



Giant *Starship* Elements Mobilize Accessory Genes in Fungal Genomes

Emile Gluck-Thaler ^{‡,1,2} Timothy Ralston,² Zachary Konkel,² Cristhian Grabowski Ocampos,³ Veena Devi Ganeshan,⁴ Anne E. Dorrance,⁵ Terry L. Niblack,² Corlett W. Wood,¹ Jason C. Slot ², Horacio D. Lopez-Nicora,^{*,2,6} and Aaron A. Vogan ^{*,7}

¹Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

²Department of Plant Pathology, The Ohio State University, Columbus, OH, USA

³Facultad de Ciencias Agrarias, Universidad Nacional de Asunción, San Lorenzo, Paraguay

⁴Arabidopsis Biological Resource Center, The Ohio State University, Columbus, OH, USA

⁵Department of Plant Pathology, The Ohio State University, Wooster, OH, USA

⁶Departamento de Producción Agrícola, Universidad San Carlos, Asunción, Paraguay

⁷Systematic Biology, Department of Organismal Biology, University of Uppsala, Uppsala, Sweden

*Corresponding authors: E-mails: lopez-nicora.1@osu.edu; aaron.vogan@ebc.uu.se.

‡Present address: Laboratory of Evolutionary Genetics, University of Neuchâtel, Neuchâtel, Switzerland.

Associate editor: Amanda Larracuent

Abstract

Accessory genes are variably present among members of a species and are a reservoir of adaptive functions. In bacteria, differences in gene distributions among individuals largely result from mobile elements that acquire and disperse accessory genes as cargo. In contrast, the impact of cargo-carrying elements on eukaryotic evolution remains largely unknown. Here, we show that variation in genome content within multiple fungal species is facilitated by *Starships*, a newly discovered group of massive mobile elements that are 110 kb long on average, share conserved components, and carry diverse arrays of accessory genes. We identified hundreds of *Starship*-like regions across every major class of filamentous Ascomycetes, including 28 distinct *Starships* that range from 27 to 393 kb and last shared a common ancestor ca. 400 Ma. Using new long-read assemblies of the plant pathogen *Macrophomina phaseolina*, we characterize four additional *Starships* whose activities contribute to standing variation in genome structure and content. One of these elements, *Voyager*, inserts into 5S rDNA and contains a candidate virulence factor whose increasing copy number has contrasting associations with pathogenic and saprophytic growth, suggesting *Voyager*'s activity underlies an ecological trade-off. We propose that *Starships* are eukaryotic analogs of bacterial integrative and conjugative elements based on parallels between their conserved components and may therefore represent the first dedicated agents of active gene transfer in eukaryotes. Our results suggest that *Starships* have shaped the content and structure of fungal genomes for millions of years and reveal a new concerted route for evolution throughout an entire eukaryotic phylum.

Key words: genome evolution, transposon, mobile genetic element, tyrosine recombinase, pangenome, horizontal gene transfer.

Introduction

A landmark realization of modern evolutionary biology is that genome content is highly variable among individuals of the same species (Brockhurst et al. 2019). Intraspecific genome content variation comprises a species' pangenome, which consists of core (i.e., consistently present) and accessory (i.e., variably present) orthologs. The pangenome concept has adjusted prior expectations of the rate of adaptive trait emergence and dispersal in populations, as differences in accessory sequence distributions across individuals have been found to arise rapidly and fuel phenotypic divergence (Würschum et al. 2015; Melnyk et al. 2019;

Yang et al. 2020). For example, bacterial animal pathogens quickly evolve resistance to antibiotics through the gain of genomic regions encoding resistance phenotypes (Kingsley et al. 2009) and the host range of many fungal plant pathogens often expands through the gain of entire chromosomes encoding host-specific virulence factors (Ma et al. 2010). Identifying the mechanisms generating selectable variation in the distribution of accessory genes among individuals is thus crucial for understanding the genetic bases of adaptation (Maistrenko et al. 2020).

Genome content variation in prokaryotes is primarily driven by mobile elements that in addition to encoding functions for their own movement, also encode accessory

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

genes (Brockhurst et al. 2019). These “cargo-carrying” mobile elements range in size, from smaller transposons to larger genomic islands to entire DNA molecules such as plasmids, and interact with their host genomes in diverse ways (Dobrindt et al. 2004; te Poele et al. 2008; Benler et al. 2021). Although some mobile elements resemble parasites because their proliferation is predicted to come at the expense of host fitness (Brown et al. 2013), cargo-carrying elements resemble mutualists that cooperate with host genomes because they often carry genes encoding adaptive host traits (Brockhurst et al. 2019). For example, several well-known integrative and conjugative elements carry accessory genes mediating advantageous symbiosis and antibiotic resistance phenotypes (Sullivan and Ronson 1998; Verma et al. 2019). Cargo-carrying mobile elements are frequently transmitted through both horizontal gene transfer (HGT) and vertical inheritance, which together generate extensive genome content variation within and across species (Brockhurst et al. 2019).

In contrast with prokaryotes, the contribution of mobile elements to genome content variation in eukaryotes remains unclear, although is of increasing interest (Inoue et al. 2018; Arkhipova and Yushenova 2019; Moniruzzaman and Aylward 2021). Eukaryotic genomes are often large and repeat-rich, which complicates the characterization of accessory genes using short-read sequencing methods (Brockhurst et al. 2019). However, long-read sequencing technologies that resolve repetitive sequence content now enable more contiguous genome assemblies in which the movements and associations of mobile elements can be tracked (Arkhipova and Yushenova 2019). Several small cargo-carrying eukaryotic mobile elements in specific species have recently been shown to generate genome content variation by increasing the copy numbers of their accessory genes (Dallaire et al. 2021; Wang et al. 2021). Larger eukaryotic mobile elements have only been identified recently, and though little is known about these elements, initial studies suggest they impact host fitness, that some are horizontally acquirable, and are pervasive across multiple species (Inoue et al. 2017, 2018; McDonald et al. 2019; Vogan et al. 2021; Urquhart et al. 2022). Yet despite compelling preliminary evidence, the diversity, phylogenetic relationships, and contributions toward genome and pangenome evolution of large mobile elements are not well understood. Fungi are ideal for investigating these aspects of large mobile element biology due to their relatively small genomes with resolvable repeat content, notable associations between several recently described large mobile elements and phenotypes (McDonald et al. 2019; Vogan et al. 2021; Urquhart et al. 2022), and a vast diversity of sequenced genomes that enable kingdom-wide analyses.

Here, we investigated the contributions of large cargo-carrying mobile elements to the evolution of Ascomycete fungi, one of the most speciose, ecologically diverse, and economically important groups of microbial eukaryotes. At first, we set out to characterize the genetic mechanisms of pathogenicity in the plant pathogen *Macrophomina phaseolina* because, whereas this fungus causes disease on hundreds of plant species, including

many economically important crops, little is known about why it is a successful pathogen. We sequenced 12 isolates from North and South America using a combination of short- and long-reads and found evidence of genome content variation even among near clonal isolates. This observed content variation, coupled with the contiguity of the assembled genomes, then allowed us to explicitly test the hypothesis that mobile element cargo carriers are a mechanism generating genome content variation among individuals. After finding several giant mobile elements within the *M. phaseolina* genomes, we searched for similar elements in other Ascomycetes to test the hypothesis that they all share a common origin. Returning to *M. phaseolina* as a case study, we ask how these elements shape the evolution of fungal genome structure, genome content, and ecologically relevant phenotypes such as plant pathogenicity.

Results and Discussion

M. phaseolina Genomes Harbor Insertions that Resemble Mobile Elements

We sequenced 12 *M. phaseolina* isolates from soybean fields in OH, USA and Paraguay using a combination of Nanopore long-read and Illumina short-read technologies. We used multiple assemblers and gene annotators to obtain near-chromosome level quality draft assemblies and annotations (avg. N50 = 2.43 mb; avg. assembly size = 52.31 mb) (supplementary fig. S1 and tables S1–S3, Supplementary Material online, see Materials and Methods). Isolates differed by at most 1% nucleotide identity across all alignable genomic sequence, as was expected given the predominance of clonal reproduction in this species (Islam et al. 2012). In contrast, we found that isolates harbor extensive variation in accessory genes (i.e., genes missing from at least one isolate), with gene content diverging 20 times faster than nucleotide sequence per unit of evolutionary time and differing by up to 36% (supplementary fig. S2 and table S4, Supplementary Material online). *Macrophomina phaseolina*'s pangenome is “open” (Heaps' $\alpha = 0.46$), meaning that the size of its pangenome is predicted to increase with each newly examined genome. *Macrophomina phaseolina*'s pangenome is more open than most other fungal species examined, suggesting that individuals within this species generally harbor high proportions of accessory genes (supplementary fig. S2 and tables S5 and S6, Supplementary Material online).

We tested whether differences in genome content arise randomly across the genome, or whether they are concentrated in specific “hotspot” regions that contain accessory genes with higher than expected rates of gain and loss (see Materials and Methods). On average, 18% of each genome is partitioned into hotspot regions dispersed across different contigs (supplementary fig. S2 and tables S7 and S8, Supplementary Material online). We find no evidence that hotspot regions are enriched in putative subtelomeric regions (Fisher's exact test; $P = 0.99$), although our analysis

was limited to only those contigs with telomeric repeats on either end (see Materials and Methods). Compared with nonhotspot regions of the genome, hotspot regions have significantly higher GC content (Wilcoxon rank-sum test; $W = 1,525,087$, $P < 2.2e-16$), greater coding sequence content per 10 kb ($W = 1,718,643$, $P < 2.2e-16$), and lower repetitive sequence content per 10 kb ($W = 819,608$, $P < 2.2e-16$). Accessory gene hotspots do not differ from the overall genome in proportion of SNPs or indels ≤ 1 kb, but are enriched for indels > 1 kb (see Materials and Methods), suggesting that gene variability is associated with large structural mutations (supplementary fig. S2 and table S9, Supplementary Material online).

We hypothesized that the variation in *M. phaseolina*'s accessory genome and the generation of large structural mutations in hotspots is facilitated by large cargo-carrying mobile elements. We searched the 12 *M. phaseolina* genomes for large insertions that had three key features typified by the recently described giant fungal transposons *Enterprise* and *HEPHAESTUS* (Vogan et al. 2021; Urquhart et al. 2022): (1) a predicted gene containing a DUF3435 domain (protein family accession: PF11917) located at the 5' end of the element in the 5'–3' direction. DUF3435 domains are homologous to and share conserved active sites with tyrosine site-specific recombinases that catalyze the excision and integration of bacterial integrative and conjugative elements (Johnson and Grossman 2015; Smyshlyayev et al. 2021; Vogan et al. 2021); (2) flanking target site duplications (TSDs) that arise when DNA transposons integrate into their target site through staggered double-stranded breaks (Wicker et al. 2007); (3) terminal inverted repeats (TIRs), which are typically required for the excision of DNA transposons by transposases, although are not present in *Enterprise* (Wicker et al. 2007; Vogan et al. 2021). We found four very large (76,706–195,340 bp) elements we call *Voyager*, *Argo*, *Phoenix*, and *Defiant* that all have a DUF3435 gene in the required position and often have TSDs and TIRs as well (fig. 1B and table 1, supplementary table S10, Supplementary Material online). These four elements are discrete and distinct entities that carry diverse arrays of accessory cargo and differ in their target sites and TIRs. However, despite this variability, DUF3435 genes in these elements often colocalize with four additional genes that are present in variable copy numbers and conserved across elements, consisting of DUF3723s, ferric reductases (FREs), patatin-like phosphatases (PLPs), and NOD-like receptors (NLRs) with different functional domains (table 2, supplementary table S11, Supplementary Material online). We refer to these genes by their predicted function throughout the text.

Starships are a Newly Discovered Group of Giant Mobile Elements Widespread Across the Pezizomycotina

We hypothesized that the conserved associations between DUF3435s, DUF3723s, FREs, PLPs, and NLRs are indicative

of a “winning” coalition of cofunctioning genes maintained through selection (Baquero 2004), much like the multi-component translocative machinery of bacterial integrative and conjugative elements (te Poele et al. 2008). We, therefore, predicted that they would appear together in other fungi. We used the genes in *M. phaseolina*'s elements to query a database of 1,649 Basidiomycotina, Saccharomycotina, and Pezizomycotina genomes, and found 134 regions from the Pezizomycotina containing homologs to DUF3435 that also colocalized to varying degrees with DUF3723s, FREs, PLPs, and NLRs (fig. 1D, supplementary tables S12 and S13, Supplementary Material online). Using whole-genome alignments, we manually annotated the insertion sites, TSDs, and TIRs for 28 of these elements sampled across all extant classes of filamentous Pezizomycotina, providing evidence of pervasive and recent mobile activity across one of the most diverse and speciose eukaryotic subphyla (fig. 1B, supplementary table S14 and figs. S4 and S5, Supplementary Material online).

We used the phylogenetic history of DUF3435 genes at the 5' element ends, which we refer to as “captains,” to characterize the evolutionary relationships between the various elements (see Materials and Methods). Not only did we find that captains from different elements fall into distinct clades (fig. 1A, supplementary figs. S5 and S6, Supplementary Material online), but we also found that the tyrosine recombinase domains of all DUF3435 captains group into a monophyletic clade that is as distantly related to bacterial tyrosine recombinases as it is to described tyrosine recombinases found in either fungal *Crypton* DNA transposons (Goodwin et al. 2003) or eukaryotic retrotransposons, such as DIRS elements (Piednoël et al. 2011) (supplementary fig. S7, Supplementary Material online). We conclude that DUF3435-containing genes represent a novel group of eukaryotic tyrosine recombinases that mobilize large amounts of DNA. Following Vogan et al. (2021), we consider regions bounded by TSDs that contain a DUF3435 sequence at the 5' end to be a new class of eukaryotic mobile element that we call *Starships* (fig. 1C). We distinguish between different types of *Starship* elements based on their insertion sites and the phylogenetic relationships between their DUF3435 captains. Regions that contain DUF3435 sequences that colocalize with DUF3723s, FREs, PLPs, and/or NLRs but lack TSDs are referred to as *Starship*-like, and regions that contain at least three hits to any of these conserved genes but lack DUF3435 sequence are considered to be degraded *Starships*.

The DUF3435 phylogeny revealed that the captains of different *Starships* likely last shared a common ancestor before the divergence of the Pezizomycotina, ca. 400 Ma (Shen et al. 2020) and thus have a single origin (fig. 1A). This phylogeny unites previously disparate phenomena in fungi, including *Enterprise* (Vogan et al. 2021), *HEPHAESTUS* (Urquhart et al. 2022), a large mobile region containing the ToxhAT transposon and ToxA effector found in numerous wheat pathogens (McDonald et al.

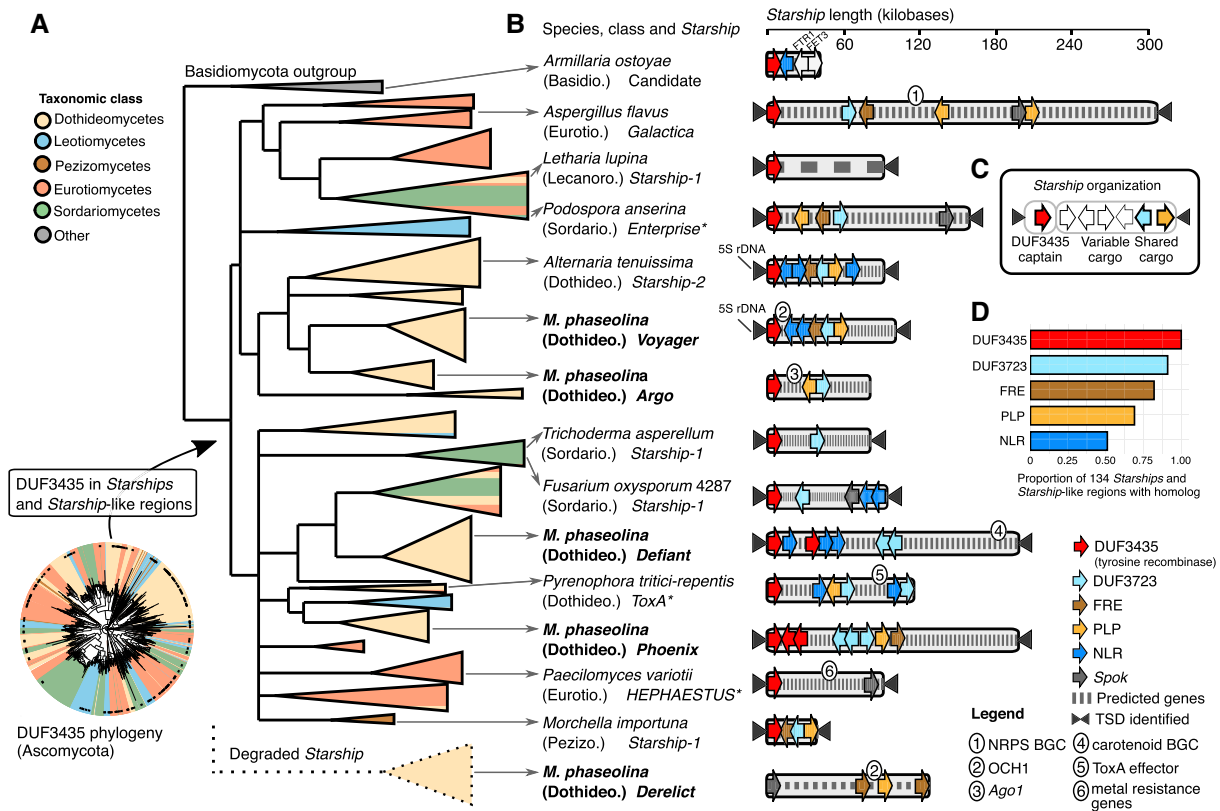


FIG. 1. Giant Starship elements found across the Ascomycota share conserved components and mobilize diverse arrays of accessory genes. (A) Bottom-left: a circular maximum likelihood (ML) tree of 1890 DUF3435 homologs retrieved from a database containing 12 Macrohomina phaseolina genomes generated for this study and 1,274 published Ascomycete genomes, where branches are color-coded according to taxonomic class. Black dots along the perimeter indicate DUF3435 sequences found in 194 Starships and Starship-like regions (60 of which are from *M. phaseolina*). Right: a ML tree of these sequences and several others of interest, where clades are depicted as triangles and all branches with <80% SH-aLRT and <95% UFboot support have been collapsed. A clade of degraded Starships that lack DUF3435 sequences is juxtaposed below the tree for visualization purposes only. (B) Fungal species, taxonomic class, Starship name, and simplified Starship schematics, with approximate locations of genes of interest depicted as colored arrows or numbered. Genes are not drawn to scale. TSDs, when indicated, were confirmed from alignments of insertion polymorphisms. Names of elements from *M. phaseolina* are in bold, and previously published mobile elements are indicated with an asterisk. (C) A schematic showing the general organization of Starship mobile elements, where captain location and orientation are always conserved. (D) Proportion of 134 Starships and Starship-like regions containing homologs to five genes of interest. Gene name abbreviations: FRE (ferric reductase); PLP (patatin-like phospholipase); NLR (Nucleotide Oligomerization Domain-like receptor); Spok (spore killer); OCH1 (alpha-1,6-mannosyltransferase); NRPS (nonribosomal peptide synthetase); Ago1 (Argonaute). Other abbreviations: BGC (biosynthetic gene cluster). All original trees can be found in the [Supplementary Material](#) online.

Table 1. Summary Statistics of Starship Elements in *Macrohomina phaseolina*.

Starship Element	Avg. Length in bp (s.d.)	Avg. Number of Genes (s.d.)	Avg. % Repeat Content (s.d.)	Copies per Isolate, min-max	Target Site
Voyager	96,567 (13,781)	37 (9)	30.4 (10.4)	0–4	GCCTGTATTCT
Argo	75,148 (14,824)	26 (4)	9.5 (2.1)	0–1	n.d.
Phoenix	74,447 (94,611)	21 (25)	44.6 (25.7)	0–3	TTTACA
Defiant	108,860 (57,441)	28 (15)	37.7 (15.5)	0–5	ACTA

2019), and multiple previously described large structural variants whose origins were unclear (Galactica_Afla, Starship-1_Afum) (Fedorova et al. 2008; Fountain et al. 2020), by demonstrating that they are all Starships. Although the original descriptions suggest Enterprise and HEPHAESTUS only share a DUF3435-containing gene, re-analyzing these elements revealed that one version of Enterprise in *Podospora anserina* (Psk-2) contains the DUF3723, FRE, and PLP genes, whereas HEPHAESTUS

possesses a pseudogenized DUF3723 domain, corroborating our findings. The majority of DUF3435s in the Ascomycota are largely restricted to clades consisting of sequences from the same taxonomic class and multiple class-specific clades exist, consistent with an evolutionary scenario dominated by the vertical inheritance of ancient homologs (supplementary fig. S4, Supplementary Material online). This same trend is recapitulated within individual taxa, such as *Macrohomina*, *Aspergillus*, and *Fusarium*

Table 2. Conserved *Starship* Cargo Genes.

Gene Name	Voyager Reference Gene(s)	Protein Family Domains	Predicted Function
DUF3435	mp102_11332	PF11917 (DUF3435)	Tyrosine recombinase
DUF3723	mp102_11356	PF12520 (DUF3723)	Unknown
FRE	mp102_11352	PF01794 (Ferric_reduc) PF08022 (FAD_binding_8) PF08030 (NAD_binding_6)	Ferric reductase
PLP	mp102_11357	PF01734 (Patatin)	Patatin-like phosphatase
NLR	mp102_11337	PF17107 (SesA) PF05729 (NACHT) PF12796 (Ank_2) PF00023 (Ank)	NOD-like receptor
	mp102_ofg737	PF06985 (HET) PF05729 (NACHT)	NOD-like receptor

that contain multiple highly diverged and distantly related DUF3435s within their genomes (supplementary fig. S5, Supplementary Material online). However, HGT has been implicated in *Starship* transfers among closely related taxa in at least two independent examples (McDonald et al. 2019; Urquhart et al. 2022) as well as in the transfer of the genomic island *Wallaby* between *Penicillium* spp. whose captain was not originally included in the DUF3435 phylogeny but otherwise meets all of the criteria to be a *Starship* (Cheeseman et al. 2014) (supplementary table S14, Supplementary Material online). Given that rates of HGT are expected to increase with donor–recipient relatedness (Andam and Gogarten 2011), this suggests that *Starship* HGT may be more prevalent at finer-grained taxonomic scales beyond the scope of this study.

Starship Accessory Cargo Suggests a Range of Impacts on Host Fitness

The diversity of accessory cargo found across *Starships* suggests their interactions with host genomes span the conflict-cooperation spectrum. For example, some predicted cargo suggests mobile element fitness could come at the host's expense: numerous *Starships* contain homologs to spore killing (*Spok*) genes that encode a toxin–antitoxin meiotic drive system capable of killing developing sexual progeny in *Enterprise* (Vogan et al. 2019, 2021) (fig. 1B, supplementary table S13, Supplementary Material online). In contrast, many cargo genes have putatively adaptive functions that would hypothetically benefit the host in certain environments: several carry biosynthetic gene clusters (BGCs) involved in the production of secondary metabolites such as carotenoids and nonribosomal peptides (e.g., *Defiant* and *Galactica*); others carry characterized virulence factors or homologs of known virulence factors (e.g., *ToxA*, McDonald et al. 2019; *Ago1* in *Argo*, Habig et al. 2021; *OCH1* in *Voyager* and *Derelect*, Bates et al. 2006; Li et al. 2014; Zhang et al. 2019); *HEPHAESTUS* contains several genes characterized to contribute to heavy metal tolerance (Urquhart et al. 2022); *Wallaby* contains multiple genes predicted to be involved in competition (Cheeseman et al. 2014); finally, FREs that are conserved across multiple *Starships* are part of a large metalloredox family with several characterized

members, suggesting some confer benefit to the host, for example, in the acquisition of iron in growth-limiting environments or contributions to pathogen virulence (Blatzer et al. 2011; Saikia et al. 2014; Rehman et al. 2018). Although the few *Starship*-like regions we identified in our Basidiomycete outgroup lacked FRE homologs, they contain homologs to *Fet3p* and *Ftr1p* which form a high affinity iron permease complex in yeast (Singh et al. 2006), suggesting a functional association convergently evolved between iron metabolism and *Starships* (fig. 1B). Carrying host-beneficial cargo may be an especially important trait for elements that are predominantly inherited through vertical means, as their own fitness is closely aligned with their host's reproductive success (Billane et al. 2022).

Starship Activity Shapes the Distribution of Accessory Genes in *M. phaseolina* Populations

We carried out an in-depth characterization of *M. phaseolina*'s *Starships* to better understand how these large elements shape variation in accessory genome content. *Starship* elements in *M. phaseolina* differ in copy number, genetic cargo, and generally have little alignable sequence in common such that they each contribute in distinct ways to content divergence across *M. phaseolina* isolates (fig. 2A and table 1, supplementary table S15, Supplementary Material online). Extensive content variation even exists within certain *Starship* element types, such as *Phoenix* and *Defiant*. For example, we found a putative cargo turnover event in two *Defiant* elements that carry either a carotenoid BGC or a nonribosomal peptide synthetase BGC (fig. 2E). Within any given genome, 73–100% of *Starships* and *Starship*-like regions overlap with accessory gene hotspots indicating that these regions are hotspots for gene gain and loss (see Materials and Methods, fig. 2C, supplementary table S16, Supplementary Material online). Although other genetic mechanisms are also likely to play an important role in shaping accessory content variation, 14.3% of all hotspots in the *M. phaseolina* population overlap with *Starship* and *Starship*-like regions and an average of 2% of each genome consists of *Starships* or *Starship*-like regions, highlighting the appreciable impact and transformative potential of this specific family on

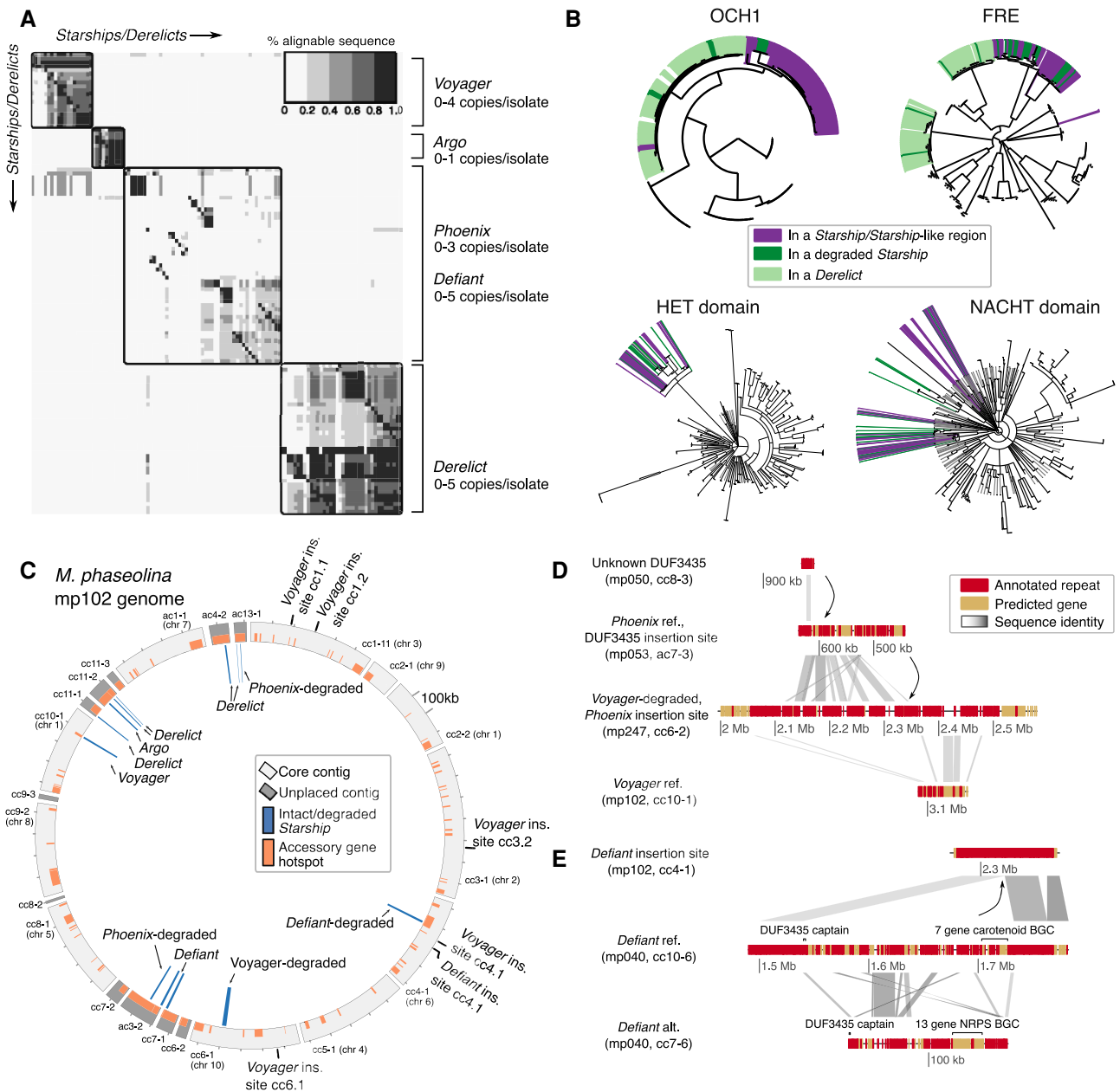


Fig. 2. Distinct *Starship* elements in *Macrophomina phaseolina* vary in copy number across individuals and generate intraspecific variation in genome content. (A) A heatmap summarizing the percentage of pairwise alignable sequence between all *Starships* and *Starship*-like regions, as well as the degraded *Starship* region *Derelict*, found across 12 *M. phaseolina* isolates, where rows and columns represent individual regions arranged by hierarchical clustering according to similarity in their alignment scores. (B) Maximum likelihood phylogenies of genes homologous to OCH1 and FREs, as well as genes containing and HET and NACHT domains found in fungal NLRs, retrieved from the *M. phaseolina* genomes, where all branches with <80% SH-aLRT and <95% UFboot support have been collapsed. Genes found in regions of interest are highlighted. (C) A circos plot depicting all contigs >100 kb in the *M. phaseolina* mp102 genome. Core contigs are partially alignable across all isolates, whereas unplaced contigs are not. Regions of interest are labeled and visualized on the inner track. Sites with resolved *Starship* insertions in other genomes, but absent in mp102, are labeled on the outside track. Accessory gene hotspots are labeled in orange on the outer track. (D) A sequential *Starship* insertion event where an unknown DUF3435 gene is found nested in *Phoenix*, which itself is nested in a degraded *Voyager* element in isolate mp247 on contig cc6-2. (E) A cargo turn-over event between two *Defiant*s carrying either a carotenoid or nonribosomal peptide synthetase BGC. Gene name abbreviations: OCH1 (alpha-1,6- mannosyltransferase); FRE (ferric reductase); NLR (Nucleotide Oligomerization Domain-like receptor).

genome content as a whole (supplementary tables S7 and S10, Supplementary Material online). Specific *Starship* elements are further linked to the expansion of accessory gene families they carry as cargo (fig. 2B). For example, over half of the OCH1 homologs (59/109) and just over

a third of the FRE homologs (112/316) in the *M. phaseolina* population are located in *Starship* or *Starship*-like regions (fig. 2B, supplementary figs. S8 and S9, Supplementary Material online). Conserved cargo genes and functional domains carried on *Starships*, including OCH1, PLP, FRE,

DUF3723, and the HET and NACHT domains in NLRs, are typically restricted to clades of *Starship*-specific sequences, indicative of cargo-*Starship* codiversification (fig. 2B, supplementary figs. S8–S13, Supplementary Material online). Content diversity and flexibility, coupled with a transposition mechanism, underscore the potential for *Starships* to shape the distribution of accessory genes among individuals.

Not only do *Starship* elements actively generate new variation, but remnants of degraded *Starship*-like regions found throughout the genomes of *M. phaseolina* suggest their impact extends beyond the life of the element. For example, we discovered a group of regions, which we call *Derelicts*, that have *Starship*-specific cargo but lack DUF3435 homologs (fig. 1A, supplementary table S11, Supplementary Material online). Like active *Starships*, *Derelicts* show large swaths of similarity among isolates and copies (avg. 41% alignable sequence), but otherwise are distinct from other elements (fig. 2A). They have homologs of FRE, DUF3723, and OCH1 that form large monophyletic clades within their phylogenies, and these sequences are in general more closely related to their homologs in other *Starships* than they are to non-*Starship*-associated sequences, suggesting that *Derelicts* are old copies of a *Starship* whose captain degraded but whose cargo has persisted (fig. 2B, supplementary figs. S8, S9, and S13, Supplementary Material online). We find many additional *Starship*-like regions that similarly lack DUF3435 homologs but have conserved *Starship* content and occasionally aligned to characterized *Starship* elements, suggesting they have recently degraded (supplementary table S11, Supplementary Material online). Across all genomes, 72–100% of these degraded regions overlapped with an accessory gene hotspot, indicating that they have remained a source of genome content variation among individuals (supplementary table S16, Supplementary Material online). Across the 1,274 Ascomycete genomes in our database, we similarly found 102 *Starship*-like regions that did not have DUF3435 genes, but had at least four hits to FRE, DUF3723, NLRs, and PLP sequences, suggesting these regions may also be degraded *Starship* elements (supplementary table S13, Supplementary Material online). Evidence of cargo that persists after element degradation implicates *Starship* birth-and-death cycles as a fundamental mechanism shaping variation in genome content among individuals.

Starships Contribute to Standing Variation in the Structure of *M. phaseolina*'s Genome

Many fungal genomes are compartmentalized into accessory chromosomes that are variably present among individuals, and core chromosomes with large tracts of conserved synteny across individuals (Stukenbrock and Croll 2014). In *M. phaseolina*, we identified ten large “core contigs” that are largely syntenic, often have telomeric repeats at either end, and are present across all individuals (fig. 3, supplementary fig. S3 and tables S17 and

S18, Supplementary Material online, see Materials and Methods). We found that on average, 88.6% of each *M. phaseolina* genome is captured by these core contigs, whereas the rest is located in unplaced contigs that do not align to the core contigs (supplementary table S19, Supplementary Material online). However, core contigs in the *M. phaseolina* population are highly rearranged and contain large tracts of variably present sequence, such that none of the core contigs in the eight highest-quality assemblies share the same structure or content across all isolates (supplementary fig. S3, Supplementary Material online, see Materials and Methods), consistent with previous observations in this species (Burkhardt et al. 2019).

We, therefore, examined associations between *M. phaseolina*'s *Starships* and large-scale differences in genome structure, which may be facilitated by the ability of *Starships* to insert into otherwise conserved sequence tracts (fig. 3B, event 3). We identified 37 large translocations, inversions, and putative chromosome fusions in the *M. phaseolina* population, and 30% of these had *Starships*, *Starship*-like regions, or degraded *Starships* overlapping or near their breakpoints (supplementary table S20, Supplementary Material online). For example, a degraded *Defiant* element is found at the junction of a putative fusion between core chromosomes 9 and 1 in *M. phaseolina* isolate mp247, and a reciprocal translocation between core chromosomes 9 and 3 in *M. phaseolina* isolate mp042 occurs precisely at a *Voyager* insertion (fig. 3B, events 1 and 2). Due to their sheer size, sequence similarity, and the fact that they are themselves colonized by multi-copy mobile elements including LTRs (supplementary table S10, Supplementary Material online), *Starship* insertions may provide opportunities for rearrangements to occur through ectopic recombination (Faino et al. 2016), although our data cannot confirm whether they are the causal mechanism. However, 95% of the rearrangements overlapped with some sort of mobile element, in general agreement with previous observations that mobile elements are associated with large-scale rearrangements in fungal genomes (Faino et al. 2016; Badet et al. 2021).

Starship activity is also linked to the active growth of large genomic regions whose presence varies among individuals. Gene phylogenies of DUF3435s and other conserved cargo suggest *Starships* can sequentially nest within each other, generating large tracts of accessory sequence (supplementary fig. S6, Supplementary Material online). For example, a degraded *Voyager* insertion in *M. phaseolina* isolate mp247 has at least three *Starship* or *Starship*-like elements that have sequentially inserted into each other to produce a repeat-rich region spanning 221,058 bp (fig. 2D). Mobile element nesting is a common mechanism of bacterial integrative and conjugative element growth (Bellanger et al. 2014) and has also been observed in *HEPHAESTUS* (Urquhart et al. 2022). In addition to generating variable regions on otherwise core contigs, *Starships* and *Starship*-like regions may also

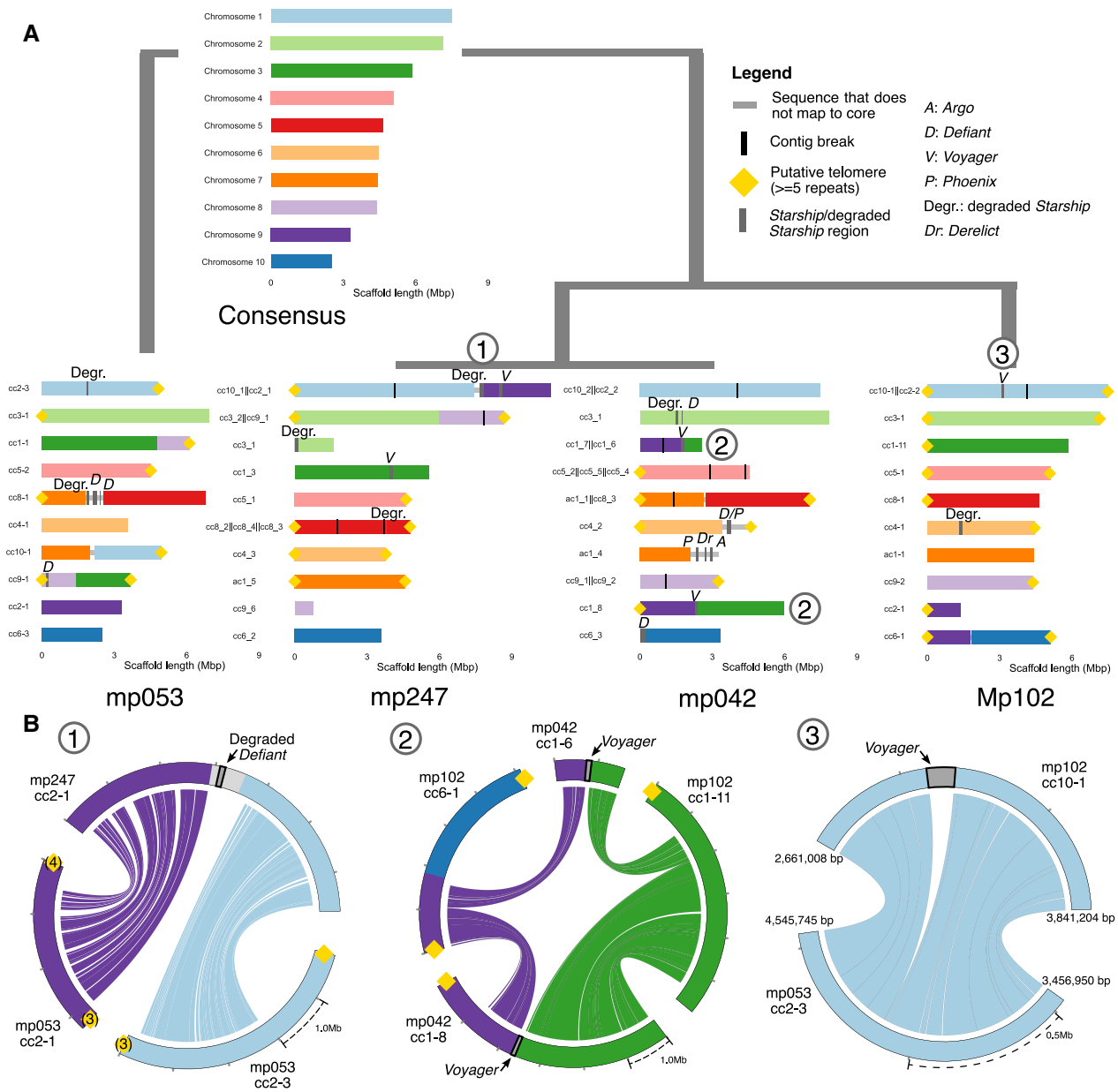


FIG. 3. Several *Starship* elements are associated with large structural variants across *Macrophomina phaseolina* genomes. (A) Alignment-based chromosome maps of four *M. phaseolina* isolates and a hypothetical ancestral genome, where contigs that share homology are colored and ordered according to their putative chromosome of origin. Putative telomeres are labeled at contigs ends, and consist of at least five sequential telomeric repeats, unless indicated otherwise by bracketed numbers. Unplaced contigs are not illustrated. (B) From left to right, circos-based nuclemer alignments of (1) a putative chromosomal fusion, (2) a reciprocal translocation, and (3) a *Voyager* insertion polymorphism. Contigs are labeled according to chromosome of origin, or otherwise colored gray. *Starships* and degraded *Starships* are drawn in dark gray, labeled, and outlined in black. Labeled numbers correspond to events depicted in both the chromosome maps and the circos alignments.

contribute to the birth and/or growth of accessory chromosomes. Contig ac11-1, a putative accessory chromosome in *M. phaseolina* isolate mp053, is ~800 kb, with *Starship*-like regions composing roughly 16% of its length.

Voyager: A Massive Mobile Element Targeting 5S rDNA

The most active *Starship* in the *M. phaseolina* population is *Voyager*, ranging in copy number from 0 to 4 across individuals (fig. 4A). We resolved six different *Voyager*

insertion sites that have clear breakpoints and high Illumina read coverage across occupied and empty sites (fig. 4C, supplementary table S21, Supplementary Material online). In isolates with a given *Voyager* insertion, flanking regions are syntenic with the corresponding insertion site in isolates that do not have the insertion, ruling out the alternative explanation that *Voyager* has moved by chromosomal translocation. Based on these resolved insertion sites, we recovered 15 full-length copies of *Voyager* with identifiable TSDs and TIRs (fig. 4B and table 1). We also found an additional seven fragmented copies from

the two lowest quality assemblies (mp124 and mp194) that had recoverable DUF3435 sequences, but were excluded from the following summary statistic calculations. *Voyager* elements range from 74,275 to 124,223 bp in length (avg. 96,567 bp), contain between 10 and 47 predicted protein-coding sequences (avg. 37), and are alignable from start to end. Variation in size and in protein-coding sequences among *Voyager* copies is largely due to variation in transposable element content, which ranges from 15% to 50% (avg. 31.2%; [supplementary table S21, Supplementary Material](#) online).

Although *Voyager* copies are found at multiple sites in any given genome, they are always inserted into a copy of the 5S rDNA gene, 20 bp from the gene's 5' start site, and 110 bp from the 3' end. The insertion of *Voyager* into the 5S rDNA gene shifts the free energy of the 5S rRNA's predicted secondary structure from an estimated -42.22 to -38.49 kcal/mol such that we predict it loses function. Due to experiencing strong purifying selection ([Rooney and Ward 2005](#)), rDNA sequences have been proposed as stable mobile element target sites enabling long-term persistence ([Eickbush and Eickbush 1995](#)). Similarly, many bacterial integrative and conjugative elements insert in or near tRNA genes which would also provide stable target sites ([Dobrindt et al. 2004](#); [Johnson and Grossman 2015](#)). We estimate that *M. phaseolina* has upwards of 45 copies of 5S rDNA genes dispersed across its genome that are fixed in the population, similar to copy numbers observed in other Ascomycete fungi ([Rooney and Ward 2005](#)). *Macrophomina phaseolina* has at least two distinct variants of the 5S rDNA gene, and *Voyager* insertions and target sites are only present in one of them, raising the possibility that changes to the 5S sequence of this species are selected to avoid being disrupted by *Voyager*. Additionally, we identified two *Starships*, one in *Alternaria tenuissima* (*Starship-2_Aten*) whose captain is closely related to *Voyager's*, and one in *Aspergillus niger* (*Starship-1_Anig*), whose target sites are similarly located inside 5S rDNA sequence ([fig. 1B](#); [supplementary table S14, Supplementary Material](#) online). This suggests that targeting conserved host genes is a broadly successful *Starship* strategy even though it may open the door to potential conflict with host genomes.

Voyager has Contrasting Associations with Pathogenic and Saprophytic Growth

Insight into the genetic bases of virulence in *M. phaseolina* is critically needed because this pathogen represents an emerging threat to food security: it can infect more than 500 plant species, including many agricultural crops, and is increasingly expanding into new geographic regions. Although mobile element insertions in rDNA tend to negatively impact growth ([Templeton et al. 1993](#); [Weider et al. 2005](#)), the majority of *Voyager* elements also carry homologs of the virulence factor OCH1, a conserved α -1,6-mannosyltransferase that contributes to cell wall integrity and pathogenic growth in diverse plant and animal pathogens ([Bates et al. 2006](#); [Li et al. 2014](#); [Zhang et al. 2019](#)). We, therefore, hypothesized that *M. phaseolina's* growth on

its host would increase with the number of OCH1-carrying *Voyager* elements present in the genome. We inoculated soybean seedlings in a greenhouse, and measured the area under the disease progress curve (AUDPC) as a proxy for pathogenic growth ($n = 12$ isolates, 12 replicates/isolate; [supplementary table S22, Supplementary Material](#) online). We similarly inoculated potato dextrose agar (PDA) plates placed at six different temperatures in an incubator, and measured the area under the growth progress curves (AUGPC) as proxies for saprophytic growth ($n = 12$ isolates, three replicates/isolate/temperature; [supplementary table S23, Supplementary Material](#) online). Using linear models corrected for phylogenetic relatedness among isolates ([supplementary table S24, Supplementary Material](#) online), we found a significant positive linear and negative quadratic relationship between *Voyager* OCH1 copy number and pathogenic growth ($y = 138.3 + 93.6x - 22.5x^2$; $P_{\text{linear}} = 0.05$, $P_{\text{quadratic}} = 0.03$) that suggests the predicted contributions of *Voyager* OCH1 to pathogenicity are greatest at intermediate copy number but then decline. Conversely, we found a marginally significant negative linear and positive quadratic relationship for saprophytic growth at 20 °C ($y = 201.1 - 31.6x + 7.3x^2$; $P_{\text{linear}} = 0.10$, $P_{\text{quadratic}} = 0.07$) and a significant negative linear and positive quadratic relationship for saprophytic growth at 30 °C ($y = 306.7 - 35.9x + 7x^2$; $P_{\text{linear}} < 0.01$, $P_{\text{quadratic}} < 0.01$) that together suggest the contributions of *Voyager* OCH1 to saprotrophy are lowest at intermediate copy number. These results are consistent with a scenario where *Voyager* incurs an ecological trade-off between *M. phaseolina's* ability to grow on or off of its host, and suggests alternating host and nonhost environments help maintain variation in virulence phenotypes in natural populations. However, our small sample size ($n = 12$) demands that observed trends be confirmed once more individuals are genotyped for *Voyager* insertions. The complex associations of *Voyager* with fungal-plant interactions may predate the origin of the *Macrophomina* genus because we identified a putatively vertically inherited copy of *Voyager* in the closely related plant pathogen *Botryosphaeria dothidea* that has the same target site and also carries OCH1, but may not be presently active due to an LTR insertion into the reading frame of the captain ([supplementary table S14 and fig. S8, Supplementary Material](#) online). Vertical inheritance of the same element strengthens the hypothesis that conserved host genes provide stable target sites, and underscores how this strategy may enable persistence through speciation events.

Are *Starships* Agents of Active Gene Transfer in Eukaryotes?

Starships are distinct from other eukaryotic mobile elements not only in their consistently large size, but in that many of them share a conserved set of accessory genes ([table 2](#)). Although the functions of these genes are not known, they present parallels to translocative and dispersive modules in bacterial integrative and conjugative

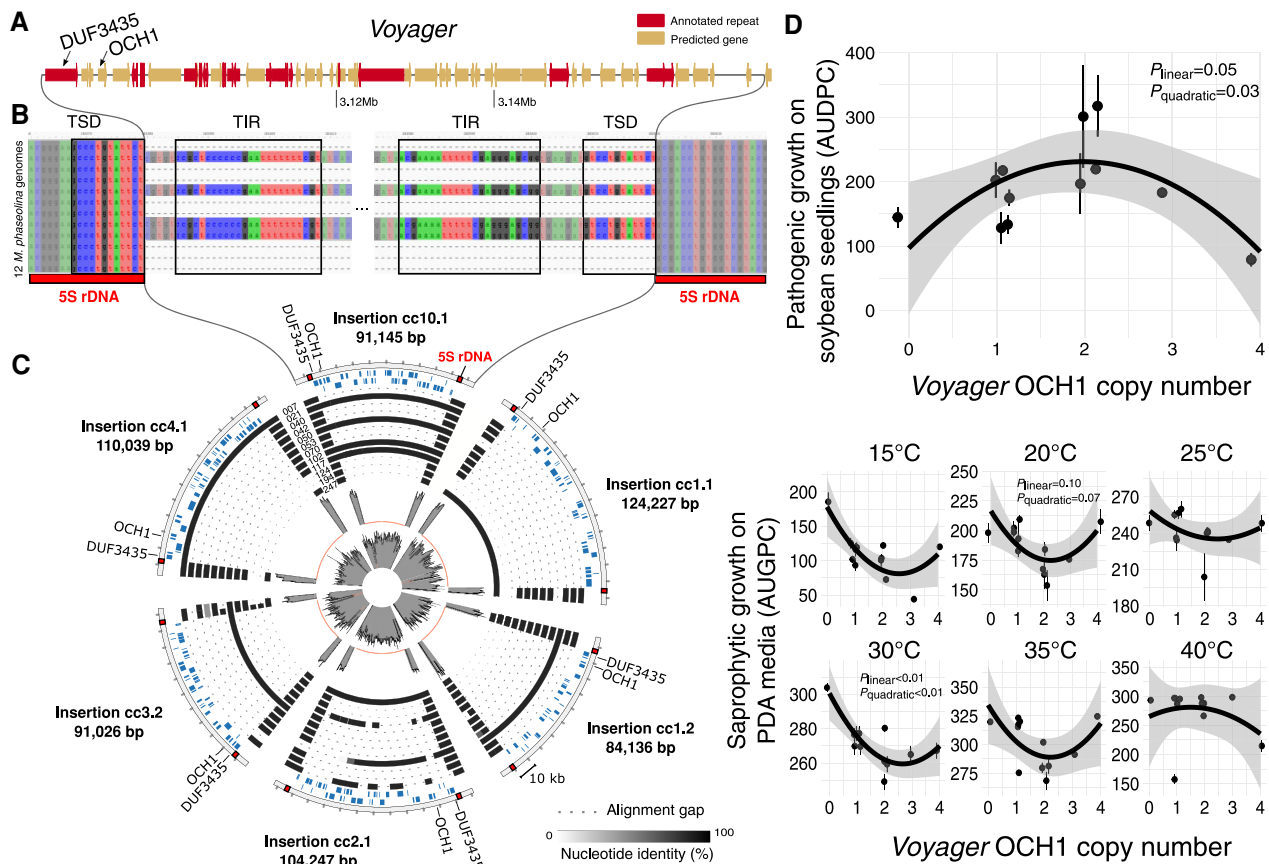


Fig. 4. Increasing *Voyager* OCH1 copy number has contrasting effects on *Macrophomina phaseolina*'s pathogenic and saprotrophic growth. (A) A schematic of the reference *Voyager* element in *M. phaseolina* mp102 depicting all predicted genes and annotated repetitive elements. (B) A multiple sequence alignment of the *Voyager* insertion cc10.1, with annotated target site duplications and terminal inverted repeats. (C) A circos plot of six fully resolved *Voyager* insertions with 10 kb flanking regions, labeled by insertion ID. Tracks 1–2 depict genes in the + and – orientation, respectively. Tracks 3–14 depict the percent nucleotide identity of sequence alignments at each locus across 12 *M. phaseolina* genomes. The density plot in Track 15 depicts median Illumina coverage across intact 5S rDNA insertion sites and ranges from 0 to 77 \times . The density plot in Track 16 depicts median Illumina coverage across inserted *Voyager* elements and their flanking regions, and ranges from 0 to 100 \times . (D) Scatterplots (jitter = 0.05) depicting measurements of pathogenic growth on soybean seedlings (mean area under the disease progress curve, AUDPC, top) and saprotrophic growth on PDA media (mean area under the growth progress curve, AUGPC, bottom) as a function of the number of OCH1 homologs found in *Voyager* *Starships*, for each of the 12 isolates. Saprotrophic growth was measured at six different temperatures. Error bars indicate \pm standard error and are visible if they exceed the area of the point. Predicted quadratic equations with shaded 95% confidence intervals are superimposed on the plots. P-values for significant ($P < 0.05$) and marginally significant ($P < 0.10$) linear and quadratic coefficients are shown, and others in [supplementary table S24, Supplementary Material](#) online. Abbreviations: OCH1 (α -1,6-mannosyltransferase); PDA (potato dextrose agar).

elements, especially those found in Actinomycetes which produce coenocytic hyphal networks resembling filamentous Ascomycete fungi. Like many of these bacterial elements, *Starships* are likely mobilized by site-specific tyrosine recombinases located at the 5' end of the element (te Poele et al. 2008; Delavat et al. 2017) and have been implicated in several well-known HGT events (Cheeseman et al. 2014; McDonald et al. 2019; Urquhart et al. 2022). DUF3723s are homologous to DNA-binding bacterial chromosome segregation genes (TIGRFAM accession: TIGR02168), and present a functional analogy to DNA translocases that catalyze bacterial element transfer (te Poele et al. 2008). NLRs with HET, NACHT, and PLP domains govern hyphal fusion events in fungi (Dyrka et al. 2014; Heller et al. 2018) and present parallels with Nudix hydrolase genes that govern Actinomycete hyphal fusion

before integrative and conjugative element transfer (te Poele et al. 2008). Genes with NUDIX domains are also found in several *Starships* (*Galactica*, *Starship-2_Afla*) (supplementary table S13, Supplementary Material online). Together, these shared features suggest *Starships* encode mechanisms not only for their own excision and integration but also for their transfer between individual genomes, and provide testable hypotheses for elucidating a dedicated mechanism of fungal HGT, which has long been searched for.

Concluding Remarks

Starships: New Frontiers for Eukaryotic Genome Evolution

Starships represent a dramatic shift in our understanding of how mobile elements shape eukaryotic genome and

pangenome evolution (Arkhipova and Yushenova 2019). They add to the growing realization that eukaryotic genomes, ranging from fungi to algae to animals, are frequently colonized by large mobile elements, and their activity casts the impact of these other elements in a new light (Inoue et al. 2018; Arkhipova and Yushenova 2019; Moniruzzaman and Aylward 2021). Using the plant pathogen *M. phaseolina* as a model, we demonstrate that *Starships* link the processes of mobile element activity, structural variant generation, and genome content evolution such that *Starship* birth-and-death lifecycles likely hold multi-faceted consequences for their hosts. Our work with *M. phaseolina* further suggests that *Starship* characterization will be important for understanding the emergence of virulence in pathogen populations. *Starships* carrying diverse accessory genes are found in fungi ranging from saprotrophs to lichens to human and plant pathogens and are further known to encode adaptive traits (Cheeseman et al. 2014; McDonald et al. 2019; Urquhart et al. 2022), which coupled with their potential role in mediating life history trade-offs, underscores their vast potential for shaping ecological outcomes across an entire kingdom. The *Starship* activity that we observe in present-day fungal populations represents only a fraction of their historical impact, as they have likely been active for millions of years. Together, our results establish a novel framework for investigating how mobile element cargo carriers shape the (pan)genomic bases of fungal adaptation, pushing our understanding of eukaryotic evolution to bold, new frontiers.

Materials and Methods

Macrophomina phaseolina Isolation

Composite soil samples were collected from six soybean fields in Paraguay (2014–2015) and six in OH, USA (2013–2014) (supplementary fig. S1, Supplementary Material online). A 10-ha area was defined in each field and with a 2.5-cm-diameter cylindrical probe, 20–25 soil cores were collected at a 25 cm depth, and thoroughly mixed. From each composite soil sample, *M. phaseolina* was isolated from a 1-g subsample of air-dried soil as previously described (Mengistu et al. 2009). Colony-forming units morphologically identified as *M. phaseolina* were transferred to a new PDA plate to obtain pure isolates.

Pathogenicity Phenotyping on Soybean Seedlings

A modified soybean cut-stem inoculation technique was used to assess the pathogenicity of *M. phaseolina* isolates (Twizeyimana et al. 2012; Pawlowski et al. 2015). In the greenhouse, soybean plants were grown to the second stage of vegetative growth (Fehr et al. 1971). The stem was cut 25 mm above the second unifoliate node with a surface-sterilized scalpel. A mycelial plug from each *M. phaseolina* isolate was placed on top of the cut stem. Control plants were prepared by placing an uninoculated PDA plug on the cut stem. A 200 μ l pipette tip was used to

cover the stem and mycelial plug to ensure contact of the fungus and the plant and to prevent the plug from drying out. Plants were arranged in a randomized complete block design with six blocks and two plants per experimental unit. Soybean seeds were sown in 48-cell 4 \times 12 plastic trays filled with potting mix (ProMix BX, Premier Horticulture Ltd.) and thinned to one plant per cell post-emergence. Two adjacent plants were considered an experimental unit, and one block was composed of a flat containing 26 plants (used for 12 *M. phaseolina* isolates and one control). A total of six blocks were placed on a greenhouse bench and covered with a plastic tent, to maintain temperature and humidity. Plants were watered as needed, temperature maintained at 27 $^{\circ}$ C \pm 3 $^{\circ}$ C, and a 16-h photoperiod. Three days post-inoculation (dpi), pipette tips were removed, and the length of the necrotic lesion was measured at 3, 6, 9, and 12 dpi. The AUDPC was calculated using the agricolae package in R (version 4.1.0) for each isolate in each block (Mendiburu and Simon 2015; Madden et al. 2017). We replicated the above experiment in 2018 and in 2021. Analysis of variance revealed that the two experiments were not statistically different ($P = 0.532$), therefore, data from both experiments were pooled. An average AUDPC for each isolate was then determined by averaging all values across the 12 blocks from the two experiments.

Saprotrophy Phenotyping on PDA Media

From each *M. phaseolina* isolate, a 6-mm-diameter mycelial plug was placed in the center of a 100-mm-diameter Petri dish containing PDA. Each Petri dish was considered an experimental unit and arranged in a randomized complete block design with three blocks, each corresponding to a different shelf in an incubator. Experimental units were incubated in the dark in the incubator for each of the above temperatures. Colony diameter was measured every 24 h for 5 days and these measurements were used to calculate the AUGPC for each isolate in each block using the agricolae package in R (version 4.1.0) as previously described (Mendiburu and Simon 2015; Sexton et al. 2016). An average AUGPC for each isolate was determined by averaging values across the three blocks. This experiment was repeated using incubators set at 15, 20, 25, 30, 35, and 40 $^{\circ}$ C.

DNA Extraction, Library Prep, and Genome Sequencing

Isolates were grown in potato dextrose broth by shaking cultures in the dark for 3–5 days at 27 $^{\circ}$ C. Mycelia were sieved through autoclaved Miracloth (EMD Millipore, Darmstadt, Germany) and rinsed with sterile distilled water. Mycelia were flash-frozen in liquid nitrogen and stored at -20° C until genomic DNA extraction. Frozen fungal mycelia were ground with a mortar and pestle and genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, GmbH, Germany) following manufacturer's protocol. DNA libraries were prepared for short-

read Illumina Paired End 300 sequencing using the Indexed Truseq DNA Library Preparation Kit (cat.#: FC-121-2001). Four samples were sequenced per Mi-Seq Illumina flow cell with 600 cycles (Illumina, San Diego, CA, USA). DNA libraries for isolates mp007, mp021, mp124, and mp194 were prepared for Oxford Nanopore long-read sequencing using the indexed Nanopore Ligation Sequencing Kit SQK-LS108, whereas all others were prepared using SQK-LS109. Two DNA samples were sequenced per MinION FLO-MIN-106 flow cell with R9.4.1 1D chemistry (Oxford Nanopore Technologies, New York, NY, USA). All library preparation and sequencing were carried out by the Molecular and Cellular and Imaging Center (Wooster, OH, USA).

Read Processing, Genome Assembly, and Chromosome Annotations

Illumina reads were trimmed and quality filtered using Trimmomatic v0.36 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:11 HEADCROP:10 CROP:285 SLIDINGWINDOW:50:25 MINLEN:100). Nanopore reads were basecalled and demultiplexed with Albacore v2.2.4 (default settings; Oxford Nanopore Technologies), and filtered with Porechop v0.2.3 to remove adaptors, to only retain reads with at least one barcode, and to split chimeric reads (default settings; <https://github.com/rrwick/Porechop>). Given that different genome assemblers often include different amounts of sequence in their final output, we merged the nonredundant output of multiple assemblers to obtain the most complete assemblies possible. First, filtered Illumina and Nanopore reads were used to generate hybrid assemblies with SPAdes v3.12 (`-careful, -k 55,77,99,127 -cov-cutoff auto`) (Antipov et al. 2016). Then, unfiltered Illumina and filtered Nanopore reads were used to generate hybrid assemblies with MaSuRCA v3.2.8 (default settings) (Zimin et al. 2013) that underwent ten rounds of polishing by Pilon v1.22 using filtered Illumina reads to correct bases, fix mis-assemblies, and fill gaps (default settings) (Walker et al. 2014). All contigs from the SPAdes assembly were then aligned to the MaSuRCA assembly using nucmer v4.0.0beta (delta-filter options: `-q`) (Marçais et al. 2018), and contigs fully unique to the SPAdes assembly were added to the MaSuRCA assembly to obtain the final assemblies, which were modestly larger than the MaSuRCA assemblies alone (supplementary table S25, Supplementary Material online). We calculated pairwise percent nucleotide identity of alignable sequence between all assemblies as “Length of alignable and identical sequence/Total length of alignable sequence” using nucmer v4.0.0beta (`-maxmatch`).

We classified contigs into provisional core chromosomes using multiple criteria. First, to leverage population-level data that could help in the annotation of telomeres from specific isolates, we mapped all long-read sequences to the ends of each contig >10 kb in each isolate. We kept all reads that aligned to at least 1,000 bp at the contig ends, and considered at least five sequential fungal telomeric repeats (TTAGGG)_n to indicate the presence of a putative telomere. We then

grouped contigs into homologous sets using a protocol inspired by graph-based methods to derive gene ortholog groups (Li et al. 2003). We conducted all possible pairwise genome alignments using the minimap2 aligner in RagTag (Alonge et al. 2019), which outputs a grouping confidence score for all contig pairs calculated as the number of base pairs a query contig covers in its assigned reference contigs divided by the total number of covered base pairs in the entire reference genome. We took the best confidence score >0.5 for each contig pair, and input them into the MCL clustering algorithm (inflation = 1.7) (van Dongen 2000) such that sets of contigs with similar scores were designated as homologous. However, this alignment-based method failed to account for large-scale rearrangements and translocations, such that contigs that were clearly rearranged chromosomes were being grouped together. We, therefore, took a conservative approach for our final classification of contigs into core chromosomes by aligning the two highest-quality genomes that were most distantly related to each other (mp102 and mp053; supplementary fig. S1, Supplementary Material online), and designating all of their shared chromosome-scale contigs as “core” (supplementary fig. S3, Supplementary Material online). We then realigned contigs >1 kb in all genomes to mp102’s core chromosomes using the nucmer aligner in RagTag (`-i 0.2 -a 0.2 -f 1000 -remove-small -nucmer-params “-mum -l 200 -c 500”`), and designated all contigs aligning to these core chromosomes as “core,” and all others as “unplaced.” Four of the assemblies (mp007, mp021, mp124, and mp194) were considerably more fragmented than the other eight, and so large-scale rearrangements could not be confidently assigned for these strains. To generate maps of putative chromosome structure among eight well-assembled isolates, contigs were manually grouped based on nucmer alignments to mp102, or to mp053 for the mp102 strain. Only contigs >500 kb were evaluated. Of note, all strains except mp053 have a break in chromosome 1 at a large track (>50 kb) of satellite sequence (satellite-7 in the repeat library) and are inferred to represent a single chromosome based on the shared presence of satellite-7 at contig ends and alignments to this region in mp053.

The features of these accessory contigs additionally provided new insight into the biology of *M. phaseolina*. Many accessory contigs showed signatures of repeat-induced point mutations (RIP), a fungal genome defense mechanism that introduces cytosine to thymine transitions into repetitive DNA, which was not previously thought to be active in *M. phaseolina* (Islam et al. 2012). Counter to the prevailing reports that *M. phaseolina* reproduces strictly through asexual propagation, the contrasting patterns of RIP across the 12 isolates suggest sexual reproduction has occurred at least once since their last common ancestor, as RIP only occurs immediately before sexual reproduction and meiosis (supplementary table S26, Supplementary Material online). In support of this hypothesis, we also observed two different mating-type loci in isolates from both geographic regions suggestive of an intact bipolar heterothallic reproductive system (supplementary table S27, Supplementary Material online).

De Novo Genome Annotations

All final assemblies were annotated using a pipeline that integrated the output of multiple de novo gene prediction tools (supplementary fig. S14, Supplementary Material online). Briefly, RepeatModeler v2.0.1 (Flynn et al. 2020) using the Repbase database (Bao et al. 2015) was used to generate a custom repeat library that was combined with previously published transcriptome and annotated proteome data from *M. phaseolina* (Islam et al. 2012) to generate initial gene predictions using MAKER2 (Holt and Yandell 2011). High-quality predictions were used to train SNAP (Korf 2004), whose output was then combined with a *M. phaseolina*-specific Augustus parameter file (Stanke et al. 2008) and the output of self-trained GeneMark (Ter-Hovhannisyan et al. 2008) in a second round of MAKER2. High-quality predictions were used to retrain parameter files for SNAP and Augustus, combined with the output of GeneMark, and used in a final round of MAKER2 to generate high-quality de novo gene annotations. To standardize annotations across all isolates, a crucial step for downstream comparative analyses (Weisman et al. 2022), OrthoFiller was used for each genome to annotate any missing genes that were present in at least two other isolates (Dunne and Kelly 2017). All predicted proteins were annotated with Protein family (Pfam) domains and Gene Ontology terms using InterProScan v5.48-83 (Jones et al. 2014). All assemblies were annotated for signatures of RIP using RIPper (default settings) (van Wyk et al. 2019). To generate a custom repeat database for *M. phaseolina*, the output of default runs from each *M. phaseolina* genome was inspected manually to produce a fully classified and annotated library. The manual steps included removing redundancies and putative cellular genes, and ensuring repeats were properly annotated from beginning to end. Due to high divergence among individual repeat families (due to RIP), individual representative sequences were used in lieu of consensus sequences. Classification of most elements was determined using the standard schema provided by Wicker et al. (2007). Noncanonical helitrons were discovered, classified, and annotated according to Chellapan et al. (2016). Initially, repeats containing DUF3435 domains were annotated as *Cryptons*, and entered into the library. Many of these were later reclassified as *Starships* and given new names. Notably, some genomes were found to possess unique repeats with little to no RIP signature, suggesting that the genomes of individual strains may have been invaded by unique transposable elements since the last occurrence of sexual reproduction. This may explain some difference in GC content observed among the strains.

Whole-Genome Alignments and SNP Tree

Whole-genome alignments of the 12 new *M. phaseolina* genomes, a previously published *M. phaseolina* reference (Islam et al. 2012) and an outgroup species *B. dothidea* (Marsberg et al. 2017) were generated with Cactus (Paten et al. 2011) and projected against isolate mp102.

MafFilter (Dutheil et al. 2014) with recommended parameters (Stukenbrock and Dutheil 2018) was used to filter alignments and extract 3,227,097 SNP sites that were then concatenated and used to generate a maximum likelihood phylogeny with IQ-TREE v1.6.9 (-alrt 1000 -bb 1000 -bnni -m MFP + ASC) (Nguyen et al. 2015). Total branch length distance between all pairs of isolates was calculated on the resulting phylogeny as a measurement of phylogenetic distance.

Pangenome Annotations and Comparisons

All pairwise sequence comparisons between predicted proteins in the 12 *M. phaseolina* genomes were determined by diamond blastp (-more-sensitive -e 0.001 -max-hsps 1) (Buchfink et al. 2015), and then used as input to the Pangloss pipeline (default settings) (McCarthy and Fitzpatrick 2019) to determine one-to-one ortholog relationships between predicted genes using a combination of synteny- and sequence-based similarity metrics. In this way, each predicted gene from each isolate was assigned to an ortholog group, and each ortholog group had at most one representative sequence per isolate. Accessory ortholog groups are those missing from at least one isolate, whereas core ortholog groups are those found in all isolates. We calculated pairwise percent identity in genome content between all isolates as the “number of shared ortholog groups/total number of nonredundant ortholog groups present across both genomes.” This process was then repeated for previously annotated genome datasets from six other fungal species in order to provide us with comparison points for evaluating *M. phaseolina*'s pangenome diversity (supplementary tables S5 and S6, Supplementary Material online). Datasets from the other species were chosen, when possible, to be composed of isolates from two distinct geographic locations, similar to the *M. phaseolina* dataset.

Accessory Gene Hotspot Analysis

We used BayesTraits v3.0.1 Multistate MCMC analysis (Pagel and Meade 2006) in conjunction with the whole-genome SNP phylogeny to estimate the number of gains and losses experienced by each accessory ortholog group since the last common ancestor of the 12 *M. phaseolina* isolates. Each analysis was run for 10 million generations with a 2 million generation burn-in, sampling every 4,000 generations until 2,000 samples were collected. Because we lacked a priori expectations of parameter distributions, we seeded the means of all exponential priors from a uniform distribution ranging from 0 to 2 by using a hyperprior, as recommended by the software authors. For each ortholog group, we derived the median probability of it being present or absent at each node in the phylogeny from the 2,000 samples of the posterior probability distribution. We then calculated the probability of gains and losses across the phylogeny by examining transitions between presence and absence states at each pair of parent and child nodes. To calculate the probability of gain of an

ortholog group at a given child node, we multiplied the probability of its absence at the parent node by the probability of its presence at the child node. Similarly, to calculate the probability of loss of an ortholog group at a given child node, we multiplied the probability of its presence at the parent node by the probability of its absence at the child node. For example, if an ortholog group had a $P(\text{absent}) = 0.75$ at a parent node, and a $P(\text{present}) = 0.75$ at one of that parent node's children, we calculated the $P(\text{gain}) = 0.5625$, and $P(\text{loss}) = 0.0625$ at the child node. We summed across all losses, and all gains in the phylogeny, to derive a phylogenetically and probabilistically weighted estimate of the number of gains and losses experienced by each ortholog group. Since the frequent gain and loss of TEs could obscure the gain and loss dynamics of host genes, we removed all ortholog groups where at least one member had either a predicted TE-associated protein domain or overlapped at least 50% with an annotated TE feature.

To identify regions of *M. phaseolina*'s genome where accessory ortholog group gains and losses were greater than expected by chance (i.e., hotspots), we derived a distribution based on the null expectation that ortholog groups are randomly distributed across the genome by sampling with replacement 5 million sets of 15 ortholog groups from randomly selected genomes. We calculated the total number of gains and losses per set, and found that $\leq 1\%$ of sets had a total number of gains and losses ≥ 24.22 . For each genome, we then calculated the total number of gains and losses in each 15 gene window on a one-gene sliding window basis, and designated all windows ≥ 24.22 as hotspots, so that any hotspot had a $\leq 1\%$ empirically derived probability of being observed by chance. We then merged all overlapping hotspots into hotspot regions. We tested whether hotspot sequence is enriched within putative subtelomeric regions. We compared the amount of hotspot and nonhotspot sequence within 100 kb of contig ends to the amounts found outside of these putative subtelomeric regions using a Fisher's exact test, limiting our analysis to the 36 contigs that have at least three putative telomeric repeats on either contig end ([supplementary table S18, Supplementary Material online](#)).

Variant Density

Genotyping structural mutations across a sample of individuals are challenging, especially on a genome-wide basis. We, therefore, developed a statistic analogous to Nei and Li's average pairwise nucleotide diversity (Nei and Li 1979), where we calculated the average length of different indel size classes in a given "focal" genome across all pairwise comparisons with other "query" genomes. We conducted all pairwise alignments of genome assemblies using minimap2 (-x asm10 -c -cs) and called SNP and indel variants using paftools (-l 1000 -L 5000) (Li 2018). Using these alignments and variant calls, we classified all sequences in a focal genome into three types: sequence

with no alignments (i.e., indels), sequence with 1:1 alignments (which included SNPs), and sequence with multiple alignments, for each pairwise comparison. We then overlaid this information with the coordinates of all hotspot regions for the focal genome, and calculated the average length of sequence impacted by SNPs and indels of various size classes (0.1–1, 1–10, >10 kb) per 10 kb of hotspot region. We then repeated this procedure for all background genomic regions. Since we were interested in aligning homologous regions, we only considered alignments to focal regions where $\leq 5\%$ sequence aligned to multiple locations in the query. To generate figures and calculate test statistics, we only considered focal regions that had valid alignments to 11 query genomes. Significant differences between the distributions of the average amount of sequence impacted by SNP and indels in hotspot and background regions were determined using the unpaired two-sample Mann–Whitney *U* test implemented in the wilcox.test function in R.

Starship Search and Classification

We initially discovered the *Voyager* element in *M. phaseolina* through progressively aligning the flanking sequences of large insertions to other genomes using MAFFT v7.407 (Katoh and Standley 2013) in order to identify "empty" insertion sites. We then used all predicted protein sequences within the *Voyager* reference element in *M. phaseolina* isolate mp102 (region ID: mp102|voyager|h11) as well as the SPOK3 sequence from *Enterprise* (Vogan et al. 2021) (GenBank accession: A0A516F180) as queries in a BLASTp search of the 12 *M. phaseolina* proteomes and 1,649 Pezizomycotina, Basidiomycotina, and Saccharomycotina proteomes curated from published databases (BLAST filtering options: min. bit score = 50, max. Evalue = $1e^{-4}$; [supplementary table S12, Supplementary Material online](#); [gitlab.com/xonq/mycotools](https://github.com/xonq/mycotools)). Sequence similarity searches present a more comprehensive approach for mobile element annotation because they do not rely on the presence of polymorphic insertion sites. In *M. phaseolina*, all filtered hits to *Voyager* queries located within 100,000 bp of each other were merged into regions, and all regions with at least four hits were designated as *Starship*-like regions. Full-length *Starship* regions were then characterized by progressively aligning flanking regions, as above. Regions whose insertion sites could not be resolved were bounded by genes at the 5' and 3' terminal ends. We aligned all regions to each other using nucmer and manually adjusted boundaries as necessary to include only sequence homologous to other regions. We repeated the above procedure for the larger genomic database, except we implemented a more conservative neighboring hit distance cutoff of 25,000 bp because in the vast majority of cases, we could not manually verify the region boundaries through pairwise alignments due to insufficient data. Genes within all regions were annotated with protein family (Pfam) domains using Interproscan v5.48-83 (Jones et al. 2014), and were additionally annotated with predicted functions and assigned to COG categories

using eggNOG-mapper (Huerta-Cepas et al. 2017). Due to requiring the presence of four *Starship*-associated hits in a single contig, many *Starships* and *Starship*-like regions were likely unannotated in genomes with poorer contiguity. Likewise, some *Starship*-like regions may represent single loci, which are broken across multiple contigs. After not finding evidence for *Starship*-like regions in either the Basidiomycotina or Saccharomycotina, these taxa were excluded from subsequent analyses that focused on the set of 1274 Pezizomycotina genomes only.

Phylogenetic Classification of *M. phaseolina* *Starships*

All *Starships* and *Starship*-like regions in *M. phaseolina* were classified into distinct element types by comparing the phylogenetic relatedness of their DUF3435 captain sequences, following established approaches for classifying bacterial integrative and conjugative elements (Ghinet et al. 2011). Relationships were additionally supported using phylogenies of DUF3723, FRE, OCH1, and HET and NACHT domains from NLRs, and were used to identify the origin of degraded *Starship* regions. We first retrieved all hits to the DUF3435, DUF3723, FRE, and OCH1 sequences from the *Voyager* reference using a BLASTp search of the 12 *M. phaseolina* proteomes, and aligned all hits using MAFFT v7.407 (–auto) (Katoh and Standley 2013). Due to the functional domain complexity of NLRs, we built individual phylogenetic trees for the HET and NACHT domains found most often in NLR genes (table 2). We used hmmsearch to search the 12 *M. phaseolina* proteomes using profile HMMs (–incE 0.001), and aligned all hits to the query profile HMM using hmalign (Eddy 2011). Alignments were trimmed using ClipKIT (default settings) (Steenwyk et al. 2020), and used as input to generate maximum likelihood phylogenies with SH-like approximate likelihood ratio test (SH-aLRT) and ultra-fast bootstrap approximation (UFboot) support using IQ-TREE v2.0.5, where the best-fit substitution model was selected using ModelFinder (–m MFP –alrt 1000 –B 1000) (Nguyen et al. 2015; Kalyanamoorthy et al. 2017).

Tyrosine Recombinase Classification

An initial alignment of all DUF3435 sequences from the fully annotated *Starships* (supplementary table S14, Supplementary Material online) was generated using MAFFT (defaults). This alignment was used as input to HHPRED (Söding et al. 2005) which identified the tyrosine recombinase domain of CRE recombinase (Protein Data Bank Accession: 1X00) as the best hit to a conserved region of the alignment. A database of repeat elements with known tyrosine recombinase domains was collected from RepBase and HHPRED was used again to define the bounds of the tyrosine recombinase domains for each repeat family. This approach was then used for prokaryotic and plasmid elements from the ICEberg 2.0 database (Liu et al. 2019). For each sequence, all amino acids within the bounds of the tyrosine recombinase homology region were extracted, aligned using MAFFT v7.407 (–auto)

(Katoh and Standley 2013), trimmed and submitted to IQ-TREE as above. The tyrosine recombinase tree strongly supports the monophyly of the DUF3435 sequences in *Starship* elements in relation to other known tyrosine recombinase domains.

Regression Models

We built all phylogenetic generalized least squares regression models and assessed the significance of their parameter estimates using the “gls” function in ape v5.0 in conjunction with the SNP-based isolate tree (correlation = corBrownian, method = “ML”) (Paradis and Schliep 2019: 0). To test for the effect of *Voyager* on *M. phaseolina* growth, we used *Voyager* OCH1 copy number as the linear and quadratic predictor variables and either mean isolate AUDPC or AUGPC as the response variable for the pathogenicity and PDA experiments, respectively (supplementary table S24, Supplementary Material online).

Statistical Analyses and Data Visualization

All statistical analyses were conducted in R v4.1.1 (R Core Team 2021). All Circos plots were generated using Circos (Krzywinski et al. 2009) and all other graphs were visualized using ggplot2 (Wickham et al. 2021: 2). All *Starship* schematics were either drawn manually or visualized using genoplots (Guy et al. 2010). All phylogenetic trees were visualized using ete3 (Huerta-Cepas et al. 2016).

Supplementary Material

(Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Megan McDonald, Hannah Wilson, Alexander Idnurm, and Andrew Urquhart for inspiring discussions on large mobile elements; S. Lorena Ament-Velásquez, for the development of scripts that were invaluable to the outcomes of this research; Paula Moolhuijzen, who supplied several genomes for the pangenome analysis; Thomas Badet, who supplied scripts for genome alignment visualization and provided helpful feedback on the manuscript; and the Canadian Fungal Network for facilitating the coming together of this collaborative research group. This work was supported by a Marie Skłodowska-Curie Individual Fellowship (grant number SEP-210615880-TRADEOFF to E.G.-T.); the National Science Foundation (grant number DEB-1638999 to J.C.S.); the Swedish Research Council Formas (grant number 2019-01227 to A.A.V.); the Swedish Research Council VR (grant number 2021-04290 to A.A.V.); and the Ohio Soybean Council (grant number 17-R-13 to T.L.N. and H.D.L.-N.). We are grateful for the use of published sequence data that were produced in collaboration with the user community by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the

Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Computational work was performed in part using the resources of the Ohio Supercomputer Center.

Author Contributions

E.G.-T., H.D.L.-N., and A.A.V. conceptualized the research. E.G.-T. and A.A.V. conducted formal analyses, with contributions from H.D.L.-N. and T.R. Z.K., C.G.O., and V.D.G. provided materials. H.D.L.-N. secured research grant funding, and J.C.S., C.W.W., A.E.D., and T.L.N. provided additional personnel support. E.G.-T. and A.A.V. wrote the original draft, whereas E.G.-T., A.A.V., J.C.S., C.W.W., and H.D.L.-N. participated in reviewing and editing the manuscript.

Data Availability

This Whole-Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the Biosample accessions SAMN17156263-SAMN17156274. All additional supporting data, including predicted protein sequences, GFFs, repeat libraries, characterized *Starship* elements, sequence alignments, and newick tree files are available from the following Figshare repository (DOI: <https://dx.doi.org/10.6084/m9.figshare.17185880>).

References

- R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing.
- Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* **20**:224.
- Andam CP, Gogarten JP. 2011. Biased gene transfer in microbial evolution. *Nat Rev Microbiol.* **9**:543–555.
- Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* **32**:1009–1015.
- Arkipova IR, Yushenova IA. 2019. Giant transposons in eukaryotes: is bigger better? *Genome Biol Evol.* **11**:906–918.
- Badet T, Fouché S, Hartmann FE, Zala M, Croll D. 2021. Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. *Nat Commun.* **12**:3551.
- Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**:11.
- Baquero F. 2004. From pieces to patterns: evolutionary engineering in bacterial pathogens. *Nat Rev Microbiol.* **2**:510–518.
- Bates S, Hughes HB, Munro CA, Thomas WPH, MacCallum DM, Bertram G, Atrih A, Ferguson MAJ, Brown AJP, Odds FC, et al. 2006. Outer chain N-glycans are required for cell wall integrity and virulence of *Candida albicans*. *J Biol Chem.* **281**:90–98.
- Bellanger X, Payot S, Leblond-Bourget N, Guédon G. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev.* **38**:720–760.
- Benler S, Faure G, Tran H-A, Shmakov S, Zheng F, Koonin E. 2021. Cargo genes of Tn7-like transposons comprise an enormous diversity of defense systems, mobile genetic elements and antibiotic resistance genes. Available from: <https://www.biorxiv.org/content/10.1101/2021.08.23.457393v1>.
- Billane K, Harrison E, Cameron D, Brockhurst MA. 2022. Why do plasmids manipulate the expression of bacterial phenotypes? *Philos Trans R Soc B Biol Sci.* **377**:20200461.
- Blatzer M, Binder U, Haas H. 2011. The metalloredoxase FreB is involved in adaptation of *Aspergillus fumigatus* to iron starvation. *Fungal Genet Biol.* **48**:1027–1033.
- Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. 2019. The ecology and evolution of pangenomes. *Curr Biol.* **29**:R1094–R1103.
- Brown CJ, Sen D, Yano H, Bauer ML, Rogers LM, Van der Auwera GA, Top EM. 2013. Diverse broad-host-range plasmids from freshwater carry few accessory genes. *Appl Environ Microbiol.* **79**:7684–7695.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**:59–60.
- Burkhardt AK, Childs KL, Wang J, Ramon ML, Martin FN. 2019. Assembly, annotation, and comparison of *Macrophomina phaseolina* isolates from strawberry and other hosts. *BMC Genomics* **20**:802.
- Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E, Debuchy R, et al. 2014. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat Commun.* **5**:2876.
- Chellapan BV, van Dam P, Rep M, Cornelissen BJC, Fokkens L. 2016. Non-canonical helitrons in *Fusarium oxysporum*. *Mob DNA.* **7**:27.
- Dallaire A, Manley BF, Wilkens M, Bista I, Quan C, Evangelisti E, Bradshaw CR, Ramakrishna NB, Schornack S, Butter F, et al. 2021. Transcriptional activity and epigenetic regulation of transposable elements in the symbiotic fungus *Rhizophagus irregularis*. *Genome Res.* **31**:2290–2302.
- Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. 2017. The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev.* **41**:512–537.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol.* **2**:414–424.
- Dunne MP, Kelly S. 2017. OrthoFiller: utilising data from multiple species to improve the completeness of genome annotations. *BMC Genomics* **18**:390.
- Dutheil JY, Gaillard S, Stukenbrock EH. 2014. MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC Genomics* **15**:53.
- Dyrka W, Lamacchia M, Durrens P, Kobe B, Daskalov A, Paoletti M, Sherman DJ, Saupe SJ. 2014. Diversity and variability of NOD-like receptors in fungi. *Genome Biol Evol.* **6**:3137–3158.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol.* **7**:e1002195.
- Eickbush DG, Eickbush TH. 1995. Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics* **139**:671–684.
- Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* **26**:1091–1100.
- Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A, et al. 2008. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet.* **4**:e1000046.
- Fehr WR, Caviness CE, Burmood DT, Pennington JS. 1971. Stage of development descriptions for soybeans, *Glycine Max* (L.) Merrill. *Crop Sci.* **11**:929–931.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* **117**:9451–9457.
- Fountain JC, Clevenger JP, Nadon B, Youngblood RC, Korani W, Chang P-K, Starr D, Wang H, Isett B, Johnston HR, et al. 2020.

- Two new *Aspergillus flavus* reference genomes reveal a large insertion potentially contributing to isolate stress tolerance and aflatoxin production. *G3 Genes Genomes Genet.* **10**:3515–3531.
- Ghinat MG, Bordeleau E, Beaudin J, Brzezinski R, Roy S, Burrus V. 2011. Uncovering the prevalence and diversity of integrating conjugative elements in actinobacteria. *PLoS One* **6**:e27846.
- Goodwin TJ, Butler MI, Poulter RTM. 2003. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiol Read Engl.* **149**:3099–3109.
- Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**:2334–2335.
- Habig M, Schotanus K, Hufnagel K, Happel P, Stukenbrock EH. 2021. *Ago1* affects the virulence of the fungal plant pathogen *Zymoseptoria tritici*. *Genes* **12**:1011.
- Heller J, Clavé C, Gladieux P, Saupe SJ, Glass NL. 2018. NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proc Natl Acad Sci U S A.* **115**:E2292–E2301.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**:491.
- Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol.* **34**:2115–2122.
- Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* **33**:1635–1638.
- Inoue Y, Kumagai M, Zhang X, Saga T, Wang D, Koga A, Takeda H. 2018. Fusion of piggyBac-like transposons and herpesviruses occurs frequently in teleosts. *Zool Lett.* **4**:6.
- Inoue Y, Saga T, Aikawa T, Kumagai M, Shimada A, Kawaguchi Y, Naruse K, Morishita S, Koga A, Takeda H. 2017. Complete fusion of a transposon and herpesvirus created the Teratorn mobile element in medaka fish. *Nat Commun.* **8**:551.
- Islam MS, Haque MS, Islam MM, Emdad EM, Halim A, Hossen QMM, Hossain MZ, Ahmed B, Rahman S, Rahman MS, et al. 2012. Tools to kill: genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC Genomics* **13**:493.
- Johnson CM, Grossman AD. 2015. Integrative and Conjugative Elements (ICEs): what they do and how they work. *Annu Rev Genet.* **49**:577–601.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.
- Katoh K, Standley DM. 2013. MAFFT multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol Biol Evol.* **30**:772–780.
- Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, Harris D, Clarke L, Whitehead S, Sangal V, et al. 2009. Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res.* **19**:2279–2287.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinform.* **5**:59.
- Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**:1639–1645.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**:2178–2189.
- Li M-H, Xie X-L, Lin X-F, Shi J-X, Ding Z-J, Ling J-F, Xi P-G, Zhou J-N, Leng Y, Zhong S, et al. 2014. Functional characterization of the gene FoOCH1 encoding a putative α -1,6-mannosyltransferase in *Fusarium oxysporum* f. sp. cubense. *Fungal Genet Biol.* **65**:1–13.
- Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou H-Y. 2019. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **47**:D660–D665.
- Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**:367–373.
- Madden LV, Hughes G, van den Bosch F. 2017. *The study of plant disease epidemics*. The American Phytopathological Society. Available from: <https://apsjournals.apsnet.org/doi/book/10.1094/9780890545058>.
- Maistrenko OM, Mende DR, Luetge M, Hildebrand F, Schmidt TSB, Li SS, Rodrigues JFM, von Mering C, Pedro Coelho L, Huerta-Cepas J, et al. 2020. Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* **14**:1247–1259.
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* **14**:e1005944.
- Marsberg A, Kemler M, Jami F, Nagel JH, Postma-Smidt A, Naidoo S, Wingfield MJ, Crous PW, Spatafora JW, Hesse CN, et al. 2017. *Botryosphaeria dothidea*: a latent pathogen of global importance to woody plant health. *Mol Plant Pathol.* **18**:477–488.
- McCarthy CGP, Fitzpatrick DA. 2019. Pangloss: a tool for pan-genome analysis of microbial eukaryotes. *Genes* **10**:521.
- 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**:e01515–e01519.
- Melnik RA, Hossain SS, Haney CH. 2019. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated *Pseudomonas*. *ISME J.* **13**:1575–1588.
- Mendiburu FD, Simon R. 2015. Agricolae – ten years of an open source statistical tool for experiments in breeding, agriculture and biology. *PeerJ Inc.* **3**:e1404v1. doi:10.7287/peerj.preprints.1404v1.
- Mengistu A, Reddy KN, Zablutowicz RM, Wrather AJ. 2009. Propagule densities of *Macrophomina phaseolina* in soybean tissue and soil as affected by tillage, cover crop, and herbicide. *Plant Health Prog.* **10**:28.
- Moniruzzaman M, Aylward F. 2021. Endogenous giant viruses shape intraspecies genomic variability in the model green alga *Chlamydomonas reinhardtii*. Available from: <https://www.biorxiv.org/content/10.1101/2021.11.30.470594v1>.
- Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* **76**:5269–5273.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* **32**:268–274.
- Pagel M, Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am Nat.* **167**:808–825.
- Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528.
- Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. 2011. Cactus: algorithms for genome multiple sequence alignment. *Genome Res.* **21**:1512–1528.
- Pawlowski ML, Hill CB, Hartman GL. 2015. Resistance to charcoal rot identified in ancestral soybean germplasm. *Crop Sci.* **55**:1230–1235.
- Piednoël M, Gonçalves IR, Higuete D, Bonnard E. 2011. Eukaryote DIRS1-like retrotransposons: an overview. *BMC Genomics* **12**:621.
- Rehman L, Su X, Li X, Qi X, Guo H, Cheng H. 2018. FreB is involved in the ferric metabolism and multiple pathogenicity-related traits of *Verticillium dahliae*. *Curr Genet.* **64**:645–659.

- Rooney AP, Ward TJ. 2005. Evolution of a large ribosomal RNA multi-gene family in filamentous fungi: birth and death of a concerted evolution paradigm. *Proc Natl Acad Sci U S A*. **102**:5084–5089.
- Saikia S, Oliveira D, Hu G, Kronstad J. 2014. Role of ferric reductases in iron acquisition and virulence in the fungal pathogen *Cryptococcus neoformans*. *Infect Immun*. **82**:839–850.
- Sexton ZF, Hughes TJ, Wise KA. 2016. Analyzing isolate variability of *Macrophomina phaseolina* from a regional perspective. *Crop Prot*. **81**:9–13.
- Shen X-X, Steenwyk JL, LaBella AL, Oplente DA, Zhou X, Kominek J, Li Y, Groenewald M, Hittinger CT, Rokas A. 2020. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci Adv*. **6**:eabd0079.
- Singh A, Severance S, Kaur N, Wiltsie W, Kosman DJ. 2006. Assembly, activation, and trafficking of the Fet3p.Ftr1p high affinity iron permease complex in *Saccharomyces cerevisiae*. *J Biol Chem*. **281**:13355–13364.
- Smyshlyaev G, Bateman A, Barabas O. 2021. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol Syst Biol*. **17**:e9880.
- Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. **33**:W244–W248.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntetically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. **24**:637–644.
- Steenwyk JL, Iii TJB, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol*. **18**:e3001007.
- Stukenbrock EH, Croll D. 2014. The evolving fungal genome. *Fungal Biol Rev*. **28**:1–12.
- Stukenbrock EH, Duthel JY. 2018. Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics*. **208**:1209–1229.
- Sullivan JT, Ronson CW. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A*. **95**:5145–5149.
- Templeton AR, Hollocher H, Johnston JS. 1993. The molecular through ecological genetics of abnormal abdomen in *Drosophila mercatorum*. V. Female phenotypic expression on natural genetic backgrounds and in natural environments. *Genetics*. **134**:475–485.
- te Poele EM, Bolhuis H, Dijkhuizen L. 2008. Actinomycete integrative and conjugative elements. *Antonie Van Leeuwenhoek*. **94**:127–143.
- Ter-Hovhannisyantsyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. **18**:1979–1990.
- Twizeyimana M, Hill CB, Pawlowski M, Paul C, Hartman GL. 2012. A cut-stem inoculation technique to evaluate soybean for resistance to *Macrophomina phaseolina*. *Plant Dis*. **96**:1210–1215.
- Urquhart AS, Chong NF, Yang Y, Idnurm A. 2022. A large transposable element mediates metal resistance in the fungus *Paecilomyces variotii*. *Curr Biol*. **32**:937–950.e5.
- van Dongen S. 2000. A cluster algorithm for graphs. Available from: <https://ir.cwi.nl/pub/4463>.
- van Wyk S, Harrison CH, Wingfield BD, De Vos L, van der Merwe NA, Steenkamp ET. 2019. The RIPper, a web-based tool for genome-wide quantification of repeat-induced point (RIP) mutations. *PeerJ*. **7**:e7447.
- Verma J, Bag S, Saha B, Kumar P, Ghosh TS, Dayal M, Senapati T, Mehra S, Dey P, Desigamani A, et al. 2019. Genomic plasticity associated with antimicrobial resistance in *Vibrio cholerae*. *Proc Natl Acad Sci U S A*. **116**:6226–6231.
- 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**:789–798.
- Vogan AA, Ament-Velásquez SL, Granger-Farbos A, Svedberg J, Bastiaans E, Debets AJ, Coustou V, Yvanne H, Clavé C, Saupe SJ, et al. 2019. Combinations of *Spok* genes create multiple meiotic drivers in *Podospora*. *eLife*. **8**:e46454.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. **9**:e112963.
- Wang C, Milgate AW, Solomon PS, McDonald MC. 2021. The identification of a transposon affecting the asexual reproduction of the wheat pathogen *Zymoseptoria tritici*. *Mol Plant Pathol*. **22**:800–816.
- Weider LJ, Elser JJ, Crease TJ, Mateos M, Cotner JB, Markow TA. 2005. The functional significance of ribosomal (r)DNA variation: impacts on the evolutionary ecology of organisms. *Annu Rev Ecol Evol Syst*. **36**:219–242.
- Weisman CM, Murray AW, Eddy SR. 2022. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. 2022.01.13.476251. Available from: <https://www.biorxiv.org/content/10.1101/2022.01.13.476251v1>.
- 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**:973–982.
- Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D. 2021. ggplot2: create elegant data visualisations using the grammar of graphics. Available from: <https://CRAN.R-project.org/package=ggplot2>.
- Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL. 2015. Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. *BMC Genet*. **16**:96.
- Yang H, Yu H, Ma L-J. 2020. Accessory chromosomes in *Fusarium oxysporum*. *Phytopathology*. **110**:1488–1496.
- Zhang J, Zhang Y, Yang J, Kang L, EloRM AM, Zhou H, Zhao J. 2019. The α -1,6-mannosyltransferase VdOCH1 plays a major role in microsclerotium formation and virulence in the soil-borne pathogen *Verticillium dahliae*. *Fungal Biol*. **123**:539–546.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics*. **29**:2669–2677.