gene content similarity across *Gaemannomyces* *Starships*. a) An upset plot indicating groups of elements which share at least 5 orthologous genes (accessory), and elements with at least 5 unique cargo genes (specific). Elements are ordered by phylogenetic relationships taken from [Supplementary Fig. 2](#). Total number of cargo orthogroups is shown in the right-hand bar plot with the proportion of accessory and specific cargo genes colored per element. Element rows are colored by host lineage. A representation of all shared accessory orthologous genes is given in [Supplementary Fig. 7](#). b) An upset plot indicating the ratio of orthologous genes shared across lineage/species boundaries. c) A tanglegram comparing the topology of *Gaemannomyces* elements taken from [Supplementary Fig. 2](#) and a hierarchical clustering of cargo orthologous gene presence-absence.

address questions about inter-relatedness of larger CME clades but should still allow for assessment of more recent divergences.

Both cargo genes and noncoding cargo content contribute to k-mer -based phylogenetic relationships between *Gaemannomyces* *Starships*

To explore the extent to which cargo gene content corresponded with the k-mer -based phylogenetic relationships, we used twenty *Starships* previously identified from 7 genomes across 3 separate lineages within the genus *Gaemannomyces*, an understudied member of the *Magnaporthaceae* ([Hill et al. 2025](#)). These genomes were sequenced from 5 strains of the wheat root pathogen species *G. tritici* (Gt) and 2 of the oat root pathogen *G. avenae* (Ga). Within the Gt strains there is further subdivision of 2 strains belonging to “type A” and 3 to “type B,” 2 distinct genetic lineages present in the species ([Palma-Guerrero et al. 2021](#)). This division is meaningful, as differences between the 2 types in terms of both virulence and genomic signatures may indicate that these 2 types actually represent cryptic species ([Hill et al. 2025](#)). As well as being a consistently amassed set of *Starships* for controlled comparison, these *Gaemannomyces* elements also provided major variability, ranging from ~32–688 Kbp in total length and containing between 1 and 156 genes ([Fig. 2a](#)). It should be noted that 6 of the elements, including both from the GtA strains, were excluded from the first phylogenetic analysis ([Fig. 1a](#)) as these elements did not have predicted flanking DRs and so there is some uncertainty as to their exact boundaries. However, we retained them here so as not to exclude potentially biologically meaningful results.

We found that *Starships* with greater numbers of shared orthologous genes were frequently sister elements or closely related in the k-mer tree, for instance, Gt-LH10_s00088, Gt-23d_s00104 and Ga-3aA1_s00046 ([Fig. 3a](#)). Most cases of more distantly related elements with high cargo gene sharedness involved the largest and most gene-rich element, Gt-23d_s00107, which incidentally also had one of the highest proportions (48%) of element-specific genes. Hierarchical clustering of cargo orthologous gene content

supported these results, with reasonable concordance between the hierarchical clustering and k-mer element tree (RF distance 0.47 = 47% differing bipartitions; [Fig. 3c](#)) and the most notable deviation between the 2 trees was the divergence of element Gt-23d_s00107. Pairs of closely related elements with evident regions of syntenic cargo genes ([Fig. 2a](#)) were often located on different chromosomes, suggesting previous mobilization (e.g. Gt-23d_s00104 and Ga-3aA1_s00046; Ga-CB1_s00036 and Ga-3aA1_s00044; Gt-4e_s00056 and Gt-23d_s00105; [Fig. 2b](#)). In contrast, there were also apparently static elements, being closely related and in the same orientation and position within different genomes (e.g. Gt-LH10_s00088 and Ga-3aA1_s00046; Gt-4e_s00058 and Gt-23d_s00103). The question of how similar elements must be to be considered “the same” is also pertinent, as there was one case of closely related elements at different locations within the same host genome, although 1 lacking predicted flanking repeats (Gt-23d_s00105 and Gt-23d_s00099). Elements becoming multiplicity in the genome may arise from mobilization of an ancestral element followed by sexual recombination between 2 hosts with the element in the original and more recent genomic location, respectively ([Urquhart et al. 2023](#)).

While cargo gene content was evidently a contributing factor to the patterns of *Gaemannomyces* element relatedness recovered from the k-mer -based phylogenies, the nature of a k-mer -based approach means that intergenic content within *Starships* must also be implicated. Indeed, repetitive DNA, introns, and presumably other noncoding regions can provide important phylogenetic signals ([Lo et al. 2022](#)). Here, the only 2 GtA elements found, 1 in each GtA genome, contained a single cargo gene despite being 61 and 73 kbp long. In the larger k-mer tree of starfish-predicted elements the GtB and Ga elements were closely related to sordariomycete elements from the pathogenic rice blast fungus *Piricularia oryzae* (syn. *Magnaporthe oryzae*) and a eurotiomycete clade, while the single-gene GtA elements were in a distinct clade more closely related to an element from the sordariomycete species *Sporothrix brasiliensis*, albeit without significant branch

support (Supplementary Fig. 2). *S. brasiliensis* is found in soils and vegetation, but is also an opportunistic mammalian pathogen, primarily of humans and cats, due to its temperature-dependent dimorphic lifestyle (Téllez et al. 2014). Despite being a similar length (57 kbp) to the GtA elements, the *S. brasiliensis* element contained 19 genes, none of which showed sequence similarity with the single gene found in the GtA elements. This suggests that it was primarily noncoding cargo content that informed *k*-mer-based relationships between the *S. brasiliensis* and GtA elements. The GtA elements were also previously found to have likely undergone repeat-induced point mutation (RIP) (Hill et al. 2025). RIP induces transition mutations in repetitive DNA, with a particular bias for C→T mutations targeting CpA dinucleotides, and so RIP-like signatures in genomic sequences manifest as biases in the relative frequencies of dinucleotides (Lewis et al. 2009; Hane et al. 2015). This raises the question of whether or to what extent signatures of RIP, such as a higher frequency of TpA dinucleotides, influence *k*-mer-based inference of element relationships, especially in cases with extensive intergenic cargo content.

The whole-element *k*-mer trees, captain tree, and the patterns of shared cargo genes indicated that there is no apparent species boundary for *Starship* content between GtB and Ga. We found no evidence of similarity with GtA elements, although there was only 1 gene-poor GtA element with which to compare. We see 2 possible scenarios: (1) elements were in the common ancestor of all 3 lineages and lost in GtA or (2) elements are readily exchanged between Ga and GtB strains, whether through HGT or interspecific hybridization. Either way, together with the fact that, unlike the other *Gaeumannomyces* elements, the GtA elements were previously found to be subject to element-wide RIP (Hill et al. 2025), *Starship* prevalence and divergence may be another symptom of cryptic speciation between Gt types. Although GtB and Ga elements appear to be closely related, there was an imbalance in how cargo genes were shared, as a higher proportion of Ga cargo genes had an orthologue in GtB elements (56%) than GtB cargo genes had in Ga elements (38%; Fig. 3b). Additionally, there were differences in how cargo genes were distributed in the genome, with more cargo gene orthogroups only found inside Ga elements that had copies integrated into the wider genome in Gt strains than the reverse (Supplementary Fig. 8a). In a similar vein, Ga *Starships* broadly had a higher proportion of orthogroups that were only inside the element compared with GtB *Starships* (Supplementary Fig. 8b). Unpicking the differences in relative levels of duplication, sharedness, and location of cargo genes on different *Starships* may be important for determining patterns of inheritance or selection.

***Gaeumannomyces* Starship cargos harbor a variety of putative plant-fungal interaction genes, but the ToxA gene was notably absent**

Most genes previously reported to be common, or notable, in *Starships* (Gluck-Thaler et al. 2022) were absent from *Gaeumannomyces* *Starship* cargos, namely DUF3723, FRE, PLP, and spore-killer (Spok1) genes. There was 1 putative NOD-like receptor (NLR) located on element Gt-23d_s00107 (Fig. 2). The NLR contained a central NACHT domain—the most common nucleotide binding and oligomerization (NOD) domain in fungal NLRs (Daskalov et al. 2020)—a WD40 repeat domain, and a sesA N-terminal domain of unknown function (PF17107) that is more common in ascomycete NLRs (Daskalov et al. 2020). This sesA-NACHT-WD structure is also found in the NWD3 gene of the model experimental fungus *Podospira anserina* (Daskalov et al. 2012). While the function of sesA is not established, other members of the *P. anserina* NWD gene family are involved

in heterokaryon/vegetative incompatibility or self/nonself-recognition, which has also been hypothesized to contribute to an innate fungal immune system (Paoletti and Saupe 2009; Uehling et al. 2017).

Of particular note was the absence of the necrosis-inducing ToxA effector in the *Gaeumannomyces* cargos, which is located in *Starships* in 3 other wheat pathogens—*Pyrenophora tritici-repentis*, *Parastagonospora nodorum*, and *Bipolaris sorokiniana* (McDonald et al. 2019; Bucknell et al. 2024). *Py. tritici-repentis* and *Pa. nodorum* are known to frequently co-infect wheat (Abdullah et al. 2020), and *Py. tritici-repentis* and *B. sorokiniana* together form a leaf blight disease complex (Kumar et al. 2002). While we could not find information on the potential co-occurrence of *Gaeumannomyces* spp. and other wheat pathogens in the literature, based on their global distributions and the global distribution of the wheat crop, it is highly likely that *Gaeumannomyces* spp. also co-occur with 1 or more of these wheat pathogens (Větrovský et al. 2020), which would have provided the opportunity to exchange *Starships*. However, all 3 species containing ToxA reside in a different class, *Dothideomycetes*, in the order *Pleosporales*. At the present time, the lack of ToxA in the *Gaeumannomyces* *Starships* is consistent with our *k*-mer tree results indicating a host relatedness boundary to *Starship* exchange.

Regarding whether the *Gaeumannomyces* *Starship* cargos exhibited a core functional role, GO term enrichment analysis of cargo genes reflected high variability as there was no significant enrichment in most elements, although ubiquinone biosynthesis and regulation of translational fidelity were significantly enriched in Ga-3aA1_s00044 and Ga-CB1_s00036, respectively. There were no cargo orthogroups that were core to all elements, but 5 orthogroups were present in at least 50% of the elements (Supplementary Fig. 7). One was predicted to be a carbohydrate-active enzyme (CAZyme) belonging to glycosyltransferase family 2 (GT2; Fig. 2). The GT2 family includes enzymes necessary for the synthesis of chitin (Lairson et al. 2008), which is required for the structural integrity of the fungal cell wall (Bowman and Free 2006). A GT2 enzyme has been demonstrated to be required for the disease-causing abilities of the wheat pathogens *Zymoseptoria tritici* and *Fusarium graminearum* (King et al. 2017). Expansion and contraction of GT2 CAZyme genes have been shown to be strong predictors of phytopathogenicity and saprotrophy, respectively (Dort et al. 2023), but GT2 genes are also expanded in mycorrhizal lineages (Rosling et al. 2024), suggesting a key role in both pathogenic and mutualistic plant-fungal interactions. In addition to the prevalent GT2 orthogroup, other CAZymes and CAZyme families were found in various elements: sterol 3 β -glucosyltransferase (GT1), glycoside hydrolase (GH) family 33, α -galactosidase (CBM35 + GH27), and glucose-methanol-choline oxidoreductase (AA3_2) in elements Ga-3aA1_s00044 and Ga-CB1_s00036; chitinase (GH18) in Gt-LH10_s00085; and another GT2 CAZyme in Gt-23d_s00099.

Multiple *Gaeumannomyces* *Starship* cargo genes had BLAST hits to genes in the PHI-base database, which compiles and curates experimentally verified genes implicated in pathogen-host interactions (Urban et al. 2025). This included 4 genes in the closely related *P. oryzae* which have been associated with virulence in barley and rice, 2 of which are implicated in calcium signaling and 2 transcription factors, and the previously mentioned GT2 CAZyme which has been associated with virulence of *Zymoseptoria tritici* and *Fusarium graminearum* in wheat leaves and floral spikes, respectively (Table 1). Intriguingly, the chitinase CAZyme cargo gene in element Gt-LH10_s00085 matched a chitinase gene in the mycoparasite *Trichoderma virens* which is associated with its virulence toward the basidiomycete plant pathogen *Rhizoctonia solani*. *Trichoderma* species are known for endophytic colonization of

Table 1. PHI-base genes with BLAST hits in *Gaeumannomyces* Starship cargos.

PHI-base ID	Gene	Function	Species	Mutant phenotype	Plant-host
PHI:7559	FgGT2	glycosyltransferase	<i>Fusarium graminearum</i>	Loss of pathogenicity	<i>Triticum aestivum</i>
PHI:2057	MoPLC1	modulator of calcium flux	<i>Pyricularia oryzae</i> ^b	Loss of pathogenicity	<i>Oryza sativa</i>
PHI:3837	Sre1	iron-sensitive transcription factor	<i>Bipolaris maydis</i>	Reduced virulence	<i>Zea mays</i>
PHI:2476	CcpelA ^a	pectate lyase	<i>Colletotrichum coccodes</i>	Reduced virulence	<i>Solanum lycopersicum</i>
PHI:222	PELB ^a	pectate lyase	<i>Colletotrichum gloeosporioides</i>	Reduced virulence	<i>Persea americana</i>
PHI:9042	nrps5 (FGSG_13878)	non-ribosomal peptide synthetase	<i>Fusarium graminearum</i>	Reduced virulence	<i>Triticum aestivum</i>
PHI:6262	FUG1	role in pathogenicity and fumonisin biosynthesis	<i>Fusarium verticillioides</i>	Reduced virulence	<i>Zea mays</i>
PHI:3315	cnx1	Zn ² Cys ⁶ transcription factors	<i>Pyricularia oryzae</i> ^b	Reduced virulence	<i>Oryza sativa</i>
PHI:3308	cnf1	Zn ² Cys ⁶ transcription factors	<i>Pyricularia oryzae</i> ^b	Reduced virulence	<i>Hordeum vulgare</i>
PHI:2113	Kin4	Ca ²⁺ /CAM-dependent serine/threonine protein kinases	<i>Pyricularia oryzae</i> ^b	Reduced virulence	<i>Hordeum vulgare</i>
PHI:144	CHT42	chitinase	<i>Trichoderma virens</i>	Reduced virulence	<i>Rhizoctonia solani</i>
PHI:3210	FfSge1	morphological switch regulator	<i>Fusarium fujikuroi</i>	Unaffected pathogenicity	<i>Oryza sativa</i>
PHI:1603	GzOB044	transcription factor	<i>Fusarium graminearum</i>	Unaffected pathogenicity	<i>Triticum</i>
PHI:1377	GzC2H040	transcription factor	<i>Fusarium graminearum</i>	Unaffected pathogenicity	<i>Triticum</i>
PHI:6639	Modnm3	dynamain	<i>Pyricularia oryzae</i> ^b	Unaffected pathogenicity	<i>Oryza sativa</i>
PHI:6613	Mocapn14	calpain	<i>Pyricularia oryzae</i> ^b	Unaffected pathogenicity	<i>Oryza sativa</i>
PHI:124206	Xyp1 (Uv8b_02447)	cell wall degrading enzymes	<i>Ustilaginoides virens</i>	Increased virulence	<i>Oryza sativa</i>

^a Pectate lyases CcpelA and PELB matched to the same orthogroup.

^b *Pyricularia oryzae* = *Magnaporthe oryzae*.

plants, particularly roots, and in some cases can reduce disease via both inducing plant resistance and direct antagonism of other fungi (Harman et al. 2004). Two further orthogroups had BLAST hits to CAZyme genes in PHI-base (Xyp1 and PELB/CcpelA); however, as these were not previously flagged during CAZyme annotation (Hill et al. 2025) there remains some uncertainty as to their function.

Also of note is that none of the biosynthetic gene clusters (BGCs) previously identified in the *Gaeumannomyces* genomes were present in any Starships, but 2 cargo genes had hits to PHI-base genes implicated in secondary metabolite synthesis in *Fusarium* species, namely nrps5 and FUG1. The latter is involved in fumonisin (FUM) synthesis in *Fusarium verticillioides* (Ridenour and Bluhm 2017), but is located on a separate locus to the FUM gene cluster, suggesting that it may play a regulatory role, as biosynthesis transcription factors can frequently be located outside of contiguous BGCs (Kwon et al. 2021). FUG1 was also previously found to have orthologs across Ascomycota, including in Gt (Ridenour and Bluhm 2017). The non-ribosomal peptide synthetase nrps5 gene is located alongside nrps9 in an 8-member BGC cluster in *Fusarium* species, which produces fusaoctaxin A and is essential to virulence of *F. graminearum* in wheat (Jia et al. 2019). However, none of the genes surrounding the nrps5-like gene in the *Gaeumannomyces* elements showed similarity to the other nrps5/9 cluster members. We also found an uncharacterized candidate secreted effector protein (CSEP) gene in 1 element (Ga-CB1_s00036). Intriguingly, this CSEP was located within a region that was highly syntenic with another element (Ga-3aA1_s00044) but the CSEP was not present in that second element (Fig. 2), underlining the dynamism of Starship cargos.

Conclusions

Here, we provide evidence of a difference in evolutionary history between Starship elements in their entirety vs their captain genes. This raises the question: is it more important to define Starships by their mode of mobilization—i.e. the tyrosine recombinase captain gene—or the cargo of genes and noncoding/repetitive content mobilized? The answer to that question will depend on the context in which the question is asked, namely, whether the inquiry at hand is

to understand the mechanism of transposition, or to understand how elements and their cargos evolve and impact host fitness. Whole-element relationships are easily assessed using *k*-mer-based phylogenetic methods, which have revealed previously hidden signals corresponding to host taxonomy. These methods also allow us to assess relationships including “degraded” elements where captains and/or DRs/TIRs have been lost. By accounting for the composition of Starships without being hampered by alignment issues caused by repeats, indels, duplications, rearrangements and inversions, or lack of available sequences in general, *k*-mer-based phylogenetic methods can help to refine the existing haplotype-based classification of CMEs. Beyond informing classification, this new approach could also provide context and new insights to address fundamental outstanding questions regarding Starships and other CMEs, such as the evolutionary origins of elements, the prevalence of HGT, and the role of elements in the host genome.

Data availability

All original data sources used in this study are cited in the text. Analysis scripts are available at <https://github.com/Rowena-h/StarshipTrees>.

Supplemental material available at G3 online.

Acknowledgments

We are very grateful to Gillian Reynolds at the Earlham Institute for sharing her insights on *k*-mer-based methods. Many thanks to Javier Palma-Guerrero and Tania Chancellor for valuable discussion as part of the bilateral Earlham Institute–Rothamsted Research take-all working group. We also thank Neil Hall at the Earlham Institute for his continued support and guidance.

Funding

RH, GC, MG, KHK, and MM were supported by the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme (ISP) grant Delivering Sustainable Wheat (BB/X011003/1) within the work package Delivering Resilience to

Biotic Stress (BBS/E/ER/230003B Earlham Institute and BBS/E/RH/230001B Rothamsted Research). DS was supported by the BBSRC Core Capability Grant (BB/CCG2280/1). GC was supported by the DEFRA funded Wheat Genetic Improvement Network (WGIN) phase 3 (CH0106) and phase 4 (CH0109). MG was supported by the BBSRC ISP grant Decoding Biodiversity (BBX011089/1) within the work package Genome Enabled Analysis of Diversity to Identify Gene Function, Biosynthetic Pathways, and Variation in Agricultural Traits (BBS/E/ER/230002B).

Conflicts of interest

The author(s) declare no conflict of interest.

References

- Abdullah AS, Gibberd MR, Hamblin J. 2020. Co-infection of wheat by *Pyrenophora tritici-repentis* and *Parastagonospora nodorum* in the wheatbelt of western Australia. *Crop Pasture Sci.* 71(2):119–127. doi:10.1071/CP19412.
- Alexa A, Rahnenfuhrer J. 2022. topGO: Enrichment Analysis for Gene Ontology. [accessed 2023 Aug 15]. <https://bioconductor.org/packages/release/bioc/html/topGO.html>.
- Arkhipova IR, Yushenova IA. 2019. Giant transposons in eukaryotes: is bigger better? *Genome Biol Evol.* 11(3):906–918. doi:10.1093/gbe/evz041.
- Auxier B, Zhang J, Marquez FR, Senden K, Van Den Heuvel J, Aanen DK, Snelders E, Debets AJM. 2024. The narrow footprint of ancient balancing selection revealed by heterokaryon incompatibility genes in *Aspergillus fumigatus*. *Mol Biol Evol.* 41(5):msae079. doi:10.1093/molbev/msae079.
- Bengtsson H. 2024. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). [accessed 2024 Apr 11]. <https://cran.r-project.org/package=matrixStats>.
- Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics.* 186(4):1085–1093. doi:10.1534/genetics.110.124180.
- Bjornson S, Verbruggen H, Upham NS, Steenwyk JL. 2024. Reticulate evolution: detection and utility in the phylogenomics era. *Mol Phylogenet Evol.* 201:108197. doi:10.1016/j.ympev.2024.108197.
- Bowman SM, Free SJ. 2006. The structure and synthesis of the fungal cell wall. *BioEssays.* 28(8):799–808. doi:10.1002/bies.20441.
- Bucknell A, Wilson HM, Do Santos KCG, Simpfendorfer S, Milgate A, Germain H, Solomon PS, Bentham A, McDonald MC. 2024. *Sanctuary: A Starship* position facilitating the movement of the virulence factor ToxA in fungal wheat pathogens. *bioRxiv* 583430. <https://doi.org/10.1101/2024.03.04.583430>, preprint: not peer reviewed.
- Burch CL, Romanchuk A, Kelly M, Wu Y, Jones CD. 2023. Empirical evidence that complexity limits horizontal gene transfer. *Genome Biol Evol.* 15(6):evad089. doi:10.1093/gbe/evad089.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 10(1):421. doi:10.1186/1471-2105-10-421.
- Campitelli E. 2024. ggnewscale: Multiple Fill and Colour Scales in “ggplot2.” [accessed 2024 Feb 08]. <https://cran.r-project.org/package=ggnewscale>.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.
- Chancellor T, Smith DP, Chen W, Clark SJ, Venter E, Halsey K, Carrera E, McMillan V, Canning G, Armer VJ, et al. 2024. A fungal endophyte induces local cell wall-mediated resistance in wheat roots against take-all disease. *Front Plant Sci.* 15:1444271. doi:10.3389/fpls.2024.1444271.
- Chomsky E, Lifshitz A. 2023. tgutil: Simple utility functions for Tanay lab code. [accessed 2023 Oct 13]. <https://github.com/tanaylab/tgutil>.
- Daskalov A, Dyrka W, Saupe SJ. 2020. 6 NLR function in fungi as revealed by the study of self/non-self recognition systems. In: Benz JP, Schipper K, editors. *Genetics and Biotechnology. The Mycota.* Cham: Springer International Publishing. p. 123–141.
- Daskalov A, Paoletti M, Ness F, Saupe SJ. 2012. Genomic clustering and homology between HET-S and the NWD2 STAND protein in various fungal genomes. *PLoS One.* 7(4):e34854. doi:10.1371/journal.pone.0034854.
- Dmitrijeva M, Tackmann J, Matias Rodrigues JF, Huerta-Cepas J, Coelho LP, von Mering C. 2024. A global survey of prokaryotic genomes reveals the eco-evolutionary pressures driving horizontal gene transfer. *Nat Ecol Evol.* 8(5):986–998. doi:10.1038/s41559-024-02357-0.
- Dort EN, Layne E, Feau N, Butyaev A, Henrissat B, Martin FM, Haridas S, Salamov A, Grigoriev IV, Blanchette M, et al. 2023. Large-scale genomic analyses with machine learning uncover predictive patterns associated with fungal phytopathogenic lifestyles and traits. *Sci Rep.* 13(1):17203. doi:10.1038/s41598-023-44005-w.
- Dupont P-Y, Cox MP. 2017. Genomic data quality impacts automated detection of lateral gene transfer in fungi. *G3 (Bethesda).* 7(4):1301–1314. doi:10.1534/g3.116.038448.
- Fan H, Ives AR, Surget-Groba Y, Cannon CH. 2015. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. *BMC Genomics.* 16(1):522. doi:10.1186/s12864-015-1647-5.
- Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol.* 24(14):3529–3545. doi:10.1111/mec.13226.
- Geiser DM, LoBuglio KF, Gueidan C. 2015. 5 Pezizomycotina: eurotiomycetes. In: McLaughlin DJ, Spatafora JW, editors. *Systematics and Evolution. The Mycota, Vol 7B.* Berlin: Springer. p. 121–141.
- Gluck-Thaler E, Forsythe A, Puerner C, Stajich JE, Croll D, Cramer RA, Vogan AA. 2024. Giant transposons promote strain heterogeneity in a major fungal pathogen. *bioRxiv* 601215. <https://doi.org/10.1101/2024.06.28.601215>, preprint: not peer reviewed.
- Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack TL, Wood CW, Slot JC, Lopez-Nicora HD, et al. 2022. Giant *Starship* elements mobilize accessory genes in fungal genomes. *Mol Biol Evol.* 39(5):msac109. doi:10.1093/molbev/msac109.
- Gluck-Thaler E, Vogan AA. 2024. Systematic identification of cargo-mobilizing genetic elements reveals new dimensions of eukaryotic diversity. *Nucleic Acids Res.* 52(10):5496–5513. doi:10.1093/nar/gkae327.
- Goodwin TJD, Butler MI, Poulter RTM. 2003. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology.* 149(11):3099–3109. doi:10.1099/mic.0.26529-0.
- Grealey J, Lannelongue L, Saw W-Y, Marten J, Méric G, Ruiz-Carmona S, Inouye M. 2022. The carbon footprint of bioinformatics. *Mol Biol Evol.* 39(3):msac034. doi:10.1093/molbev/msac034.
- Hackl T, Ankenbrand MJ, van Adrichem B. 2024. gggenomes: A Grammar of Graphics for Comparative Genomics. [accessed 2024 Jun 19]. <https://github.com/thackl/gggenomes>.
- Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP. 2015. Repeat-induced point mutation: a fungal-specific, endogenous mutagenesis process. In: van den Berg MA, Maruthachalam K, editors. *Fungal Biology: Genetic Transformation Systems in Fungi.* Springer International Publishing. p. 55–68.

- Harman GE, Howell CR, Viterbo A, Chet I, Lorito M. 2004. *Trichoderma* species—opportunistic, avirulent plant symbionts. *Nat Rev Microbiol.* 2(1):43–56. doi:[10.1038/nrmicro797](https://doi.org/10.1038/nrmicro797).
- Hill R, Grey M, Fedi MO, Smith D, Canning G, Ward SJ, Irish N, Smith J, McMillan VE, Hammond J, et al. 2025. Evolutionary genomics reveals variation in structure and genetic content implicated in virulence and lifestyle in the genus *Gaeumannomyces*. *BMC Genomics.* 26(1):239. doi:[10.1186/s12864-025-11432-0](https://doi.org/10.1186/s12864-025-11432-0).
- Irber L, Pierce-Ward NT, Abuelanin M, Alexander H, Anant A, Barve K, Baumler C, Botvinnik O, Brooks P, Dsouza D, et al. 2024. Sourmash v4: a multitool to quickly search, compare, and analyze genomic and metagenomic data sets. *JOSS.* 9(98):6830. doi:[10.21105/joss.06830](https://doi.org/10.21105/joss.06830).
- Jia L-J, Tang H-Y, Wang W-Q, Yuan T-L, Wei W-Q, Pang B, Gong X-M, Wang S-F, Li Y-J, Zhang D, et al. 2019. A linear nonribosomal octapeptide from *Fusarium graminearum* facilitates cell-to-cell invasion of wheat. *Nat Commun.* 10(1):922. doi:[10.1038/s41467-019-08726-9](https://doi.org/10.1038/s41467-019-08726-9).
- Kassambara A. 2023. ggpubr: “ggplot2” Based Publication Ready Plots. [accessed 2023 Oct 10]. <https://cran.r-project.org/package=ggpubr>.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version : improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780. doi:[10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Katz LS, Griswold T, Morrison SS, Caravas JA, Zhang S, den Bakker HC, Deng X, Carleton HA. 2019. Mashtree: a rapid comparison of whole genome sequence files. *JOSS.* 4(44):1762. doi:[10.21105/joss.01762](https://doi.org/10.21105/joss.01762).
- King R, Urban M, Lauder RP, Hawkins N, Evans M, Plummer A, Halsey K, Lovegrove A, Hammond-Kosack K, Rudd JJ. 2017. A conserved fungal glycosyltransferase facilitates pathogenesis of plants by enabling hyphal growth on solid surfaces. *PLoS Pathog.* 13(10):e1006672. doi:[10.1371/journal.ppat.1006672](https://doi.org/10.1371/journal.ppat.1006672).
- Knowles LL, Huang H, Sukumaran J, Smith SA. 2018. A matter of phylogenetic scale: distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. *Am J Bot.* 105(3):376–384. doi:[10.1002/ajb2.1064](https://doi.org/10.1002/ajb2.1064).
- Kojima KK, Jurka J. 2011. Crypton transposons: identification of new diverse families and ancient domestication events. *Mob DNA.* 2(1):12. doi:[10.1186/1759-8753-2-12](https://doi.org/10.1186/1759-8753-2-12).
- Koski LB, Golding GB. 2001. The closest BLAST hit is often not the nearest neighbor. *J Mol Evol.* 52(6):540–542. doi:[10.1007/s002390010184](https://doi.org/10.1007/s002390010184).
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 35(21):4453–4455. doi:[10.1093/bioinformatics/btz305](https://doi.org/10.1093/bioinformatics/btz305).
- Krassowski M. 2022. ComplexUpset. [accessed 2024 Jun 19]. <https://github.com/krassowski/complex-upset/tree/v1.3.5>.
- Kumar J, Schäfer P, Hückelhoven R, Langen G, Baltruschat H, Stein E, Nagarajan S, Kogel K. 2002. *Bipolaris sorokiniana*, a cereal pathogen of global concern: cytological and molecular approaches towards better control. *Mol Plant Pathol.* 3(4):185–195. doi:[10.1046/j.1364-3703.2002.00120.x](https://doi.org/10.1046/j.1364-3703.2002.00120.x).
- Kurland CG, Canback B, Berg OG. 2003. Horizontal gene transfer: a critical view. *Proc Natl Acad Sci U S A.* 100(17):9658–9662. doi:[10.1073/pnas.1632870100](https://doi.org/10.1073/pnas.1632870100).
- Kwon MJ, Steiniger C, Cairns TC, Wisecaver JH, Lind AL, Pohl C, Regner C, Rokas A, Meyer V. 2021. Beyond the biosynthetic gene cluster paradigm: genome-wide coexpression networks connect clustered and unclustered transcription factors to secondary metabolic pathways. *Microbiol Spectr.* 9(2):e0089821. doi:[10.1128/Spectrum.00898-21](https://doi.org/10.1128/Spectrum.00898-21).
- Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 77(1):521–555. doi:[10.1146/annurev.biochem.76.061005.092322](https://doi.org/10.1146/annurev.biochem.76.061005.092322).
- Lanier HC, Knowles LL. 2015. Applying species-tree analyses to deep phylogenetic histories: challenges and potential suggested from a survey of empirical phylogenetic studies. *Mol Phylogenet Evol.* 83:191–199. doi:[10.1016/j.ympev.2014.10.022](https://doi.org/10.1016/j.ympev.2014.10.022).
- Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schübeler D, Selker EU. 2009. Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. *Genome Res.* 19(3):427–437. doi:[10.1101/gr.086231.108](https://doi.org/10.1101/gr.086231.108).
- Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, Spatafora JW, Groenewald M, Dunn CW, Hittinger CT, et al. 2021. A genome-scale phylogeny of the kingdom fungi. *Curr Biol.* 31(8):1653–1665.e5. doi:[10.1016/j.cub.2021.01.074](https://doi.org/10.1016/j.cub.2021.01.074).
- Lo R, Dougan KE, Chen Y, Shah S, Bhattacharya D, Chan CX. 2022. Alignment-free analysis of whole-genome sequences from symbiodiniaceae reveals different phylogenetic signals in distinct regions. *Front Plant Sci.* 13:815714. doi:[10.3389/fpls.2022.815714](https://doi.org/10.3389/fpls.2022.815714).
- Luczak BB, James BT, Girgis HZ. 2019. A survey and evaluations of histogram-based statistics in alignment-free sequence comparison. *Brief Bioinform.* 20(4):1222–1237. doi:[10.1093/bib/bbx161](https://doi.org/10.1093/bib/bbx161).
- McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, Milgate A, Solomon PS. 2019. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *mBio.* 10(5):e01515–e01519. doi:[10.1128/mBio.01515-19](https://doi.org/10.1128/mBio.01515-19).
- Milgroom MG, Smith ML, Drott MT, Nuss DL. 2018. Balancing selection at nonself recognition loci in the chestnut blight fungus, *Cryphonectria parasitica*, demonstrated by trans-species polymorphisms, positive selection, and even allele frequencies. *Heredity.* 121(6):511–523. doi:[10.1038/s41437-018-0060-7](https://doi.org/10.1038/s41437-018-0060-7).
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17(1):132. doi:[10.1186/s13059-016-0997-x](https://doi.org/10.1186/s13059-016-0997-x).
- Palma-Guerrero J, Chancellor T, Spong J, Canning G, Hammond J, McMillan VE, Hammond-Kosack KE. 2021. Take-all disease: new insights into an important wheat root pathogen. *Trends Plant Sci.* 26(8):836–848. doi:[10.1016/j.tplants.2021.02.009](https://doi.org/10.1016/j.tplants.2021.02.009).
- Paoletti M, Saupe SJ. 2009. Fungal incompatibility: evolutionary origin in pathogen defense? *BioEssays.* 31(11):1201–1210. doi:[10.1002/bies.200900085](https://doi.org/10.1002/bies.200900085).
- Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 35(3):526–528. doi:[10.1093/bioinformatics/bty633](https://doi.org/10.1093/bioinformatics/bty633).
- Pedersen TL. 2024. ggforce: Accelerating “ggplot2.” [accessed 2024 Feb 19]. <https://cran.r-project.org/web/packages/ggforce/index.html>.
- Pederson TL. 2024. patchwork: The Composer of Plots. [accessed 2024 Jan 08]. <https://CRAN.R-project.org/package=patchwork>.
- Pfister DH. 2015. 2 Pezizomycotina: pezizomycetes, orbiliomycetes. In: McLaughlin DJ, Spatafora JW, editors. *Systematics and Evolution. The Mycota, Vol 7B*. Berlin: Springer. p. 35–55. doi:[10.1007/978-3-662-46011-5_2](https://doi.org/10.1007/978-3-662-46011-5_2).
- Popa O, Landan G, Dagan T. 2017. Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. *ISME J.* 11(2):543–554. doi:[10.1038/ismej.2016.116](https://doi.org/10.1038/ismej.2016.116).
- Rahman Hera M, Pierce-Ward NT, Koslicki D. 2023. Deriving confidence intervals for mutation rates across a wide range of evolutionary distances using FracMinHash. *Genome Res.* 33(7):1061–1068. doi:[10.1101/gr.277651.123](https://doi.org/10.1101/gr.277651.123).

- Ravenhall M, Škunca N, Lassalle F, Dessimoz C. 2015. Inferring horizontal gene transfer. *PLoS Comput Biol*. 11(5):e1004095. doi:10.1371/journal.pcbi.1004095.
- R Core Team. 2023. R: A language and environment for statistical computing. [accessed 2023 Jun 16]. <https://www.r-project.org/>.
- Revell LJ. 2024. Phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*. 12:e16505. doi:10.7717/peerj.16505.
- Ridenour JB, Bluhm BH. 2017. The novel fungal-specific gene FUG1 has a role in pathogenicity and fumonisin biosynthesis in *Fusarium verticillioides*. *Mol Plant Pathol*. 18(4):513–528. doi:10.1111/mpp.12414.
- Rosling A, Eshghi Sahraei S, Kalsoom Khan F, Desirò A, Bryson AE, Mondo SJ, Grigoriev IV, Bonito G, Sánchez-García M. 2024. Evolutionary history of arbuscular mycorrhizal fungi and genomic signatures of obligate symbiosis. *BMC Genomics*. 25(1):529. doi:10.1186/s12864-024-10391-2.
- Round ER. 2021. [glottoTrees]: Phylogenetic trees in Linguistics. [accessed 2024 Jul 01]. <https://github.com/erichround/glottoTrees>.
- Schliep K, Potts AJ, Morrison DA, Grimm GW. 2017. Intertwining phylogenetic trees and networks. *Methods Ecol Evol*. 8(10):1212–1220. doi:10.1111/2041-210X.12760.
- Shen X-X, Steenwyk JL, Rokas A. 2021. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Syst Biol*. 70(5):997–1014. doi:10.1093/sysbio/syab011.
- Slowikowski K. 2024. ggrepel: Automatically Position Non-Overlapping Text Labels with “ggplot2.” [accessed 2024 Jan 10]. <https://cran.r-project.org/package=ggrepel>.
- Smith ML, Hahn MW. 2021. New approaches for inferring phylogenies in the presence of paralogs. *Trends Genet*. 37(2):174–187. doi:10.1016/j.tig.2020.08.012.
- Steenkamp ET, Wingfield MJ, McTaggart AR, Wingfield BD. 2018. Fungal species and their boundaries matter—definitions, mechanisms and practical implications. *Fungal Biol Rev*. 32(2):104–116. doi:10.1016/j.fbr.2017.11.002.
- Steenfels J, Gallone B, Verstrepen KJ. 2021. Interspecific hybridization as a driver of fungal evolution and adaptation. *Nat Rev Microbiol*. 19(8):485–500. doi:10.1038/s41579-021-00537-4.
- Steenwyk JL, Li Y, Zhou X, Shen X-X, Rokas A. 2023. Incongruence in the phylogenomics era. *Nat Rev Genet*. 24(12):834–850. doi:10.1038/s41576-023-00620-x.
- Téllez MD, Batista-Duharte A, Portuondo D, Quinello C, Bonne-Hernández R, Carlos IZ. 2014. *Sporothrix schenckii* complex biology: environment and fungal pathogenicity. *Microbiology*. 160(11):2352–2365. doi:10.1099/mic.0.081794-0.
- Tralamazza SM, Gluck-Thaler E, Feurtey A, Croll D. 2024. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. *Nat Commun*. 15(1):5728. doi:10.1038/s41467-024-49913-7.
- Uehling J, Deveau A, Paoletti M. 2017. Do fungi have an innate immune response? An NLR-based comparison to plant and animal immune systems. *PLoS Pathog*. 13(10):e1006578. doi:10.1371/journal.ppat.1006578.
- Urban M, Cuzick A, Seager J, Nonavinakere N, Sahoo J, Sahu P, Iyer VL, Khamari L, Martinez MC, Hammond-Kosack KE. 2025. PHI-base—the multi-species pathogen–host interaction database in 2025. *Nucleic Acids Res*. 53(D1):D826–D838. doi:10.1093/nar/gkae1084.
- Urquhart AS, Chong NF, Yang Y, Idnurm A. 2022. A large transposable element mediates metal resistance in the fungus *Paecilomyces variotii*. *Curr Biol*. 32(5):937–950.e5. doi:10.1016/j.cub.2021.12.048.
- Urquhart AS, Gluck E, Vogan AA. 2024. Gene acquisition by giant transposons primes eukaryotes for rapid evolution via horizontal gene transfer. *Sci Adv*. 10(49):eadp8738. doi:10.1126/sciadv.adp8738.
- Urquhart AS, Vogan AA, Gardiner DM, Idnurm A. 2023. *Starships* are active eukaryotic transposable elements mobilized by a new family of tyrosine recombinases. *Proc Natl Acad Sci U S A*. 120(15):e2214521120. doi:10.1073/pnas.2214521120.
- Van Etten J, Stephens TG, Bhattacharya D. 2023. A k-mer-based approach for phylogenetic classification of taxa in environmental genomic data. *Syst Biol*. 72(5):1101–1118. doi:10.1093/sysbio/syad037.
- VanWallendael A, Alvarez M. 2022. Alignment-free methods for polyploid genomes: quick and reliable genetic distance estimation. *Mol Ecol Resour*. 22(2):612–622. doi:10.1111/1755-0998.13499.
- Větrovský T, Morais D, Kohout P, Lepinay C, Algora C, Awokunle Hollá S, Bahnmann BD, Bílohnědá K, Brabcová V, D’Alò F, et al. 2020. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Sci Data*. 7(1):228. doi:10.1038/s41597-020-0567-7.
- Vogan AA, Ament-Velásquez SL, Bastiaans E, Wallerman O, Saupe SJ, Suh A, Johannesson H. 2021. The *Enterprise*, a massive transposon carrying *Spok* meiotic drive genes. *Genome Res*. 31(5):789–798. doi:10.1101/gr.267609.120.
- Wells JN, Feschotte C. 2020. A field guide to eukaryotic transposable elements. *Annu Rev Genet*. 54(1):539–561. doi:10.1146/annurev-genet-040620-022145.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 8(12):973–982. doi:10.1038/nrg2165.
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. 2019. Welcome to the tidyverse. *J Open Source Softw*. 4(43):1686. doi:10.21105/joss.01686.
- Wickham H, Seidel D. 2023. scales: Scale Functions for Visualization. [accessed 2023 Nov 28]. <https://cran.r-project.org/package=scales>.
- Wilke CO. 2024. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.” [accessed 2024 Jan 22]. <https://cran.r-project.org/package=cowplot>.
- Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, Tang W, Feng T, Chen M, Zhan L, et al. 2021. ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol Biol Evol*. 38(9):4039–4042. doi:10.1093/molbev/msab166.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 8(1):28–36. doi:10.1111/2041-210X.12628.
- Zielezinski A, Girgis HZ, Bernard G, Leimeister C-A, Tang K, Dencker T, Lau AK, Röhling S, Choi JJ, Waterman MS, et al. 2019. Benchmarking of alignment-free sequence comparison methods. *Genome Biol*. 20(1):144. doi:10.1186/s13059-019-1755-7.
- Zielezinski A, Vinga S, Almeida J, Karlowski WM. 2017. Alignment-free sequence comparison: benefits, applications, and tools. *Genome Biol*. 18(1):186. doi:10.1186/s13059-017-1319-7.