Orphan genes are clustered with allorecognition loci and may be involved in incompatibility and speciation in *Neurospora*

Zheng Wang[1], Yaning Wang[1,2], Takao Kasuga[3], Francesc Lopez-Giraldez[4], Yang Zhang[5], Zhang Zhang[5], Rudy Diaz[1], Caihong Dong[2], Anita Sil[6], Frances Trail[7], Oded Yarden[8], and Jeffrey P. Townsend[1, 9*]

[1]*Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut 06511, USA*

[2]*Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China*

[3]*College of Biological Sciences, University of California, Davis, CA95616, USA*

[4]*Yale Center for Genomic Analysis, New Haven, Connecticut 06511, USA*

[5]*National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China*

[6]*Department of Microbiology and Immunology, University of California, San Francisco, CA94115, USA*

[7]*Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA*

[8]*Department of Plant Pathology and Microbiology, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel*

[9]*Department of Ecology and Evolutionary Biology, Program in Microbiology, and Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06511, USA*

* Correspondence: jeffrey.townsend@yale.edu;

Abstract:

As *de novo* elements of the genome, orphan genes have long been postulated to play roles in the establishment of genetic barriers to intercrossing and speciation. However, there is a lack of working hypotheses as to what role they play. Systematic investigation of the evolutionary history, genome structures, expression dynamics and regulation of orphan genes in well-studied models can shed light on their functions in reproductive isolation. The 670 orphan genes that we identified in the genome of the fungal model *Neurospora crassa* were aggregated adjacent to the telomeres, and over 63% of them formed clusters with 61% of the *het*-like genes—genes that regulate self-recognition and define vegetative compatibility groups. Nearly all orphan-*het* clusters are syntenic within closely related species, suggesting relatively recent colocalization. Analysis of transcriptomic data from *N. crassa* conditions associated with growth and reproduction reveals 342 orphan genes that are dynamically expressed during both asexual and sexual phases. Among these, 37% were expressed in a detectable manner when the fungus was cultured on common carbon resources, but 64% were detectable on unusual carbon sources such as furfural and HMF—wildfire-produced chemicals that are a strong inducer of sexual development in *N. crassa*. Expression of a significant portion of the orphan genes was sensitive to light and temperature, factors that regulate asexual and sexual reproduction, among other fungal activities. Coordinate expression in orphan-*het* gene clusters was detected during early hyphal branching. Furthermore, orphan genes and clustered *het*-like genes respond similarly to mutations in transcription factors *adv-1* and *pp-1* that regulate hyphal communication, and expression of more than one quarter of the orphan genes was affected by a mating locus mutant. Functional interactions between orphan and *het*-like genes likely consolidate vegetative

incompatibility during asexual reproduction, but possibly promote crossings between vegetative compatibility groups during sexual reproduction. Their involvement in the balance between genome homogeneity and heterogeneity in vegetative compatibility likely contribute to *Neurospora* speciation. Orphan genes are potential targets for the manipulation of fungal growth, a key aspect of both the control of fungal pathogenicity and the iterative improvement of fungal biotechnology.

Introduction

Since the emergence of life, molecular evolution has contributed to the accumulation of novel and diverse features in all kinds of organisms. Two fundamental components of that molecular evolutionary novelty are new genes and novel gene functions, which have long been considered to be emergent properties of gene duplication and rearrangement. Nevertheless, genomes often harbor numerous orphan genes—novel genes that have no homologues in distantly- or closely-related lineages and that cannot be tracked to ancestral lineages. These orphan genes manifest in large numbers across a diversity of organisms, such that they represent nearly one-third of the genes in all genomes, including phages, archaea, bacteria, and eukaryotic organisms (Tautz and Domazet-Lošo 2011). Three key challenges have thus been the focus of studies of orphan genes: how to identify de novo genes, how to track their evolutionary histories, and perhaps most importantly, how these genes are integrated into pre-existing gene interaction networks (McLysaght and Hurst 2016).

Accurate identification of *de novo* genes can, at times, be difficult. They originate via three evolutionary processes: (1) rapid gene evolution, which is refractory to tracking homology on the basis of sequence conservation; (2) intragenomic gene loss and gain or horizontal gene transfer, which can convey higher fitness in response to genetic/environmental changes; and (3) accumulated mutations that establish novel function, evolving slowly but long enough on an independent lineage that the gene phylogeny cannot be tracked back to its distant ancestor or to related lineages (Su and Townsend 2015; Dornburg et al. 2019).

Studies of the genomic characteristics of de novo elements in several model organisms have revealed the likely origins of orphan genes from gene duplication, non-coding sequences, as

well as fast-evolution at conserved genomic positions. One example supporting the frequency of origin by rapid divergence after gene duplication and rearrangement can be found in yeast, where the presence of 55% to 73% percent of the orphan genes can be explained by sufficient divergence from sister species (Weisman et al. 2020). Some *de novo* protein-coding genes might have directly evolved from non-coding regions in the genome (Ruiz-Orera et al. 2015), as has been described in the tests of fruit flies (Begun et al. 2006; Begun et al. 2007). One hundred seventy five *de novo* genes in Asian rice corresponded to recognizable non-genic sequences in closely related species (Zhang et al. 2019). These investigations using model organisms confirmed unique characteristics of these *de novo* genes, making good frameworks for investigating orphan genes in other species. However, linkages between these revealed genomic characteristics and the integrative functions of orphan genes remain unclear. Therefore, systematic approaches that combine study of comparative genomics with functional assays of gene expression and gene-perturbation phenotypes using well-established model systems are critical to integrate the investigation of the orphan gene's origination and function.

The most frequently used approach to identify orphan genes is using phylostratigraphy (Cai et al. 2006; Domazet-Loso et al. 2007)Precise identification of the origins of *de novo* genes using a phylostratigraphic approach is critically dependent on accurate gene annotation and extensive comparison among proper representative genomes (Casola 2018). It is difficult to distinguish whether genes with no homologues in closely related lineages are true orphan genes as opposed to lacking homologues in closely related lineages that have few genomes available. An alternative to phylostratigraphy is AKA gene synteny, which compares each gene’s position

relative to its neighbors. A recent study suggested that if the neighbors of a gene are in a conserved order in other species, then the gene is likely to correspond to whatever is at the orthologous position in the other species as well—even if the sequences do not match (Vakirlis et al. 2020).

With detailed pangenomic analyses, identification of orphan genes and *de novo* protein-coding genes need to be further verified with a systematic approach focusing on possible functional novelty and genetic signals that may be associated with such a novelty (McLysaght and Hurst 2016). Systematic assessment of the putative orphan and *de novo* gene function can track their behaviors during the growth and development and verify their possible roles by examination of corresponding knockdown or knockout phenotypes (McLysaght and Hurst 2016). As orphan genes appear along specific lineages, they are naturally thought to be important to species- or genus-level adaptations of development to taxon-specific ecology.

The well-annotated model species in the genus *Neurospora*, *N. crassa, N. tetrasperma* and *N. discreta* of the class Sordariomycetes, provide a set of three closely-related genomes enabling investigation of possible genetic novelties associated with recent and rapid ecological divergences (Gladieux et al. 2020), such as responses to nutrients and other environmental factors and developmental divergences such as heterothallic, pseudohomothallic and homothallic outcrossing during sexual reproduction as well as incompatibility during vegetative growth. *Neurospora* species are highly adapted to the postfire environment, capable of fast asexual growth and reproduction on simple nutrients and have long been genetic models for eukaryotic metabolic regulation and for mating, meiosis and morphological development during reproduction (Mitchell 1965; Ebbole 2000; Davis and Perkins 2002). In fact, fungi in the

Sordariomycetes exhibit remarkable ecological, biological, and morphological diversity (Nowrousian et al. 2004; Zhang and Wang 2015; Zámocký et al. 2016). These fungi also exhibit diverse reproductive modes, including heterothallism, pseudohomothallism, and homothallism with or without asexual reproduction (Ebbole 2000; Lehr et al. 2014; Wang et al. 2014; Corcoran et al. 2016; Wang, Miguel-Rojas, et al. 2019; Wang, López-Giráldez, et al. 2019). Such diversities in ecology and development that have evolved in parallel or convergently within closely related species in a single fungal class, may provide avenues toward an understanding of the G × E impacts of de novo elements on the evolutionary process. Due to the repeat-induced point mutation (RIP) genome-defense system, *Neurospora* is known to lack recent gene duplications—a major source of evolutionary novelty (Galagan and Selker 2004; Gladyshev 2017). Nevertheless, comparing representative genomes in prokaryotes, plants & animals, Basidiomycota, major lineages of Ascomycota, and *Chaetomium globosum,* which is closely related to *N. crassa*, 2219 orphan genes were identified in *N. crassa* by phylostratigraphy (Kasuga et al. 2009). In the past decade, many more fungal genomes have been sequenced. Their sequence, along with the advances in genome sequencing and annotation techniques (Grigoriev 2011; Grigoriev et al. 2014; Haridas et al. 2018), provide a more inclusive comparison for identifying *de novo* genes. Therefore, to understand how important the roles that orphan genes play in genome-wide regulation during the whole life history, we investigated signals of recent selection pressures in orphan genes, focusing on genes involved in environmental responses and fertility or mating in *Neurospora*.

Several orphan genes are allorecognition loci, often referred to as *het* (heterokaryon incompatibility) or *vic* (vegetative incompatibility) genes. As their names indicate, *het* or *vic*

genes regulate allorecognition during vegetative growth, and only individuals with compatibility at all of their *het*-loci can fuse and simply expand their colonies (Zhao et al. 2015). Some *het*-genes have pleiotropic effects in sexual development in some fungal species and play direct roles in reproductive isolation and speciation within sympatry (Ament-Velásquez et al. 2022). *Het*-like genes can be predicted with conserved HET domain in protein sequences, and there were 69 *het*-like genes identified in the genome strain of *N. crassa* (Zhao et al. 2015). Characters such as copy number, frequencies of gain and loss, and evolutionarily itinerant chromosomal location are shared traits of *het*-like genes and orphan genes, suggesting the merits of integrative investigations of both gene groups. Our findings provide insight into recent molecular evolutionary selection on *de novo* genes. In addition, transcriptomic profiles during the life cycle of *N. crassa* and its closely related species combined with knockout phenotypes provide insights on how *de novo* genes are integrated into pre-existing regulatory networks and become essential in incompatibility and speciation in the *N. crassa* model.

Results

***670 Orphan genes identified in Neurospora genomes*—**A total of 1872 *Neurospora* orphan genes, using *N. crassa* as reference, were identified using genomic phylostratigraphy on representative taxa for major fungal lineages and a few non-fungal reference genomes (**Fig. 1, Table S1**). Within these 1872 *Neurospora* orphan genes, a total of 695 genes are shared between the genomes of *Neurospora* and the sole species *S. macrospora* in the sister genus (**Table S1**). Further reciprocal-BLAST searches for the 1871 genes against genomes of species within the genus including *N. crassa, N. tetrasperma* and *N. discreta*, identified 670 genes that are

Neurospora-specific orphans (**Table S1**). Among the 670 *Neurospora* orphans identified in the *N. crassa* genome, 241 are *N. crassa* unique and 405 are shared between *N. crassa* and *N. tetrasperma* with 248 showing no orthologs in *N. discreta*. 181 *Neurospora* orphan genes are shared between *N. crassa* and *N. discreta* with 26 showing no orthologs in *N. tetrasperma*. Because of the special status of *N. crassa* as a model species, here we specifically investigate and report as orphan genes those that are specific to *N. crassa* in this study. With more genomes in this fungal class being sequenced and annotated and more non-classified genes in the phylostratigraphy being analyzed, these numbers are expected to be slightly changed. Of the 670 orphan genes, 515 have at least one intron (average: ~2 introns, maximum: 8 introns in NCU07480). These orphan genes encoded proteins ranging from 26 to 1310 amino acids (NCU05561 and NCU04852, respectively). The average orphan gene length (~192 amino acids) was significantly shorter than the average length of the non-orphan gene (~528 amino acids).

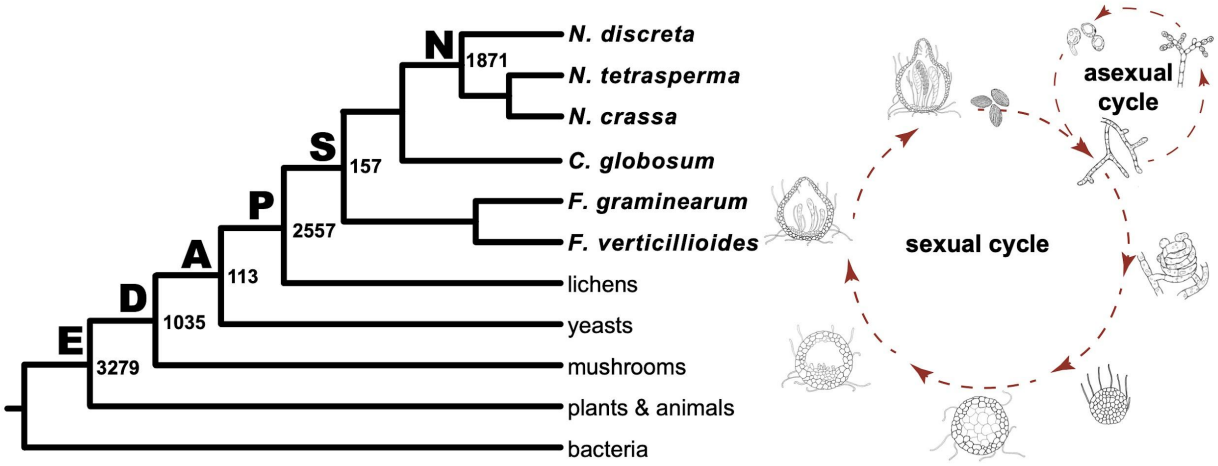

Figure 1. Systematic study of *Neurospora* orphan genes and their roles in fungal growth and development. (**A**) Genomic phylostratigraphy of genes (enumerated at ancestral nodes) in the *N. crassa* genome; (**B**) Life history of *N. crassa*. These developmental processes have been transcriptionally profiled (dashed red arrows; Wang, Gudibanda, et al. 2018).

Orphan genes are aggregated in the telomere regions and clustered with het genes—Orphan genes are distributed in all seven chromosomes of the *N. crassa* genome, with paralogs from duplicates often clustered together (**Fig. 2**). Window-free maximum-likelihood model averaging of the gene-regionalized probability of *Neurospora* orphan genes and *het*-like genes revealed that orphan genes were clustered, with significant ($P < 0.05$) clustering on chromosomes I, II, III, IV, and V. Large orphan-gene clusters are typically aggregated toward the telomeres of each chromosome, especially in chromosomes I, III, IV, V , VI and VII (**Fig. 2 A–G, S2**). Detailed clustering revealed with the Cluster `Locator` (Pazos Obregón et al. 2018) tallied 67% of orphan genes as present in clusters with a `max-gap` of five (48% with a `max-gap` of one). About 30% of orphan genes were in clusters with more than four genes (**Tables 1 & S3**), including six clusters hosting 9, 10, 16, and 23 genes.

A total of 69 genes with HET domains (*het*-like genes) were identified and mapped to the *N. crassa* FGSC2489 strain (Figure 1 in Zhao et al. 2015). Many of these *het*-like genes exhibited nonrandom distributions and were clustered near the end of the linkage groups, largely overlapping with clusters of orphan genes (**Fig. 2**). Additional analysis included NCU03125 (het-C; Saupe et al. 1996). 42 of the 68 *het*-like genes were clustered with at least one orphan gene within a range of five genes (`max-gap = 5`), and 23 and 14 *het*-like genes were clustered with at least one orphan gene within a range of three (`max-gap = 1`) or two genes (`max-gap = 0`) separately (**Fig. 2A–G, Table 1, S4**). In fact, many of *het*-like genes were only syntenic within *Neurospora* and very closely related species in the Sordariales, and most *het*-like genes were embraced by *Neurospora* orphan genes and comparatively "young" genes (**Fig. 2H, Table S5**).

Table 1. Orphan and *het*-like genes are present in clusters in the *N. crassa* genome.

Cluster size	Orphan genes[a]			Orphan and *het*-like genes[b]		
	# Clusters	# Genes	% Genes	# Clusters	# Genes	% Genes
2	57	114	17.1%	63 (15 *het*)	126	17.1%
3	28	84	12.6%	28 (7 *het*)	84	11.4%
4	8	32	4.8%	12 (6 *het*)	48	6.5%
5	14	70	10.5%	13 (3 *het*)	65	8.8%
6	5	30	4.5%	11 (7 *het*)	66	8.9%
7	3	21	3.2%	2	14	1.9%
8	2	16	2.4%	2	16	2.2%
9	3	27	4.1%	1	9	1.2%
10–11	1	11	1.7%	2 (2 *het*)	21	2.9%
16–18	1	16	2.4%	2 (2 *het*)	35	4.8%
23	1	23	3.5%	1	23	3.1%
Total	123	444	66.8%	129 (42 *het*)	474 (44 *het*)	64.4%

[a] 670 genes; [b] 736 genes.

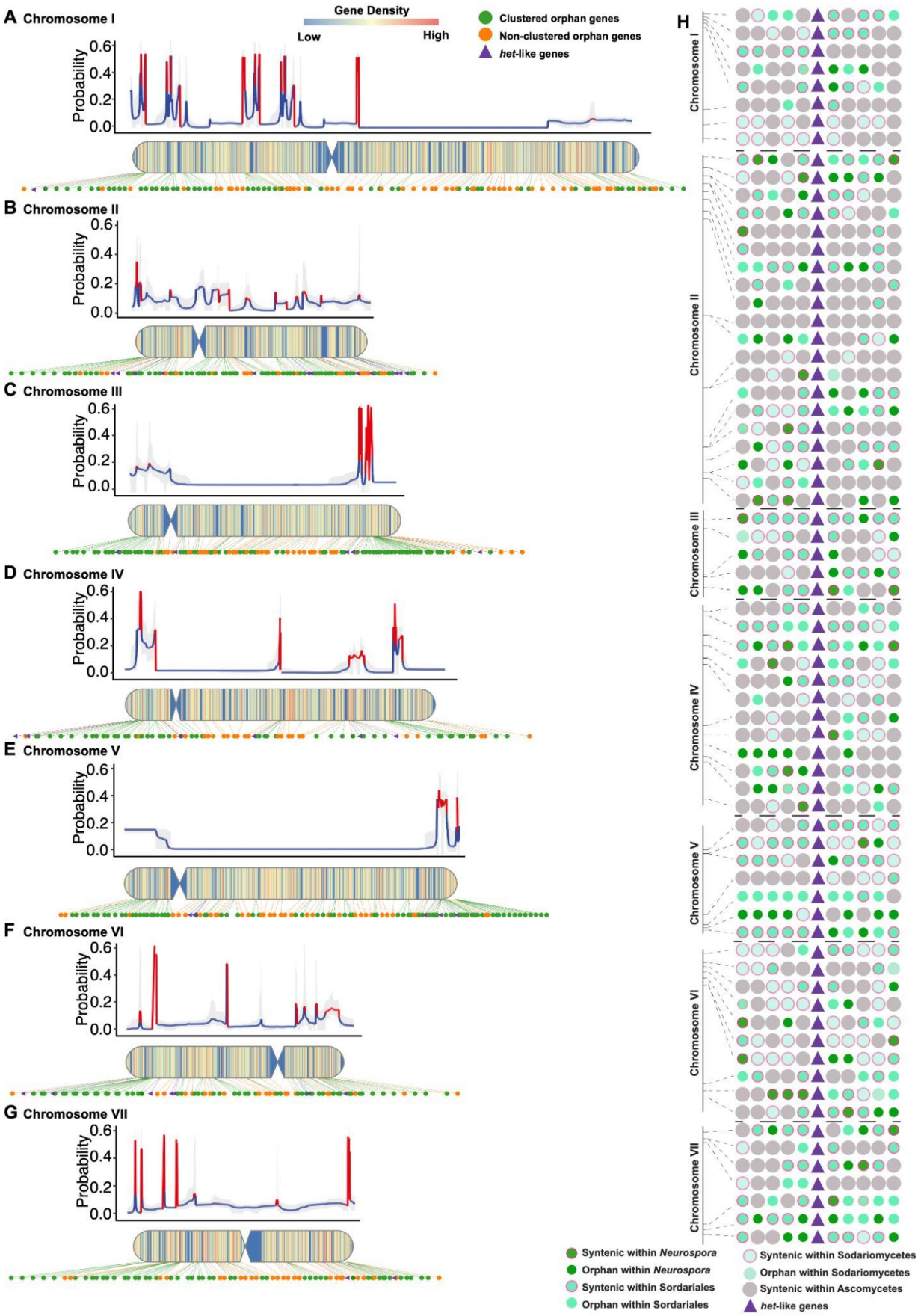
A Chromosome I
B Chromosome II
C Chromosome III
D Chromosome IV
E Chromosome V
F Chromosome VI
G Chromosome VII
H
Probability
0.0
0.2
0.4
0.6
Gene Density
Low
High
Clustered orphan genes
Non-clustered orphan genes
het-like genes
Syntenic within Neurospora
Orphan within Neurospora
Syntenic within Sordariales
Orphan within Sordariales
Syntenic within Sodariomycetes
Orphan within Sodariomycetes
Syntenic within Ascomycetes
het-like genes

Figure 2. Regionalized probability of the orphan status of a gene (grey band: 95% model uncertainty interval) inferred using window-free maximum-likelihood model averaging spatial distribution of *Neurospora* orphan genes and *het*-like genes in orphan gene clusters, chromosomal gene-density heat maps over windows of 10000 base pairs, and distributions of orphan genes (clustered: green [$P < 0.01$], and non-clustered [$P \geq 0.01$]: orange) and *het*-like genes (purple triangles) across *N. crassa* (**A**) chromosome I, (**B**) chromosome II, (**C**) chromosome III, (**D**) chromosome IV, (**E**) chromosome V, (**F**) chromosome VI, and (**G**) chromosome VII. (**H**) At four taxonomic depths, phylogenetic synteny (larger grey circles for deeper synteny, i.e. Ascomycetes > Sordariomycetes > Sordariales > Neurospora) and orphan status (greener hue for orphans found only in lesser taxonomic depths, e.g. Neurospora > Sordariales > Sordariomycetes > Ascomycetes) of five neighbor genes on either side of 69 *het*-like genes (**Table S5**) predicted in the original genome-sequenced *N. crassa* strain (FGSC2489; Galagan et al. 2003).

Expression dynamics of orphan and het-like genes in response to developmental and environmental changes—Genome-wide gene expression was measured in key stages of the *N. crassa* life cycle (**Figs. 1**, **S1, S2**; **Table S6**). During sexual development on SCM and asexual growth on BM, similar trends were observed for proportions of orphan genes and non-orphan genes that exhibited measurable expression, but increased proportion of orphan genes were expressed during the early hyphal branching on MSM (**Fig. S1**). At one or more time points during perithecial development, 238 exhibited at least 5-fold ($P < 0.05$) expression changes during perithecial development, indicating potential roles in the regulation of sexual reproduction. 35 genes exhibited no measurable expression in all sampled life history stages, suggesting either function only in unusual circumstances or mis-annotation as expressed genes (**Fig. S2, Table S6**). During conidial germination and early asexual growth, respectively, 95 and 137 orphan genes exhibited at least 5-fold ($P < 0.05$) expression changes for cultures on Bird medium (BM) and maple sap medium (MSM), and 213 and 181 orphan genes were expressed during none or only one of the four sampled stages for cultures on BM or MS separately. Expression of 342 orphan genes was detected during sexual and during asexual growth on

asexual specific BM and MSM (**Fig. 3**). Transcriptomics were also profiled for *N. crassa* growing on seven different carbon conditions including, absence of carbon, solo glucose carbon source, and a complex carbon source with five commonly available crop residues, including barley, corn, rice, soybean, and wheat straws, found in the field (Wang et al. 2015). Expression of 464 orphan genes was too low to be detected under any of these conditions; this large portion of *Neurospora* orphan genes are not required for vegetative growth associated with carbon metabolism regulation (**Tables S7, S8**). Expression of 22 orphan genes required that at least one type of carbon resource was present in the media, while expression of 56 other orphan genes was only detected in the absence of carbon. Revisiting expression data collected from experiments investigating *N. crassa*'s tolerance to furfural (Feldman et al. 2019) identified that 245, 239, 257, and 232 orphan genes exhibited measurable levels of expression in the wild-type condition, DMSO (the carbon blank control), furfural, and HMF treatments (**Tables S7, S8**). Within the 291 genes expressed in either furfural or HMF cultures or both, 61 were not expressed in the wild-type condition. Furfural is derived from lignocellulosic biomass and enriched in a post-fire environment. *N. crassa* sexual spore germination can be induced by furfural presence (Emerson 1948; Eilers and Sussman 1970). Furfural also inhibits conidia germination. Compared with cultures under wild-type conditions, 12 orphan genes exhibited a 3-fold or higher expression in response to furfural, with NCU09604, 07323, and 01153 showing 6- to 22-fold up-regulation in furfural cultures.

Environmental factors, including light and temperature that regulate fungal growth and development, also dramatically affect expression of orphan genes. *N. crassa* genome is equipped with a set of light sensors responding to different light spectrums, duration and intensities

(Kritsky et al. 2005; Káldi et al. 2006; Chen et al. 2009; Wu et al. 2014; Wang et al. 2016; Wang, Wang, et al. 2018), and orphan genes exhibited sensitive responses to light conditions. When *N. crassa* cultures were exposed to light up to 4 h, genes were classified as short light responsive genes or long light responsive genes based on their expression profile changes (Wu et al. 2014). Revisiting the previous data disclosed that out of 488 genes induced by light stimulus, 106 were orphan genes (significantly enriched, $P < 0.01$), 59 of which were in the predicted clusters, including all genes in three 2-orphan clusters, including cluster #25, 61, and 81 as well as three genes in a 4-gene cluster #127. Among 49 genes whose expression halted upon exposure to light, six were orphan genes, including two genes NCU05052 and 05058 in a 3-gene cluster (**Fig. 4A, Table S8**). During conidial germination at a high temperature of 37 C on BM, expression of 270 orphan genes was completely inhibited, and this was significantly enriched ($P < 0.01$) for orphan genes as only a total of 941 out of 10592 genes exhibited non-detectable expression. In detail, there were 148 orphan genes exhibiting non-detectable expression under both at 25 C and 37 C conditions. There were 152 orphan genes exhibiting detectable expression in cultures at 25 C but being turned off at 37 C, 100 of which were clustered orphan genes, including 24 clusters with more than 2 genes being turned off at 37 C. In contrast, 13 exhibited detectable expression in cultures at 37 C, but not at 25 C (**Fig. 4B, Table S8**). The orphan cluster #25 of NCU02144–02145 was not expressed in dark conditions or at 37 C.

Expression of *het*-like genes as a whole exhibited no clear patterns in response to environmental conditions or developmental stages **(Fig. 4, Tables S6, S8**). However, 11 *het*-like genes were not expressed in cultures on furfural or HMF, and 28 *het*-like genes were expressed neither in the dark nor during a shift from the dark to light for a duration of up to 2 h. Both of

these conditions are known to promote sexual reproduction in *N. crassa*. More genes (5968 vs. 2935) were significantly up-regulated ($P < 0.05$) during the first branching of the germ tube on MSM, which supports both asexual and sexual development, than those on BM, which is specifically designed for promoting asexual reproduction and inhibiting sexual development. Accordingly, many more *het*-like genes (46 vs. 15) were significantly up-regulated ($P < 0.05$) on MSM than on BM during that first branching stage.

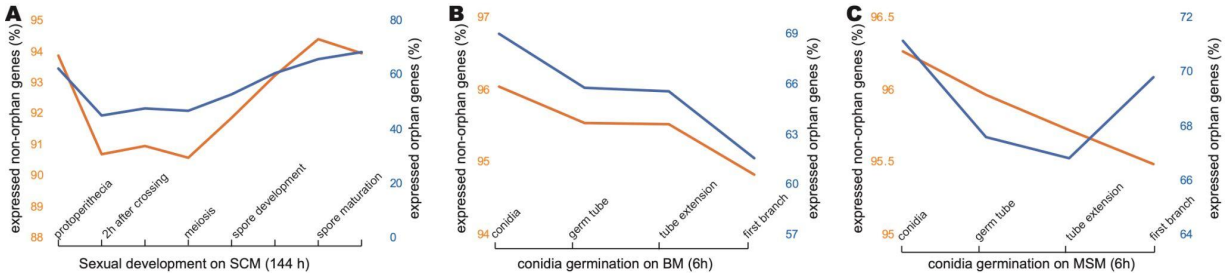

Figure S1. Proportion changes of orphan (blue) and non-orphan genes (orange) being expressed at sampling points from sexual development on SCM and conidial germination and growth on BM and MSM. (**A**) sexual development from protoperithecia (starting stage) to mature perithecia at 144 h (Wang et al. 2014); (**B**) asexual growth from conidial germination to the first hyphal branching on Bird medium supporting only asexual development; (**C**) asexual growth from conidial germination to the first hyphal branching on maple sap medium supporting both asexual and sexual reproduction (conidial germination; Wang, Miguel-Rojas, et al. 2019).

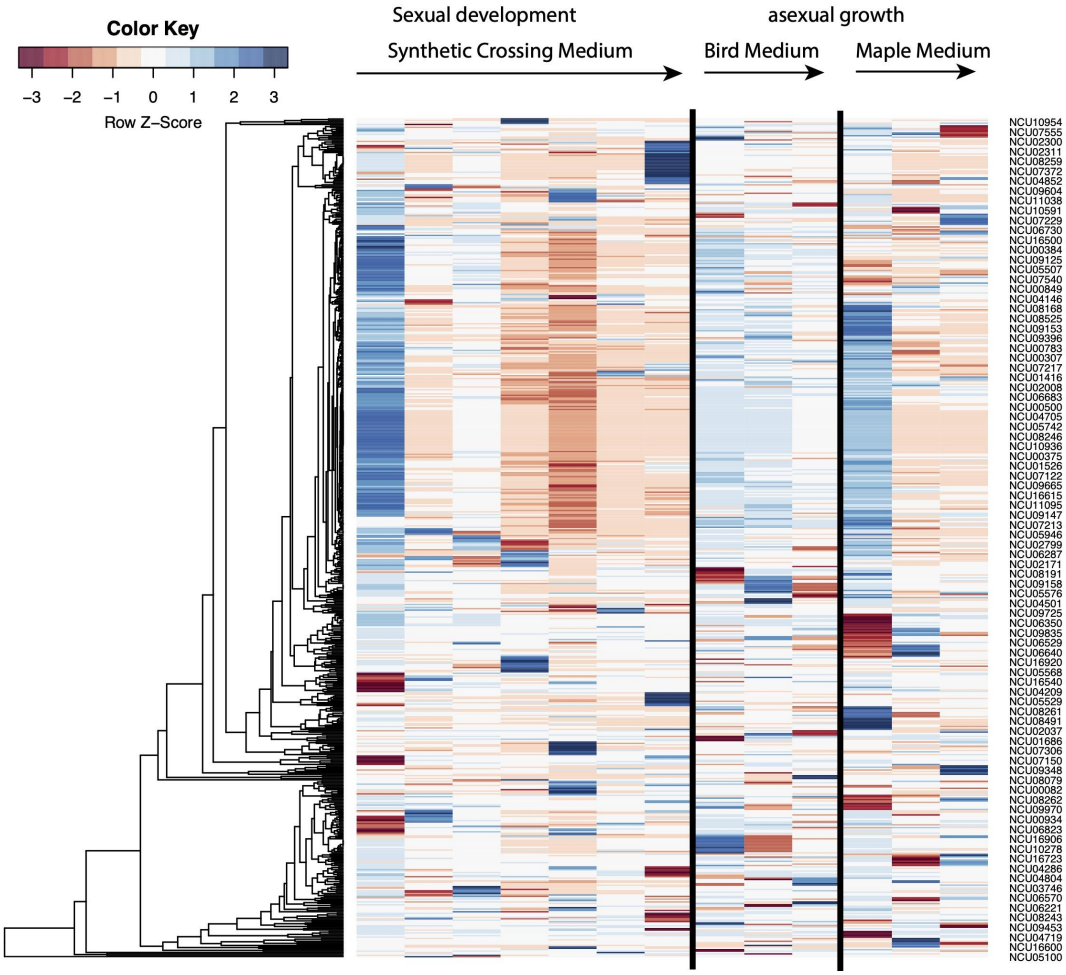

Figure S2. Gene expression dynamics of orphan genes across sexual development from protoperithecia (starting stage) to mature perithecia at 144 h (Wang et al. 2014) and asexual growth from conidial germination to the first hyphal branching, on Bird medium supporting only asexual development, and on a maple sap medium supporting both asexual and sexual reproduction (conidial germination; Wang, Miguel-Rojas, et al. 2019).

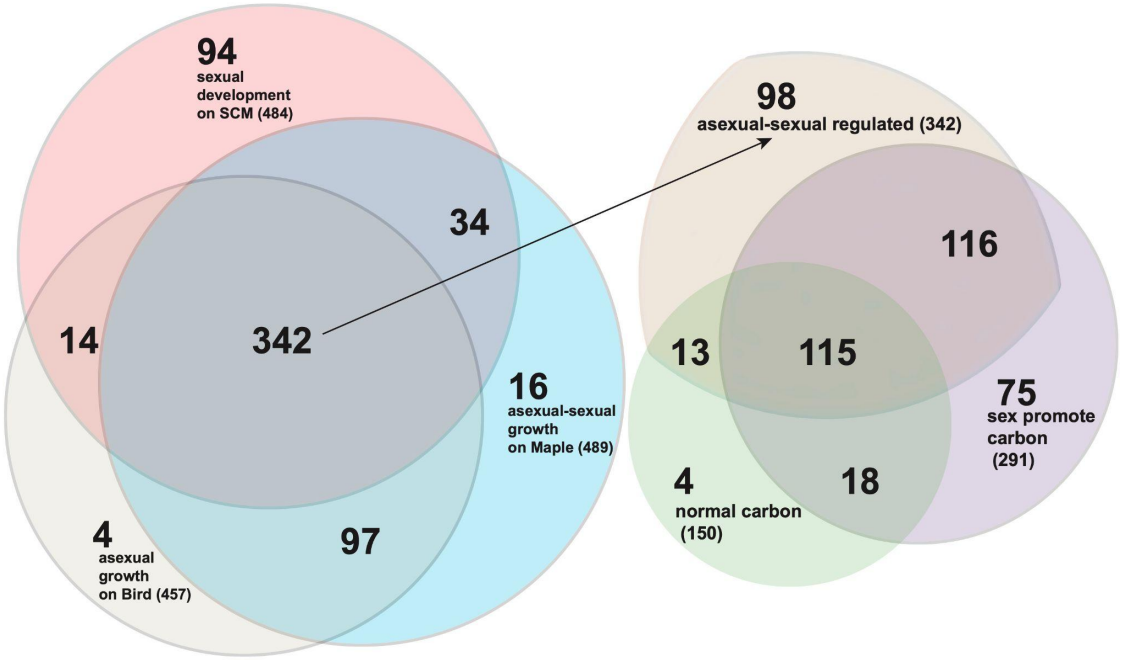

Figure 3. Differential expression of orphan genes in *N. crassa* growth during three developmental processes (Wang et al. 2014; Wang, Miguel-Rojas, et al. 2019) on media with various carbon resources (Wang et al. 2015; Feldman et al. 2019). 342 orphan genes were expressed at measurable levels in at least two stages in each of the three developmental processes (**Tables S6–S7**), including eight stages of sexual development on SCM (salmon pink, totaling 484 orphan genes with measurable expression), four stages of asexual growth on BM (beige, totaling 457 orphan genes with measurable expression) and four stages of asexua-sexual growth on MSM (light blue, totaling 489 orphan genes with measurable expression). Among the 342 orphan genes (arrow-linkws), 231 were detectably expressed when cultured on either 2-furaldehyde furfural and/or 5-hydroxymethyl furfural (HMF), substrates that promote sexual development (purple, total 291 orphan genes); and 128 were detectably expressed on sucrose and/or residues of at least one of five common crop straws (barley, corn, rice, soybean, and wheat; **Tables S6–S7**), substrates that support asexual growth and sproulation (light green, totaling 150 orphan genes).

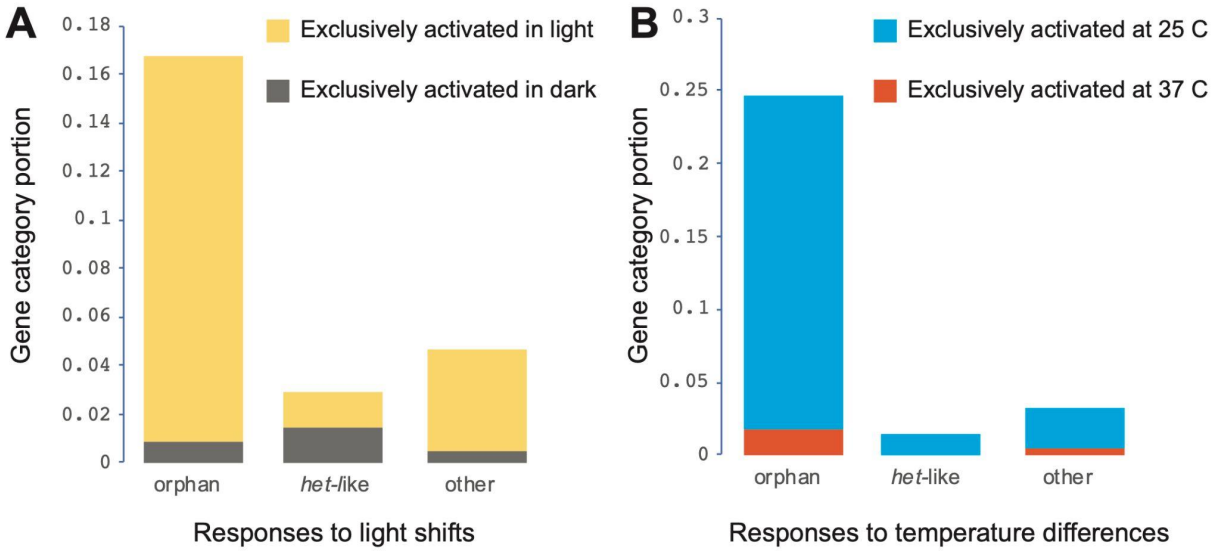

Figure 4. Expression responses of orphan, het-like, and other genes to shifting of culture environment. (**A**) Proportions of 670 orphan, 69 *het*-like, and other genes in the *N. crassa* genome, exhibited expression only in the dark (gray) or only in response to 15 to 240 minutes of light exposure (yellow). (**B**) Proportions of 670 orphan genes, 69 het-like genes, and other non-orphan genes in the *N. crassa* genome that are expressed across 4-stages of conidial germination and vegetative growth only at 25 C (blue) or 37 C (red) on Bird medium.

Coordinate expression of clustered orphan and het-like genes**—**The majority of orphan genes were clustered into physically linked groups—64% with *het* or *het*-like genes (**Table S4**). Forty-two *het*-like genes clustered with at least one orphan gene, including *het*-like orphan genes NCU03378, 07596, and 10839. Other than these three *het*-like orphan genes all clustered *het*-like genes exhibited measurable expression in at least three out of the four sampled stages in conidia germination as well as six out of eight stages sampled during sexual development.

Coordinated expression among genes within the clusters during *N. crassa* asexual and sexual growth and development was not common. Among more than 20 cases where gene-expression dynamics were highly coordinated across several developmental stages, two are worth mentioning (**Fig. 5**). They are cluster #69 of *het-14* and two orphan genes (NCU07510 and 08191) and the cluster #117 of a *het*-like gene (NCU11054) and four orphan genes (NCU03467, 03469, 03474, and 16509). Genes in the cluster with *het-14* exhibited highly coordinated expression during sexual development, even with a large non-coding sequence of over 15000 bp separating them from *het-14* and NCU07510 in that cluster (**Fig. 5A–C**). Interestingly, the non-orphan genes embraced by the rest of the cluster exhibited expression dynamics that were similar to the clustered orphan and *het*-like genes during those stages. Genes in the cluster #117 exhibited highly coordinated expression during asexual growth on MSM (**Fig. 5D–F**). Orphan genes clustered with *het*-like genes NCU07335, 10142, and 16851 also exhibited coordinate expression during condial germination and asexual growth on BM and on MSM (**Fig. S3**).

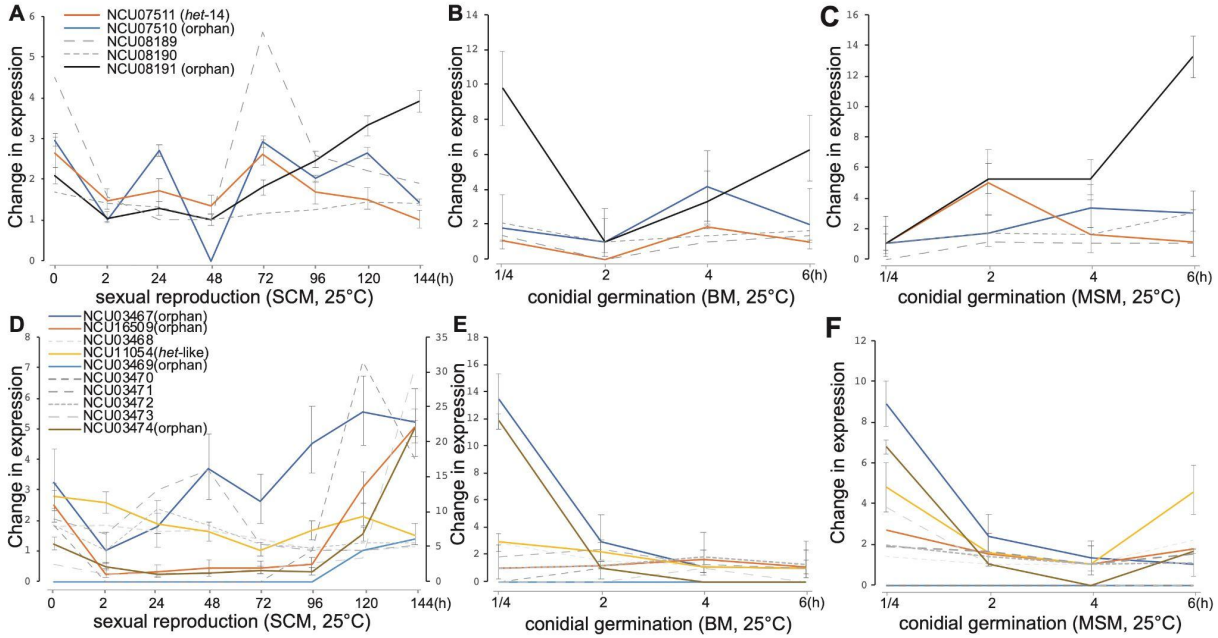

Figure 5. Expression profiles of two orphan-*het* gene clusters across asexual and sexual growth in *N. crassa*. expression profiles are plotted for orphan and *het*-like genes clustered with *het-14* (color coded, with 95% credible intervals)—as well as the non-orphan genes NCU08189 and NCU08910 embraced by the cluster (grey dashed)—cultured through (**A**) eight stages of sexual development on synthetic crossing medium, (**B**) four stages of conidial germination and asexual growth on Bird medium, and (**C**) four stages of conidial germination and asexual growth on maple sap medium. Expression profiles for genes clustered with *het-14* (color coded, with 95% credible intervals), including the five non-orphan genes NCU03468, 03470, 03471, 03472, and 03473 (grey dashed) embraced by the cluster, across (**D**) eight stages of sexual development on synthetic crossing medium, (**E**) four stages of conidial germination and asexual growth on Bird medium, and (**F**) four stages of conidial germination and asexual growth on maple sap medium.

Genes in cluster #112 exhibited no measurable expression in least at three out of the four sampled stages across conidia germination and across asexual growth, and genes in cluster #172 exhibited no measurable expression in at least six out of eight sampled stages in sexual development in *N. crassa*. However, expression was observed to be coordinated in a few of the clusters (**Fig. S3**): three clusters exhibited coordinated expression across sexual development (**Fig. S3A–C**), eight exhibited coordinated expression across conidial germination and asexual

growth on Bird medium (**Fig. S3D–K**), and 15 exhibited coordinated expression across conidial germination and asexual growth on maple sap medium (**Fig. S3L–Z**).

Two of the three orphan-gene clusters, including the cluster #24 (**Fig. S3A**) and the cluster #131 (**Fig. S3C**) exhibited coordinated expression among the genes within each cluster during sexual development, and the coordinate expression was observed during the early stages of sexual development before meiosis (about 48–72 h after crossing). Of the eight clusters, expression during asexual growth on BM was coordinately down-regulated in seven of them (**Fig. S3D–K**). The exception was the cluster #50 of NCU07335 (*het*-like) and 07336; expression of these genes was up-regulated during the process (**Fig. S3I**). Coordinate expression during asexual growth on MSM exhibited very different regulation direction: 12 out of 15 clusters exhibited up-regulated expression patterns toward the extention of first hyphal branch (**Fig. S3L–Z**), including two orphan-*het* clusters: the cluster #50 (**Fig. S3S**) and the cluster #51 of NCU16851 (*het*-like), 07316, 07317, and 07323 (**Fig. S3T**).

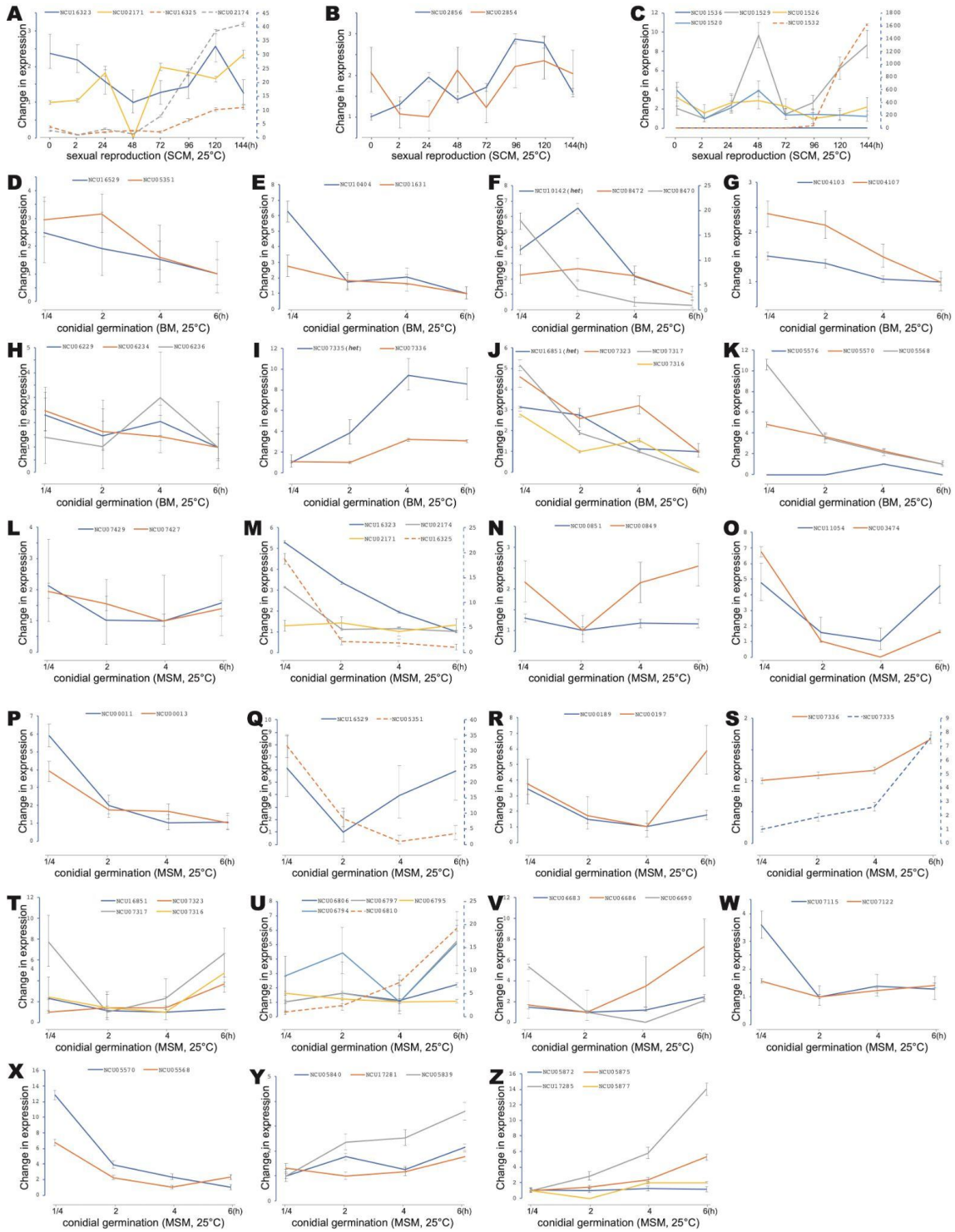

Figure S3. Expression profiles of 21 orphan-gene clusters and orphan-*het* gene clusters (Table S4) across asexual and sexual growth in *N. crassa*. Expression and 95% credible intervals for (**A–C**) genes in clusters 24, 37, and 131 during sexual development, (**D–K**) genes in clusters 121, 128, 133, 88, 62, 50, 51 and 8 during conidial germination and asexual growth on Bird medium, and (**L–Z**) genes in

clusters 22, 24, 33, 117, 63, 121, 65, 50, 51, 125, 87, 1, 8, 109 and 110 during conidial germination and asexual growth on maple sap medium.

Cell communication transcription factors affect expression of orphan and het-like genes — The transcription factors *adv-1* and *pp-1* play multiple roles in asexual and sexual development, cell growth and fusion, and cell communication in *N. crassa* and closely related fungi (Colot et al. 2006; Fu et al. 2011; Dekhang et al. 2017; Fischer et al. 2018; Lan et al. 2021). Functions and regulatory networks involving *adv-1* and *pp-1* were systematically investigated (Fischer et al. 2018), and in that study 155 genes were identified that were likely positively regulated by both transcription factors, including two *het*-like genes (NCU03494 and 09954) and one orphan gene (NCU17044). We reanalyzed their RNA sequencing data collected during conidial germination from knockout mutants of *adv-1* and *pp-1* and from wild type strain (data from Fischer et al. 2018), and in general, knocking out the two transcription factors had comparative high impacts on the expression of orphan genes (**Figs. 6, S3, Tables S10, S11**). Unlike *het*-like genes and other non-orphan genes, a significantly large portion of 173 orphan genes exhibited no expression in the knockout mutants and the wild type strain, including 196 not expressed in both *Δpp-1* and the wild type, and 195 not expressed in both *Δadv-1* and wild type. At the same time, significant numbers of orphan genes were turned on or off by the mutations to the two transcription factors. Namely, the same number of 44 orphan genes that were expressed in wild type were inactivated in the mutant strains *Δpp-1* or *Δadv-1*, and 23 orphan genes were inactivated in both the mutant strains *Δpp-1* and *Δadv-1.* Among the 23 orphan genes, 20 were within the predicted orphan-het clusters, including NCU04700 and 04710 that clustered with *het*-like gene NCU04694 and three core genes NCU08822, 08829, and 08830 in a six-gene cluster. At the same time, expression of

significant numbers of orphan genes (53 and 54 separately) went from undetectable to detectable in the *Δpp-1* or *Δadv-1* knockout strains, with expression of 31 orphan genes becoming detectable in both knockout mutants. However, only 14 out of the 31 newly detectably expressed genes were clustered orphan genes. Therefore, knocking out these two transcription factors did not positively affect the orphan-*het* gene clusters.

Twenty-seven *het*-like genes are expressed at higher level in the wildtype strain than in both mutants (41 for *Δpp-1*, and 35 for *Δadv-1*), including four *het*-like genes (NCU03494, 06583, 09037, and 09954) exhibited more than five-fold higher expression in the wild type strain than in the *Δpp-1* and/or *Δadv-1* strains. However, NCU03494, 06583, and 09037 are *het*-like genes not clustered with any orphan genes. For orphan and *het*-like genes with well measured expression in mutants and wild type, many genes exhibited similar up- or down-regulation between the *Δpp-1* and *Δadv-1* compared with their expression in the wild type (**Fig. S3**).

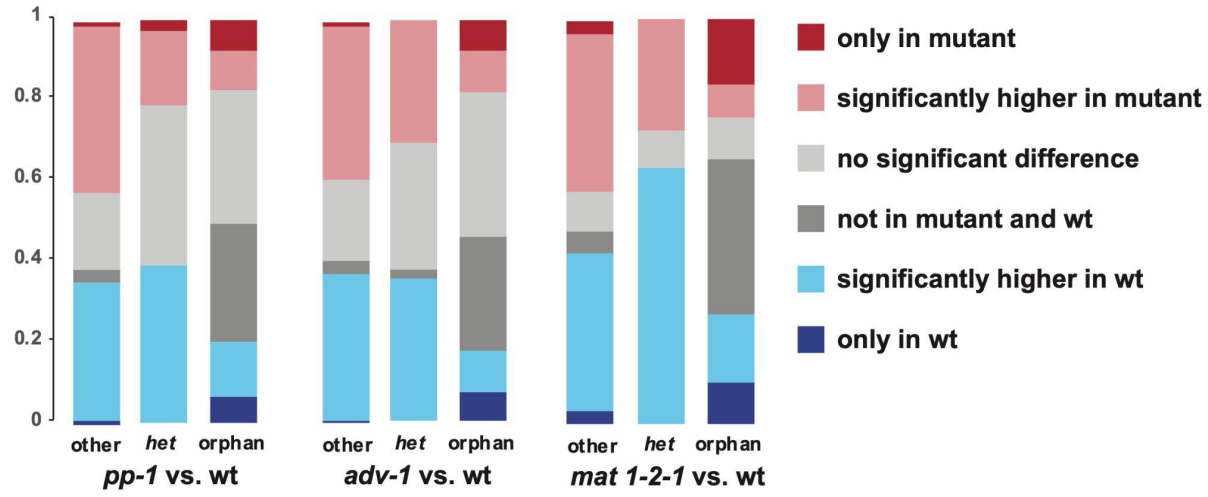

Figure 6. Impacts of transcription factor (TF) deletions on expression of *het*-like and orphan genes within the *N. crassa* genome. Comparatively larger portions of orphan genes are expressed exclusively in the mutants (dark red) or in the wildtype strains (dark blue). Expression profiles were classified in five categories (green: only measurable expression in mutant; light blue: higher in mutant with $P < 0.05$; yellow: no significant difference with $P \geq 0.05$; grey: significantly higher in wildtype with $P < 0.05$;

orange: only present in wildtype; dark blue: no measurable expression in mutant and wildtype), and comparative portions of each category in *het*-like genes, orphan genes, and other genes in the genome were reported. Data from Fischer et al. (2018) were reanalyzed to assess the impacts of *pp-1* and *adv-1* knockout mutants.

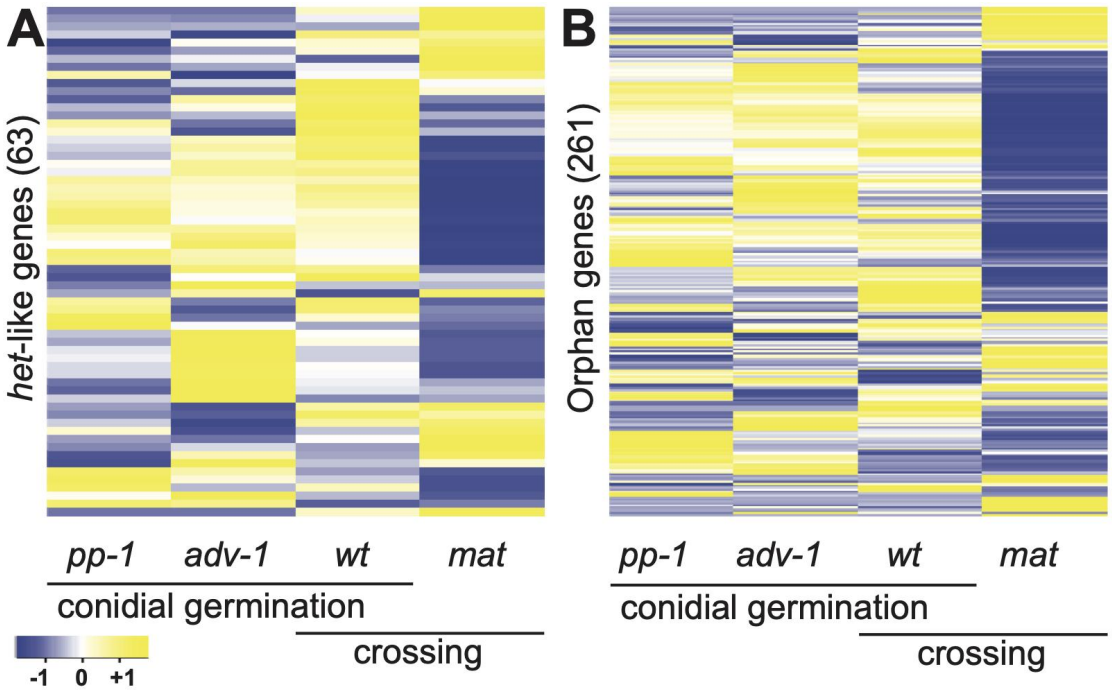

Figure S4. Heatmaps demonstrating divergent expression between wild-type strains and deletions of *pp-1* and *adv-1* in their expression of *het*-like and orphan genes in conidial germlings, and in mutants of *mat 1-2-1* in fertilized protoperithecia (crossing). (**A**) Expression divergence of *het*-like genes in three mutants *vs*. wild type, expression levels sampled in crossing were scaled according to the wild-type expression measured in germlings, and *het*-like genes with 0 measurements in wild type were excluded; (**B**) Expression divergence of orphan genes in three mutants *vs*. wild type, expression levels sampled in crossing were scaled accordingly to the wild type expression measured in germlings, and orphan genes with 0 measurements in wild type were excluded; (**C**) Expression divergence of orphan genes in three mutants *vs*. wild type, with orphan genes exhibited 0 expression in multiple samples, and 133 orphan genes exhibited 0 expression in all samples were excluded.

N. crassa mating loci regulate crossing and sexual development, and opposite mating-type strains are ordinarily vegetatively heterokaryon-incompatible (Newmeyer 1970; Jacobson 1992;

Pöggeler and Kück 2000; Xiang and Louise Glass 2004; Wang, López-Giráldez, et al. 2019). Transcriptome profiles were compared between six-day cultures of wildtype strain and a mating locus *mat 1-2-1* mutant (FGSC#4564, *mat a*[m1]s-3B cyh-1) on synthetic crossing medium (SCM). Of 9758 measured genes, a total of 836 genes exhibited undetectable expression only in either the mutant or wildtype (**Figs 6, S4. Tables S10, S11**), including 179 orphan genes (109 expressed and 70 undetected in the *mat 1-2-1* mutant) and 657 non-orphan genes (336 expressed and 321 undetected in the *mat 1-2-1* mutant). For orphan genes, lack of detectable expression occurring only in the wildtype or only in the mutant was significantly enriched ($P < 0.00001$, chi-squared test). Of the 109 orphan genes that were exclusively expressed in the *mat 1-2-1* mutant, 71 were located within 55 predicted orphan-*het* clusters. Of the 70 orphan genes that were only expressed in the wild type, 59 were located within 47 predicted orphan-*het* clusters. Only 17 clusters were common between the two groups, and larger clusters with more than three genes presented behaved differently between the *mat 1-2-1* mutant and wild type, with three genes of five-gene cluster #51 (NCU07306–07323) being inactivated in the mutant, while four genes of seven-gene cluster #124 (NCU05480–06949) and five genes of the nine-gene cluster #94 (MCU07144–07152) being activated in the mutant. Some two-gene clusters exhibited coordinated expression that was detectable only in the mutant or the wild type. There were 53 orphan genes expressed significantly higher ($P < 0.05$) in the mating type mutant, and 111 expressed significantly higher in the wild type. However, mutations in the mating locus have no such impacts on basic activities for *het*-like genes, with 18 and 41 *het*-like genes being expressed significantly higher in the mating locus mutant, and 41 *het*-like genes being expressed significantly higher in the wild type (**Figs. 6, S4**).

Binding-site enrichment analysis using CiiiDER (Gearing et al. 2019) identified no significant enrichment of binding sites for *mat 1-2-1*, *adv-1*, or *pp-1* in the upstream 5000 bp of orphan genes that exhibited activity divergence in the mutants, specifying orphan genes that whose expression was unchanged between the mutant and wildtype as the background. Orphan genes and *het*-like genes were not significantly enriched in genes that are potentially bound by *adv-1* or *pp-1*, based on data from DAP-seq (SRP133627 from Fischer et al. 2018). However, knocking out a key transcription factor encoding gene, *ada-6* (Sun et al. 2019)—whose product regulates asexual and sexual growth—inhibited expression of 30 orphan genes and promoted expreession of 25 orphan genes during conidiation and protoperithecial production. Therefore, a substantial number of orphan genes are at least peripherally involved in those regulatory networks, abundant enough that they may play fine regulatory roles, and sparsely distributed enough to impact diverse regulatory pathways.

***Selection profiles of orphan genes*—**an original study reported 135,000 SNPs in transcriptomic sequence from 48 individuals (Ellison et al. 2011), yet a total of 1,086,579 SNPs are reported from just 26 *N. crassa* isolates whose transcriptomic sequence data are available in FungiDB. The discrepancy likely arises as a consequence of the exclusion of singletons from the original data (Ellison et al. 2011), which were explicitly left unreported in the original paper, but were frequently observed (C. Ellison personal communication, July 13, 2020). 97.8% (655 out of 670) orphan genes feature at least one non-synonymous SNP, which is moderately but statistically significantly higher (Fisher exact test, $P < 0.01$) than the 91.6% (8271 out of 9027) non-orphan genes, as revealed by the inclusive SNP data in FungiDB (**Fig. S5**). Unlike orphan genes that

generally encode less than 250 amino acids, most *het*-like genes encode more than 300. Non-synonymous SNPs were observed at similar frequencies between orphan genes and *het*-like genes with similar gene length (**Fig. 7)**. Excluding singletons from the FungiDB data markedly reduced the number of SNPs—only eight *Neurospora* orphan genes then featured SNPs (1.3%) compared to 4192 non-orphan genes (45.9%). Therefore, SNP singletons are abundant within *Neurospora* orphan genes. A high density of SNP singletons in these typically short orphan genes corresponds to a high allelic polymorphism of orphan genes, present in a comparative low proportion in the population; such a pattern can arise from selection for rare alleles.

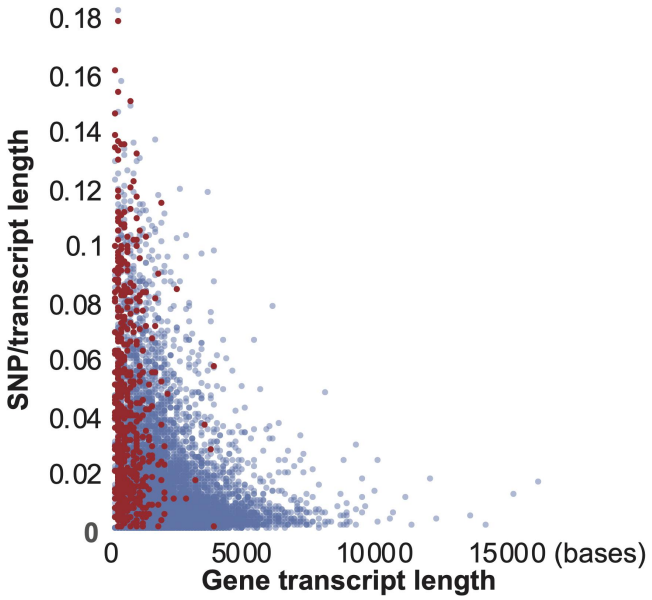

Figure S5. SNPs per nucleotide of transcript—including singletons—for 9756 *N. crassa* genes (orphan, red; non-orphan, blue) from 26 sequenced strains in FungiDB.

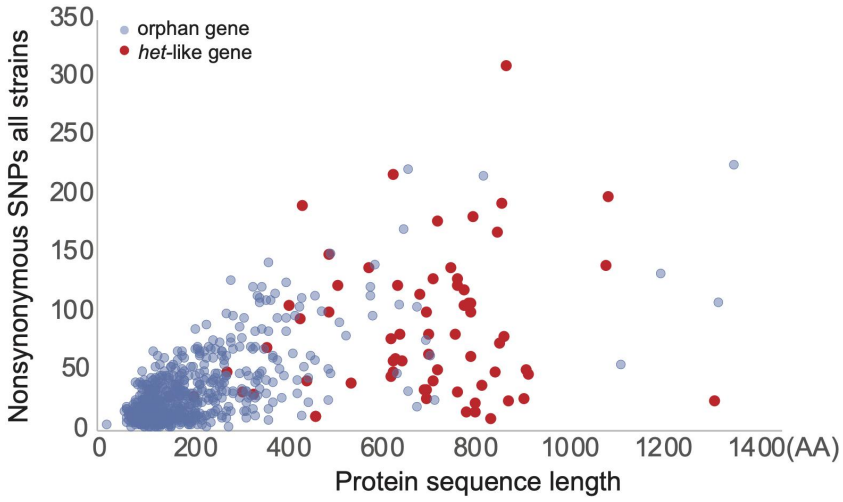

Figure 7. Nonsynonymous SNP distribution comparison between orphan genes (blue circles) and *het*-like genes (red circles) from *N. crassa* population genetic sequencing (Ellison et al. 2011). Nonsynonymous SNPs are totaled across all 26 *N. crassa* strains deposited at FungiDB.

Of all 603 *Neurospora* orphan genes, including those with singleton SNPs, only NCU07210 exhibited statistically significant strong gene-regional positive selection during the divergence of *N. crassa* from its common ancestor with *N. discreta* (**Fig. S6A**). No abnormal phenotype resulting from the knockout of this gene (*mat-a* strain FGSC14291) was observed in our assays. For 695 orphan genes shared within *Neurospora-Sordaria*, 16 genes exhibited statistically significant evidence of strong gene-regional positive selection, and five exhibited statistically significant evidence of moderately strong gene-regional positive selection (**Table S12**). Five of 16 strongly positively selected genes were identified in six biological processes related to the response to stress or stimulus (**Table S13**), including the hypothetical protein coding gene NCU02932 with the ortholog being annotated as a regulator of ATP-sensitive K^+ channels in *N. discreta* and likely play roles in stress responses (**Fig. S6B**). Further exploration suggested that positively selected *Neurospora-Sordaria* orphan genes NCU05395, 07618, 09562, and 09693 are likely involved in the regulation of sexual development (Leeder et al. 2013). Furthermore, regulation of the moderately positively selected

genes NCU01306, 00496, and 00748 is likely associated with conidiation in *N. crassa* (Greenwald et al. 2010; Sun et al. 2012; Sun et al. 2019).

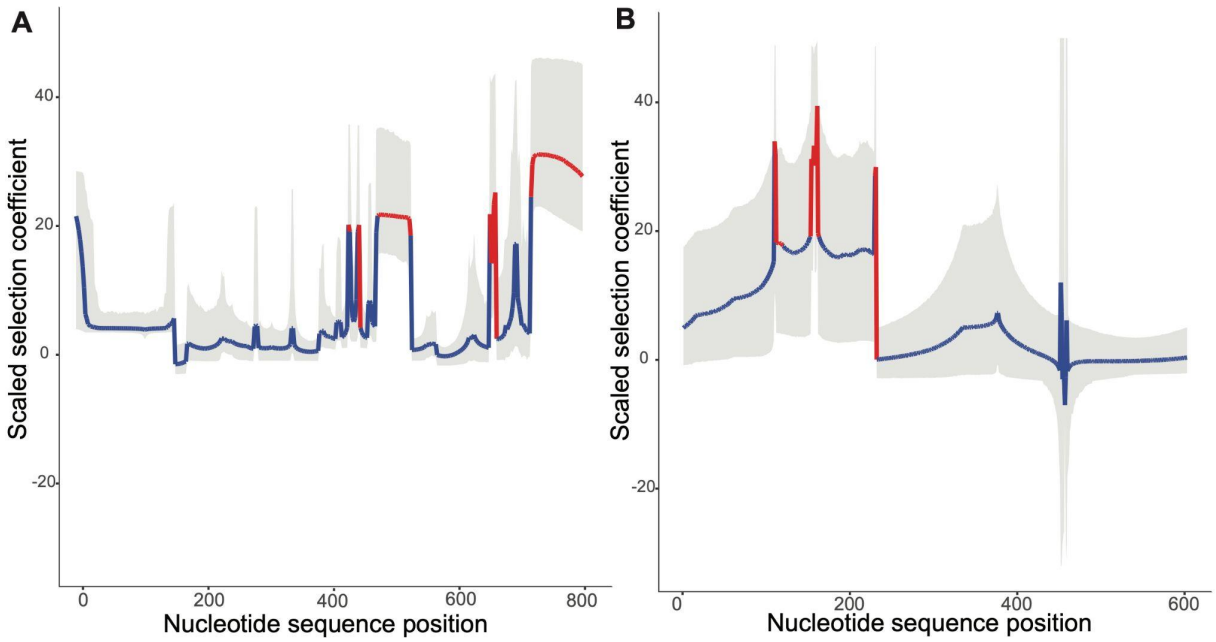

Figure S6. Window-free regionalized nucleotide site-specific selection profiles in two orphan genes showing statistically significant gene-regional positive selection during divergence of *N. crassa* from its common ancestor with *N. discreta*. Scaled selection coefficients (model-averaged γ) are plotted (red: lower bound of $\gamma > 4$; blue: lower bound of $\gamma \leq 4$; gray band: 95% model uncertainty interval) by nucleotide position for (**A**) NCU07210, a *Neurospora* orphan gene, and (**B**) NCU02932, a *Neurospora-Sordaria* orphan gene.

KO phenotypes of orphan genes—In an examination of phenotypes of crosses of 367 available KO strains of the *Neurospora* orphan genes to the KO of the opposite mating type (or the WT of the opposite mating type if the KO of the opposite mating type was not available), two *Neurospora* orphan genes, NCU00176 and 00529, and one *Neurospora-Sordaria* orphan gene, NCU00201, showed a distinct knockout phenotype in sexual development (**Fig. 8**). All these knockout mutants exhibited arrested development with the protoperithecia failing to develop into perithecia. Interestingly, both NCU00176 and NCU00201 were expressed at significantly higher

levels in protoperithecia than in mycelia after crossing. Expression of NCU00529 was observed only in the late stage of conidial germination on maple medium. NCU00529, 00530, and 00531 are homologs, and NCU00529 and 00530 exhibited no expression in sexual development and conidial germination. No abnormal phenotypes were identified for NCU00531 knockout mutants (FGSC13078 and 13079), and no knockout mutants for NCU00530 are available. Knockouts of NCU00375, 00384, 00485, and 00491 (*Neurospora* orphan genes) and NCU01623, 05395, 07618 (*Neurospora-Sordaria* orphan genes) have been reported to exhibit minor phenotypic anomalies during asexual growth, especially at 37 C (Fungidb.org/fungidb). We examined knockout mutants of these genes and confirmed increased pigment production in NCU00485 and dense and slow growth in NCU00491 and 016223 at 37 C.

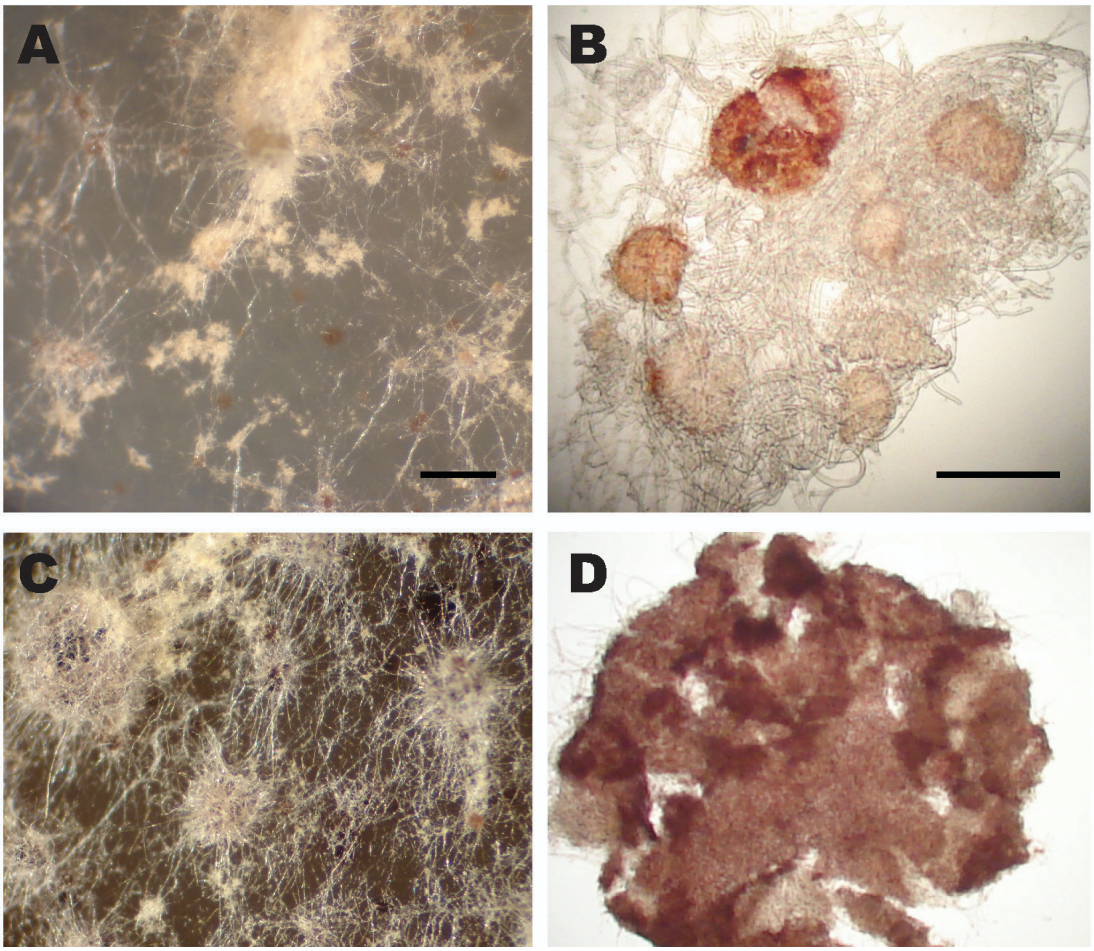

Figure 8. Knockout mutants of NCU00176, NCU00201, and NCU00529 produced arrested protoperithecia that failed to develop into perithecia and to produce sexual spores on SCM medium. Knockout cross ΔNCU00176 (FGSC12195 × 12196) exhibited (**A**) small protoperithecia (scale bar: 1 mm) and (**B**) squashed protoperithecia exhibiting an abortive ascogenous center (scale bar: 10 μm). Knockout cross ΔNCU00201 (FGSC18867 × 18868) exhibited (**C**) normal-sized protoperithecia (scale bar: 1 mm) with (**D**) abortive ascogenous centers (scale bar: 10 μm). Knockout cross ΔNCU00529 (FGSC13076 × 13077) exhibited (**E**) normal-sized protoperithecia with (**F**) abortive ascogenous centers.

Discussion

In this study, we systematically investigated orphan genes in *Neurospora* genomes to determine their location and organization on chromosomes and possible roles in biology and ecology. Using genomic phylostratigraphy and reciprocal BLAST searching, we identified 670 genes that are *Neurospora* orphan genes, only shared within *Neurospora* species. More than 63% of the 670 *Neurospora* orphan genes form clusters of 2–21 orphan genes, interspersed with *het*-like putative allorecognition genes. Many of the larger clusters aggregate near the telomeres of the seven chromosomes, and a majority of *Neurospora* orphan genes are actively regulated during asexual and sexual reproduction in reponse to carbon sources, light, and temperature changes. Knockout mutant strains were phenotyped for 367 *Neurospora* orphan genes, and arrested protoperithecia were observed for mutants of three *Neurospora* orphan genes. Our data indicate that regulation of asexual and sexual reproduction surprisingly empower evolution of *de novo* elements more than the regulation of core metabolism and cellular biology in *N. crassa*. More specifically, this regulation is associated with reproduction modes affected by carbon

resources—a critical environmental factor for this post-fire fungus. However, many orphan genes were not identified as essential for *N. crassa* development and biology. Therefore, it is possible that orphan genes play key roles in fine regulation in response to environmental factors inducing reproduction processes that have been understudied.

Neurospora orphan genes and het-like genes are tightly clustered—*Neurospora* orphan and *het*-like genes exhibit several organizational features on chromosomes. These organizational features include gene clustering, large non-coding spaces enclosed by flanking condensed coding regions, high frequencies of gene duplications, and proximity to telomeres. If clustered genes were derived from relocation and rearrangement, especially for functionally associated genes, it would probably be easy for these genes to functionally integrate back altogether into the existing system. Indeed, coordinated expression across clustered orphan and *het*-like genes was sometimes detected—especially during *N. crassa* development in response to various environmental factors. However, orphan-*het* gene clusters and their neighbor genes were syntenic mainly within very closely related taxa, and many orphan genes exhibited no expression under common laboratory settings—even across a range of nutrient conditions and developmental stages. Interestingly, heterochromatic interactions of intra- and inter-telomeric contacts were reported as common in *N. crassa* (Rodriguez et al. 2022), implying potential *cis*- and *trans*-regulation at chromosomal level of expression integration in these orphan-gene enriched regions. It is possible that integrating orphan-*het* gene clusters into existing regulatory systems temporally and spatially modifies the original functions that are expected only under common growth conditions. Therefore, specific experiments conducted in extreme growth

conditions, with rare nutrient types, and at understudied stages of the life cycles should be conducted.

A previous study characterizing chromosome ends in *N. crassa* demonstrated that highly AT-rich sequences in the telomeres are likely products of RIP and that subtelomeric elements common in other fungi are absent in *N. crassa*. Telomere repeats are required for H3K27 methylation, which would repress the transcription activities and functionally silent genes in these regions (Jamieson et al. 2018). More importantly, the telomeric regions have potential significance in niche adaptation and probably harbor hotspots for novel sequences due to abrupt sequence divergence involving repeats (Wu et al. 2009). Many genes of an annotated *het*-domain also locate near the ends of some *N. crassa* chromosomes (Zhao et al. 2015), and 42 of the 69 *het*-like proteins, which account for heterokaryon incompatibility, are actually clustered with at least one *Neurospora* orphan gene. In fact, neighboring regions of *het*-like genes are abundant with "young" genes that lack homologs in lineages beyond *Neurospora* and closely related species in the Sordariales, and most neighbor genes are syntenic within Sordariomycetes. Most orphan genes were not active in sampled stages in the *N. crassa* life cycle. However, a few orphan-*het* gene clusters exhibited coordinate expression during early sexual development as well as during active hyphal tip growth on maple sap medium that support both asexual and seuxal developments. Therefore, investigation of gene activities in the telomeric and sub-telomeric regions during mycelium development and allorcognition between same and different vegetative compatibility groups will shed light on possible concerted associations between these two functional groups in heterokaryon imcompatibility, pre-mating process and even during early sexual developmen in *Neurospora*.

Actually, three orphan genes, including NCU03378, 07596, and 10839, were annotated as a *het*-domain, and NCU03378 was annotated as *tol* (tolerant, Newmeyer 1970) and shared several conserved sequence regions with *het*-6 that is involved in incompatibility in the *N. crassa* population (Mir-Rashed et al. 2000). Functional associations among some *het*-like genes and orphan genes within the orphan-*het* clusters were evidenced with similar gene expression regulation pattern during asexual and sexual development in *N. crassa*. However, compared with the *het*-like genes that were all expressed, many orphan genes were not expressed in the sampled stages during the *N. crassa* life cycle. Further investigation on epigenetic modification of gene expression near telomeric neighborhoods could be critical to capture the right moments of associated functions between orphan genes and *het*-like genes during allorecognition and to understand the evolution of orphan genes in *N. crassa*.

Neurospora orphan genes play roles in response to key regulatory environmental factors— Orphan genes are generally not functionally annotated, mainly due to the lack of homologous references in well studied genomes. To determine the functional roles of *Neurospora* orphan genes ,we revisited recent high quality transcriptomics studies on *N. crassa*, covering almost all morphological stages in the *N. crassa* life cycle, cultures under different conditions and carbon or nitrogen resources as well as knockout mutants of key regulatory genes (Coradetti et al. 2012; Znameroski et al. 2012; Coradetti et al. 2013; Lehr et al. 2014; Wang et al. 2014; Wu et al. 2014; Craig et al. 2015; Wang et al. 2015; Xiong et al. 2017; Feldman et al. 2019; Sun et al. 2019; Wang, Miguel-Rojas, et al. 2019). We observed a significant enrichment of clustering of orphan and *het*-like genes and, within some of those clusters, highly coordinated regulation in response to arbon resources, light, and temperature conditions. This study provides evidence that some

orphan genes and clustered *het*-like genes are associated with metabolic adaptation to environmental factors that are critical indicators of successful fungal asexual and sexual growth and reproduction.

Three hundred forty-two orphan genes that were actively expressed during sexual development were also actively regulated in both asexual growth on two different media, including Bird medium, which supports only asexual reproduction, and maple sap medium, which supports both asexual and sexual growth. Nearly two thirds of the 342 genes were actively regulated in samples collected from the media supplied with carbon resources that promote sexual growth of *N. crassa*, supporting their possible roles in sexual development and the asexual-sexual switch. In favor of their roles in sexual development in *N. crassa*, more than one third of the orphan genes were actively regulated on the presence of specific carbohydrates, including HMF and furfural, compounds that powerfully stimulate initiation of sexual reproduction. *N. crassa* has been shown to respond differentially to the two furans and to possess a high tolerance to furfural, which is present in its natural habitat (Feldman et al. 2019). It is conceivable that some of the orphan genes that are uniquely expressed upon exposure to HMF or furfural could be further engineered to provide increased tolerance to atypical carbon resources for *N. crassa*, a trait of significant interest in the pursuit of robust biofuel production.

Neurospora orphan genes were observed to be active only at the hyphal tips without measurable expression from the colony samples, and thus these genes were considered to be responsible for environmental sensing and interaction with microbes (Kasuga and Glass 2008). Overall, many orphan genes are probably associated with reproductive development and growth in response to environmental conditions, especially carbon resources, light and temperature.

Furthermore, we observed that orphan genes in predicted clusters coordinately respond to these environmental factors. How these genes may possibly contribute together as fine quantitative switches for reproduction requires further investigation with multiple gene manipulations.

A significant proportion of orphan genes are regulated by key developmental transcription factors—Transcription factors, including *pp-1* and *adv-1* that play key roles in cellular communication during asexual growth and mating loci that regulate sexual crossing, have been previously reported (Pöggeler and Kück 2000; Bobrowicz et al. 2002; Fischer et al. 2018). From the previous transcriptomics data from knockout mutants of *pp-1* and *adv-1* and newly generated transcriptomics data from mutant of *mat 1-2-1*, we observed that the expression of a significant portion of orphan genes was affected by mutations in these transcription factors. Interestingly, over 95% orphan genes that were turned off in both *pp-1* and *adv-1*mutants belonged to predicted clusters, but less than 50% orphan genes that were turned on in both mutants belonged to predicted clusters. The expression of six genes in the NCU07144–07152 orphan cluster was actively regulated during sexual development, in knockout mutants of *adv-1* and *pp-1* that regulate cell communication with knockout phenotypes in sexual development (Colot et al. 2006; Dekhang et al. 2017), as well as in samples in the absence of carbon. Expression of some orphan genes was also affected in knockout mutants of other regulatory genes, such as *ada-6* and *gul-1* that are critical for sexual and asexual development in *N. crassa*. The most dramatic impacts to the expression of orphan genes were from the mutation in the mating locus, suggesting likely functional associations of mating process and sexual development initiation for orphan genes. However, no binding sites of transcription factors were observed enriched in the promote and up-stream sequences of orphan genes being turned on or off by those factors, and coordinate

expression regulation was only detected in a few orphan-*het* gene clusters. Instead of possible cis or trans regulations, an alternative explanation for the associations between the orphan genes and transcription factors would be that the orphan genes were coordinated with the sampled developmental stages when the transcription factors were actively engaged in promoting desirable gene expression. For example, the orphan genes exhibited different expression activities during hyphal branching on natural medium MSM, when cell-to-cell communication is regulated by *adv-1*. Therefore, success in further investigations will be aided by focusing on the epigenetics of the orphan-*het* gene clusters during specific periods of cell-to-cell communication.

Knockout phenotypes suggested that some orphan genes play essential roles in Neurospora development—our phenotyping of available knockout mutants of 367 orphan genes at the Fungal Genetics Stock Center (McCluskey et al. 2010) yielded three genes with a phenotype in sexual development. Within these three genes, NCU00529 and its homologs NCU00530 and 00531 form a cluster in the subtelomere of chromosome I. The other two genes with mutant phenotypes in sexual development, NCU00176 and 00201, exist on chromosome III. Their expression decreased after protoperithecia. NCU00176 was reported as one of the eight *Neurospora* orphan genes that were among the 300 down-regulated hypothetical proteins in the *gul-1* knockout mutant (Herold et al. 2019), and GUL-1 plays multiple roles in *N. crassa* hyphal morphology development (Herold et al. 2021). We also confirmed increased pigment production in NCU00485 and dense and slow growth in NCU00491 and 016223 at 37 C.

A recent study reported 40 biologically relevant clusters (BRCs) for 1168 *N. crassa* genes being phenotyped for 10 growth and developmental processes (Carrillo et al. 2020). Eleven *Neurospora* and 31 *Neurospora-Sordaria* orphan genes were included in this analysis.

Interestingly, 4 out of the 11 *Neurospora* ($P < 0.05$) and 5 out of the 31 *Neurospora-Sordaria* orphan genes were concentrated in one of the 40 BRCs (Cluster 4 of 81 genes) with minimum yeast orthologs and generally no significant phenotypes. An explanation for the lack of apparent phenotypes may be that these genes 1) were not investigated under the conditions for the phenotypes; 2) were functionally non-essential and/or not fully integrated into the regulatory networks, or 3) were functionally redundant within clusters or with non-orphan paralogs.

Orphan genes clustered with allorecognition loci to consolidate the incompatibility and speciation in Neurospora—Clustering of orphan genes would facilitate the integration of beneficial genes with existing regulatory networks, especially in genes associated with the regulation of asexual and sexual reproduction in response to environmental differences. The co-location of orphan genes with *het-like* genes—fast-evolving multicopy allorecognition loci—may indicate entwined functional interactions and evolutionary histories between these two gene categories. Functional interactions between the two gene groups were supported by coordinate responses to environmental and genetic regulatory factors. However, entwined evolutionary histories evidenced with the collocation of orphan and *het*-like genes within recent lineages requires further investigation at species and population levels. Therefore, SNP data from various populations, especially from different VCGs would be of great interest regarding the evolutionary associations between the two gene groups. In fact, balancing selections were detected in *het*-like genes (Zhao et al. 2015), and abundant rare singletons were observed for orphan genes, of which a similar occurrence of SNP polymorphisms was observed between *het*-like genes and orphan genes within the similar molecular size, based on data from limited *N. crassa* populations.

To guide investigation of associations between orphan genes and co-clustered allorecognition elements, we propose a working model based on discoveries on how orphan genes and clustered *het*-like genes respond to environmental factors, developmental stages, and key regulatory transcription factors (**Fig. 9**). This model suggests that orphan-gene associated mechanisms link vegetative incompatibility during asexual colonization to sexual incompatibility, enabling crossing between compatible mating types from different VCG backgrounds (**Fig. 9**). Vegetative incompatibility promotes the local aggregation and anastomosis of identical VCGs, which are probably associated with specific environmental preferences in nutrients, temperature, light, and osmosis, as well as parasite evasion, or other competitive advantages. In the absence of other mechanisms to maintain genetic diversity, purifying selection would lead to genomic homogeneity in allorecognition loci, which would be further enhanced supposing nearby crossings and sexual reproduction. However, the shift from vegetative incompatibility to sexual incompatibility during sexual reproduction could allow long-distance crossing to occur between VCGs, which would promote allelic diversity in VCGs via meiotic recombination (**Fig. 9**).

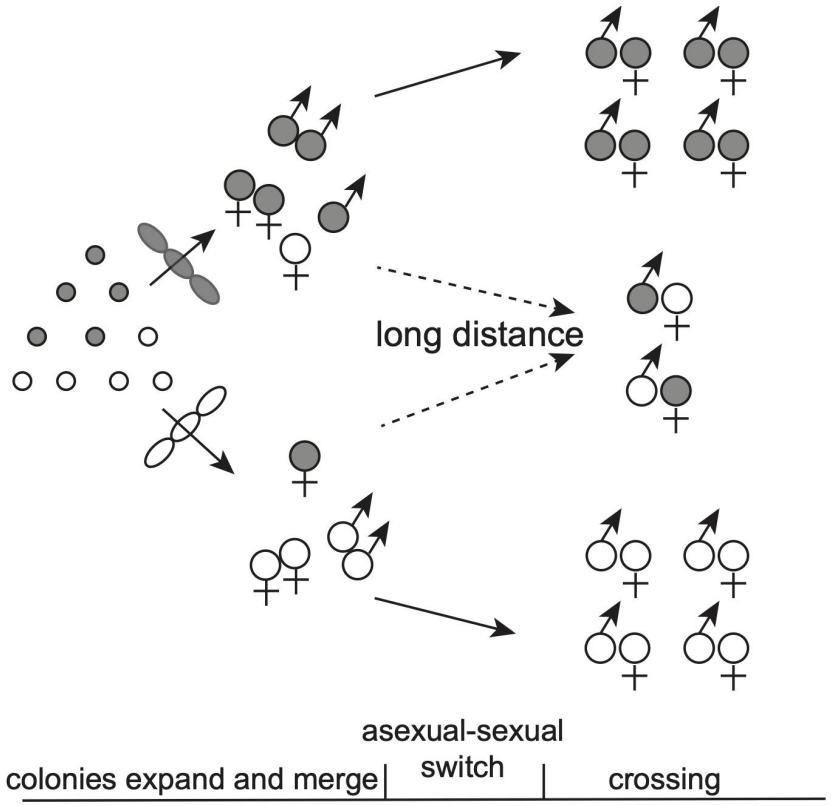

Figure 9. Model of possible interactions among *het*-loci, orphan genes, and mating loci in *N. crassa* speciation. Dark dots and white dots represent vegetative compatibility groups (VCGs) of popular *het*-allele and rare *het*-allele separately, and being regulated by the *het*-orphan clusters (three-circle chains), the two VCGs are adaptive to different vegetative growth environments and are vegetatively incompatible to each other. VCG colonies expand via aggregation and anastomosis under preferred vegetaive growth conditions until environmental turned on the asex-sex switch and shift into sexual reproduction with sexual incompatibility. Nearby crossing would promote segregation of VCGs, while long distance crossing would contribute to new haplotypes of *het*-allele. Orphan-*het* gene cluster impacts, represented with circle chains, likely consolidate vegetative incompatibility when environmental conditions preferred for asexual growth. Their roles also requires further investigation in response to mating loci and promote crossing between VCGs when environmental conditions preferred for sexual growth.

Materials and Methods

Mutant and culture conditions—Protoperithecia were sampled for *mat 1-2-1* mutant (FGSC#4564, *mat a*[m1]s-3B cyh-1). The experiments were performed with macroconidia, which were harvested from 5-day cultures on Bird medium (BM). 1×10^5 spores were placed onto the surface of a cellophane-covered synthetic crossing medium (SCM) in Petri dishes (60 mm, Falcon, Ref. 351007). Dark-colored protoperithecia were abundantly rippen in 6 days after inoculation. Tissue samples were flash frozen in liquid nitrogen and stored at －80℃. Biological replicates included all tissues collected from multiple plates in one collection process. Three biological replicates were prepared for each sampled point.

RNA isolation and transcriptome profiling, data acquisition and analysis—Total RNA was extracted from homogenized tissue with TRI REAGENT (Molecular Research Center) as in Clark et al. (2008), and sample preparation and sequencing followed our previous works (Trail et al. 2017; Wang, Miguel-Rojas, et al. 2019; Wang, López-Giráldez, et al. 2019). Briefly, mRNA was purified using Dynabeads oligo(dT) magnetic separation (Invitrogen). RNAseq Library Prep: mRNA was purified from approximately 200 ng of total RNA with oligo-dT beads and sheared by incubation at 94 C in the presence of Mg (Roche Kapa mRNA Hyper Prep Catalog # KR1352).

Following first-strand cDNAtA-tailing was performed with dUTP to generate strand-specific sequencing libraries. Indexed libraries were quantified by qRT-PCR using a commercially available kit (Roche KAPA Biosystems Cat# KK4854). The quality of cDNA samples was verified with a bioanalyzer (Agilent Technologies 2100).

The cDNA samples were sequenced at the Yale Center for Genomics Analysis (YCGA). The libraries underwent 76-bp single-end sequencing using an Illumina NovaSeq 6000 (S4 flow cell) according to Illumina protocols. Adapter sequences, empty reads, and low-quality sequences were removed. Trimmed reads were aligned to the *N. crassa* OR74A v12 genome from the Broad Institute (Borkovich et al. 2004) using `HISAT2` v2.1, indicating that reads correspond to the reverse complement of the transcripts and reporting alignments tailored for transcript assemblers. Alignments with a quality score below 20 were excluded from further analysis. Reads were counted for each gene with `StringTie` v1.3.3 and the Python script `prepDE.py` provided in the package. `StringTie` was limited to report reads that matched the reference annotation. Sequence data and experiment details were made available (GEO199259) at the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Statistical analysis of the sequenced cDNA tallies for each sample was performed with `LOX` v1.6 (Zhang et al. 2010), ignoring raw reads that mapped ambiguously or to multiple loci.

Genomic phylostratigraphy of Neurospora crassa genome—Genomic phylostratigraphy for the *N. crassa* genome was estimated as described in (Kasuga et al. 2009). Briefly, we used the SIMAP (similarity matrix of proteins) database developed by MIPS (http://mips.gsf.de/simap/) and calculated the pairwise similarity of protein-coding sequences by Smith-Waterman pairwise comparison (Arnold et al. 2005; Rattei et al. 2006). Final Smith-Waterman scores ≥80 were stored in the database. To correct biased similarity scores due to sequence length, we compensated for the dependency of hits on gene length using the average similarity score per pair-mapped amino acid, $C_v = (S * M) / L$, where S is the similarity-matrix score, M is the

number of pair-mapped sites between query and subject sequences, and L is the length of the protein. Only protein-coding sequences with $C_v > 30$ were included in this study. We retrieved homologous proteins in hierarchical taxonomic units for the MIPS-curated *N. crassa* using the taxonomy search tool at the MIPS *N. crassa* database. Homologous clusters (Pellegrini et al. 1999; Cai et al. 2006) for each of the *N. crassa* genes were constructed based on the presence or absence of homologous sequences in the taxonomic units. The genes were classified into mutually exclusive groups ranked in association with their phylostratigraphy, including Euk/Prok-core, Dikarya-core, Ascomycota-core, Pezizomycotina-specific, *N. crassa*-orphans, and others (not classified). To ensure quality of the retrieved dataset, homologous sequences of *N. crassa* PCGs were searched in genomes of *C. globosum*, *Saccharomyces cerevisiae*, *Phanerochaete chrysosporium*, *Drosophila melanogaster*, and *Arabidopsis thaliana*, respectively, using the same criteria used for the SIMAP database. Less than 1% of the SIMAP data and our homology-search data showed discrepancies in terms of groupings.

***Verification of Neurospora crassa orphan genes among closely related genomes*—**Many fungal genomes were recently published due to the efforts of the 1000 fungal genome project launched by the DOE Joint Genome Institute (Grigoriev et al. 2014). These fungal genomes— especially those well-sampled among closely related species—make it possible to be confident that orphan genes are likely the product of de novo gene evolution rather than birth-and-death processes (Schmid and Aquadro 2001). Many ortholog groups have been determined among *Neurospora* species, including *N. tetrasperma* and *N. discreta* as well as *Sordaria macrospora,* the species with an available genome (Nowrousian et al. 2004) most closely related to species in the genus

Neurospora (Stajich et al. 2012). Presence and absence of *Neurospora* orphan genes identified in the *N. crassa* genome were further examined in the genomes of *N. tetrasperma*, *N. discreta*, and *S. macrospora* in the FungiDB database. Species-specific genes in *N. tetrasperma* and *N. discreta* were not investigated in this study, except for some specific cases mentioned in the text. In addition, a 3-way reciprocal `blastp` and a `tblastx` were used to search for homologous genes among available genomes in NCBI. For the search for orthologous and paralogous genes, a BLAST expectation (E) value of 1×10^{-10} was used as a cutoff.

Expression and functional analyses of orphan genes—Genome-wide gene expression was investigated in *N. crassa* along multiple stages of its life cycle, including conidial germination on different media (Wang, Miguel-Rojas, et al. 2019) and production of meiotic propagules (ascospores) on synthetic crossing medium (Wang et al. 2014). Transcriptomic data of GSE41484 was reanalyzed with the latest annotation of the *N. crassa* genome. The tally for each sample was processed with `LOX` v1.6 (Zhang et al. 2010) to analyze gene expression levels across all data points. To assess environmental impacts on expression of orphan genes, recent available data on 37C bird from this lab (GSE168995) and a 240-minute time course of asexual growth in response to darkness and light stimulation (Wu et al. 2014) were revisited. To assess possible roles of orphan genes in metabolic regulation, transcriptomics data from mycelia exposed to 5 different carbon resources from crop residues (Wang et al. 2015) and from mycelia in response to non-preferred carbon sources such as furfural and 5-hydroxymethyl furfural (HMF; (Feldman et al. 2019) were also examined separately. To assess mutation in transcription factors, including

adv-1, *pp-1*, and *ada-6*, previous transcriptomics data (Fischer et al. 2018; Sun et al. 2019) were also analyzed with `LOX` v1.6 (Zhang et al. 2010).

***Model-averaged clustering of heterogeneous variation sites*—**The chromosomal distribution and clustering of orphan genes—as well as orphan genes and *het*-like genes—were analyzed with Cluster Locator (Pazos Obregón et al. 2018) using a default `Max-Gap = 5`, `1`, and `0` (`Max-Gap` ensures that the gap between adjacent genes is never larger than its parameter value). Significant clusters ($P < 0.01$) were reported. To analyze heterogeneous patterns of gene cluster variation, a vector of 0s and 1s (representing non-orphan genes and orphan genes, respectively), was generated as an input sequence for MACML, which is a powerful algorithm to profile clustering of site types (Zhang et al. 2009). This algorithm calculates all likely models of linear clustering by partitioning the entire sequence into all possible clusters and subclusters, and all models are statistically evaluated for information-optimality via Akaike Information Criterion (Akaike 1974), 'corrected' Akaike Information criterion (Hurvich and Tsai 1989), or Bayesian Information Criterion (Raftery et al. 1997). In this study, the weighted likelihood was computed based on the conservative Bayesian Information Criterion for each model. All estimates of 'orphan-gene cluster probability' as well as ninety-five percent confidence intervals of the probability were calculated as the weighted averages across all models.

***Knockout strains and phenotype identification*—**Knockout strains for more than 9600 genes (Colot et al. 2006), including deletion cassettes for genes in either of the two mating types, were acquired from the Fungal Genetic Stock Center (FGSC: McCluskey et al. 2010). Identified

Neurospora orphan genes were examined for altered phenotypes during conidia germination on Bird Medium (BM) and sexual development on Synthetic Crossing Medium (SCM) from protoperithecium differentiation to ascospore release. Genotype *mat A* strains were assayed for phenotypes when available; otherwise, *mat a* strains were used. All available KO strains were phenotyped on BM and on SCM with three replicates. For each investigated strain, 3000–5000 conidia were plated onto 90 mm diameter plates and monitored, and crossing was conducted between opposite mating types. Three independent phenotyping experiments were performed with each knockout strain using stored conidia supplied by the FGSC.

Acknowledgements: We thank the Broad Institute, FungiDB and JGI for making *Neurospora* related fungi genomic data available. All authors have declared that no competing interests exist. This study was supported by funding to JPT from the National Institutes of Health R01 grant AI146584, by the National Science Foundation grant IOS 1457044 to JPT, National Science Foundation IOS 1456482 to FT, and NSF-BSF-2018712 to OY, ZW, and JPT. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

References cited

Akaike H. 1974. A New Look at the Statistical Model Identification. *Springer Series in Statistics* [Internet]:215–222. Available from: http://dx.doi.org/10.1007/978-1-4612-1694-0_16

Ament-Velásquez SL, Vogan AA, Granger-Farbos A, Bastiaans E, Martinossi-Allibert I, Saupe SJ, de Groot S, Lascoux M, Debets AJM, Clavé C, et al. 2022. Allorecognition genes drive reproductive isolation in Podospora anserina. *Nat Ecol Evol* [Internet]. Available from: http://dx.doi.org/10.1038/s41559-022-01734-x

Arnold R, Rattei T, Tischler P, Truong M-D, Stümpflen V, Mewes W. 2005. SIMAP—The similarity matrix of proteins. *Bioinformatics* 21:ii42–ii46.

Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for de Novo Evolution of Testis-Expressed Genes in the Drosophila yakuba/Drosophila erecta Clade. *Genetics* [Internet] 176:1131–1137. Available from: http://dx.doi.org/10.1534/genetics.106.069245

Begun DJ, Lindfors HA, Thompson ME, Holloway AK. 2006. Recently Evolved Genes Identified FromDrosophila yakubaandD. erectaAccessory Gland Expressed Sequence Tags. *Genetics* [Internet] 172:1675–1681. Available from: http://dx.doi.org/10.1534/genetics.105.050336

Bobrowicz P, Pawlak R, Correa A, Bell-Pedersen D, Ebbole DJ. 2002. The Neurospora crassa pheromone precursor genes are regulated by the mating type locus and the circadian clock. *Molecular Microbiology* [Internet] 45:795–804. Available from: http://dx.doi.org/10.1046/j.1365-2958.2002.03052.x

Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, Seiler S, Bell-Pedersen D, Paietta J, Plesofsky N, et al. 2004. Lessons from the genome sequence of Neurospora crassa: tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* 68:1–108.

Cai JJ, Woo PCY, Lau SKP, Smith DK, Yuen K-Y. 2006. Accelerated evolutionary rate may be responsible for the emergence of lineage-specific genes in ascomycota. *J. Mol. Evol.* 63:1–11.

Carrillo AJ, Cabrera IE, Spasojevic MJ, Schacht P, Stajich JE, Borkovich KA. 2020. Clustering analysis of large-scale phenotypic data in the model filamentous fungus Neurospora crassa. *BMC Genomics* 21:755.

Casola C. 2018. From de novo to “de nono”: The majority of novel protein coding genes identified with phylostratigraphy are old genes or recent duplicates. *Genome Biology and Evolution* [Internet]. Available from: http://dx.doi.org/10.1093/gbe/evy231

Chen C-H, Ringelberg CS, Gross RH, Dunlap JC, Loros JJ. 2009. Genome-wide analysis of light-inducible responses reveals hierarchical light signalling in Neurospora. *EMBO J.* 28:1029–1042.

Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC. 2006. A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 103:10352–10357.

Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C, Glass NL. 2012. Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc. Natl. Acad. Sci. U. S. A.* 109:7397–7402.

Coradetti ST, Xiong Y, Glass NL. 2013. Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in Neurospora crassa. *Microbiologyopen* 2:595–609.

Corcoran P, Anderson JL, Jacobson DJ, Sun Y, Ni P, Lascoux M, Johannesson H. 2016. Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus Neurospora tetrasperma. *Genome Res.* 26:486–498.

Craig JP, Coradetti ST, Starr TL, Glass NL. 2015. Direct target network of the Neurospora crassa plant cell wall deconstruction regulators CLR-1, CLR-2, and XLR-1. *MBio* 6:e01452–15.

Davis RH, Perkins DD. 2002. Neurospora: a model of model microbes. *Nature Reviews Genetics* [Internet] 3:397–403. Available from: http://dx.doi.org/10.1038/nrg797

Dekhang R, Wu C, Smith KM, Lamb TM, Peterson M, Bredeweg EL, Ibarra O, Emerson JM, Karunarathna N, Lyubetskaya A, et al. 2017. The Neurospora Transcription Factor ADV-1 Transduces Light Signals and Temporal Information to Control Rhythmic Expression of Genes Involved in Cell Fusion. *G3* 7:129–142.

Domazet-Loso T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 23:533–539.

Dornburg A, Su Z, Townsend JP. 2019. Optimal Rates for Phylogenetic Inference and Experimental Design in the Era of Genome-Scale Data Sets. *Syst. Biol.* 68:145–156.

Ebbole D. 2000. Neurospora: a new (?) model system for microbial genetics. *Trends in Genetics* [Internet] 16:291–292. Available from: http://dx.doi.org/10.1016/s0168-9525(00)02040-0

Eilers FI, Sussman AS. 1970. Conversion of furfural to furoic acid and furfuryl alcohol by Neurospora ascospores. *Planta* 94:253–264.

Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc. Natl.*

Acad. Sci. U. S. A. 108:2831–2836.

Emerson MR. 1948. Chemical Activation of Ascospore Germination in Neurospora crassa. *J. Bacteriol.* 55:327–330.

Feldman D, Kowbel DJ, Cohen A, Louise Glass N, Hadar Y, Yarden O. 2019. Identification and manipulation of Neurospora crassa genes involved in sensitivity to furfural. *Biotechnology for Biofuels* [Internet] 12. Available from: http://dx.doi.org/10.1186/s13068-019-1550-4

Fischer MS, Wu VW, Lee JE, O'Malley RC, Glass NL. 2018. Regulation of Cell-to-Cell Communication and Cell Wall Integrity by a Network of MAP Kinase Pathways and Transcription Factors in. *Genetics* 209:489–506.

Fu C, Iyer P, Herkal A, Abdullah J, Stout A, Free SJ. 2011. Identification and characterization of genes required for cell-to-cell fusion in Neurospora crassa. *Eukaryot. Cell* 10:1100–1109.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus Neurospora crassa. *Nature* 422:859–868.

Galagan JE, Selker EU. 2004. RIP: the evolutionary cost of genome defense. *Trends Genet.* 20:417–423.

Gearing LJ, Cumming HE, Chapman R, Finkel AM, Woodhouse IB, Luu K, Gould JA, Forster SC, Hertzog PJ. 2019. CiiiDER: A tool for predicting and analysing transcription factor binding sites. *PLoS One* 14:e0215495.

Gladieux P, De Bellis F, Hann-Soden C, Svedberg J, Johannesson H, Taylor JW. 2020. Neurospora from Natural Populations: Population Genomics Insights into the Life History of a Model Microbial Eukaryote. In: Dutheil JY, editor. Statistical Population Genomics. New York, NY: Springer US. p. 313–336.

Gladyshev E. 2017. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. *Microbiol Spectr* [Internet] 5. Available from: http://dx.doi.org/10.1128/microbiolspec.FUNK-0042-2017

Greenwald CJ, Kasuga T, Glass NL, Shaw BD, Ebbole DJ, Wilkinson HH. 2010. Temporal and spatial regulation of gene expression during asexual development of Neurospora crassa. *Genetics* 186:1217–1230.

Grigoriev IV. 2011. JGI Fungal Genomics Program. Available from: http://dx.doi.org/10.2172/1012482

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42:D699–D704.

Haridas S, Salamov A, Grigoriev IV. 2018. Fungal Genome Annotation. *Methods Mol. Biol.* 1775:171–184.

Herold I, Kowbel D, Delgado-Álvarez DL, Garduño-Rosales M, Mouriño-Pérez RR, Yarden O. 2019. Transcriptional profiling and localization of GUL-1, a COT-1 pathway component, in Neurospora crassa. *Fungal Genet. Biol.* 126:1–11.

Herold I, Zolti A, Garduño-Rosales M, Wang Z, López-Giráldez F, Mouriño-Pérez RR, Townsend JP, Ulitsky I, Yarden O. 2021. The GUL-1 Protein Binds Multiple RNAs Involved in Cell Wall Remodeling and Affects the MAK-1 Pathway in Neurospora crassa. *Frontiers in Fungal Biology* [Internet] 2. Available from: http://dx.doi.org/10.3389/ffunb.2021.672696

Hurvich CM, Tsai C-L. 1989. Regression and time series model selection in small samples. *Biometrika* 76:297–307.

Jacobson DJ. 1992. Control of mating type heterokaryon incompatibility by the tol gene in Neurospora crassa and N. tetrasperma. *Genome* [Internet] 35:347–353. Available from: http://dx.doi.org/10.1139/g92-053

Jamieson K, McNaught KJ, Ormsby T, Leggett NA, Honda S, Selker EU. 2018. Telomere repeats induce domains of H3K27 methylation in Neurospora. *Elife* [Internet] 7. Available from: http://dx.doi.org/10.7554/eLife.31216

Káldi K, González BH, Brunner M. 2006. Transcriptional regulation of the Neurospora circadian clock gene wc-1 affects the phase of circadian output. *EMBO Rep.* 7:199–204.

Kasuga T, Glass NL. 2008. Dissecting colony development of Neurospora crassa using mRNA profiling and comparative genomics approaches. *Eukaryot. Cell* 7:1549–1564.

Kasuga T, Mannhaupt G, Glass NL. 2009. Relationship between phylogenetic distribution and genomic features in Neurospora crassa. *PLoS One* 4:e5286.

Kritsky MS, Belozerskaya TA, Sokolovsky VY, Filippovich SY. 2005. Photoreceptor Apparatus of the Fungus Neurospora crassa. *Mol. Biol.* 39:514–528.

Lan N, Ye S, Hu C, Chen Z, Huang J, Xue W, Li S, Sun X. 2021. Coordinated Regulation of Protoperithecium Development by MAP Kinases MAK-1 and MAK-2 in. *Front. Microbiol.* 12:769615.

Leeder AC, Jonkers W, Li J, Glass NL. 2013. Early colony establishment in Neurospora crassa requires a MAP kinase regulatory network. *Genetics* 195:883–898.

Lehr NA, Wang Z, Li N, Hewitt DA, López-Giráldez F, Trail F, Townsend JP. 2014. Gene expression differences among three Neurospora species reveal genes required for sexual reproduction in Neurospora crassa. *PLoS One* 9:e110398.

McCluskey K, Wiest A, Plamann M. 2010. The Fungal Genetics Stock Center: a repository for 50 years of fungal genetics research. *J. Biosci.* 35:119–126.

McLysaght A, Hurst LD. 2016. Open questions in the study of de novo genes: what, how and why. *Nature Reviews Genetics* [Internet] 17:567–578. Available from: http://dx.doi.org/10.1038/nrg.2016.78

Mir-Rashed N, Jacobson DJ, Dehghany MR, Micali OC, Smith ML. 2000. Molecular and functional analyses of incompatibility genes at het-6 in a population of Neurospora crassa. *Fungal Genet. Biol.* 30:197–205.

Mitchell MB. 1965. A Model Predicting Characteristics of Genetic Maps in Neurospora crassa. *Nature* [Internet] 205:680–682. Available from: http://dx.doi.org/10.1038/205680a0

Newmeyer D. 1970. A suppressor of the heterokaryon-incompatibility associated with mating type in Neurospora crassa. *Can. J. Genet. Cytol.* 12:914–926.

Nowrousian M, Würtz C, Pöggeler S, Kück U. 2004. Comparative sequence analysis of Sordaria macrospora and Neurospora crassa as a means to improve genome annotation. *Fungal Genetics and Biology* [Internet] 41:285–292. Available from: http://dx.doi.org/10.1016/j.fgb.2003.10.005

Pazos Obregón F, Soto P, Lavín JL, Cortázar AR, Barrio R, Aransay AM, Cantera R. 2018. Cluster Locator, online analysis and visualization of gene clustering. *Bioinformatics* 34:3377–3379.

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* 96:4285–4288.

Pöggeler S, Kück U. 2000. Comparative analysis of the mating-type loci from Neurospora crassa and Sordaria macrospora: identification of novel transcribed ORFs. *Mol. Gen. Genet.* 263:292–301.

Raftery AE, Madigan D, Hoeting JA. 1997. Bayesian Model Averaging for Linear Regression Models. *J. Am. Stat. Assoc.* 92:179–191.

Rattei T, Arnold R, Tischler P, Lindner D, Stümpflen V, Mewes HW. 2006. SIMAP: the similarity matrix of proteins. *Nucleic Acids Res.* 34:D252–D256.

Rodriguez S, Ward A, Reckard AT, Shtanko Y, Hull-Crew C, Klocko AD. 2022. The genome organization of Neurospora crassa at high resolution uncovers principles of fungal chromosome topology. *G3* [Internet] 12. Available from: http://dx.doi.org/10.1093/g3journal/jkac053

Ruiz-Orera J, Hernandez-Rodriguez J, Chiva C, Sabidó E, Kondova I, Bontrop R, Marqués-Bonet T, Albà MM. 2015. Origins of De Novo Genes in Human and Chimpanzee.

PLoS Genet. 11:e1005721.

Saupe SJ, Kuldau GA, Smith ML, Glass NL. 1996. The product of the het-C heterokaryon incompatibility gene of Neurospora crassa has characteristics of a glycine-rich cell wall protein. *Genetics* 143:1589–1600.

Schmid KJ, Aquadro CF. 2001. The evolutionary analysis of "orphans" from the Drosophila genome identifies rapidly diverging and incorrectly annotated genes. *Genetics* [Internet]. Available from: https://www.genetics.org/content/159/2/589.short

Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, et al. 2012. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res.* 40:D675–D681.

Sun X, Wang F, Lan N, Liu B, Hu C, Xue W, Zhang Z, Li S. 2019. The Zn(II)2Cys6-Type Transcription Factor ADA-6 Regulates Conidiation, Sexual Development, and Oxidative Stress Response in Neurospora crassa. *Frontiers in Microbiology* [Internet] 10. Available from: http://dx.doi.org/10.3389/fmicb.2019.00750

Sun X, Yu L, Lan N, Wei S, Yu Y, Zhang H, Zhang X, Li S. 2012. Analysis of the role of transcription factor VAD-5 in conidiation of Neurospora crassa. *Fungal Genet. Biol.* 49:379–387.

Su Z, Townsend JP. 2015. Utility of characters evolving at diverse rates of evolution to resolve quartet trees with unequal branch lengths: analytical predictions of long-branch effects. *BMC Evol. Biol.* 15:86.

Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12:692–702.

Trail F, Wang Z, Stefanko K, Cubba C, Townsend JP. 2017. The ancestral levels of transcription and the evolution of sexual phenotypes in filamentous fungi. *PLoS Genet.* 13:e1006867.

Vakirlis N, Carvunis A-R, McLysaght A. 2020. Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. *eLife* [Internet]. Available from: http://dx.doi.org/10.1101/735175

Wang B, Cai P, Sun W, Li J, Tian C, Ma Y. 2015. A transcriptomic analysis of Neurospora crassa using five major crop residues and the novel role of the sporulation regulator rca-1 in lignocellulase production. *Biotechnol. Biofuels* 8:21.

Wang Z, Gudibanda A, Ugwuowo U, Trail F, Townsend JP. 2018. Using evolutionary genomics, transcriptomics, and systems biology to reveal gene networks underlying fungal development. *Fungal Biology Reviews* [Internet] 32:249–264. Available from: http://dx.doi.org/10.1016/j.fbr.2018.02.001

Wang Z, Li N, Li J, Dunlap JC, Trail F, Townsend JP. 2016. The Fast-Evolving phy-2 Gene

Modulates Sexual Development in Response to Light in the Model Fungus Neurospora crassa. *MBio* 7:e02148.

Wang Z, Lopez-Giraldez F, Lehr N, Farré M, Common R, Trail F, Townsend JP. 2014. Global gene expression and focused knockout analysis reveals genes associated with fungal fruiting body development in Neurospora crassa. *Eukaryot. Cell* 13:154–169.

Wang Z, López-Giráldez F, Wang J, Trail F, Townsend JP. 2019. Integrative Activity of Mating Loci, Environmentally Responsive Genes, and Secondary Metabolism Pathways during Sexual Development of Chaetomium globosum. *MBio* [Internet] 10. Available from: http://dx.doi.org/10.1128/mBio.02119-19

Wang Z, Miguel-Rojas C, Lopez-Giraldez F, Yarden O, Trail F, Townsend JP. 2019. Metabolism and Development during Conidial Germination in Response to a Carbon-Nitrogen-Rich Synthetic or a Natural Source of Nutrition in Neurospora crassa. *MBio* [Internet] 10. Available from: http://dx.doi.org/10.1128/mBio.00192-19

Wang Z, Wang J, Li N, Li J, Trail F, Dunlap JC, Townsend JP. 2018. Light sensing by opsins and fungal ecology: NOP-1 modulates entry into sexual reproduction in response to environmental cues. *Mol. Ecol.* 27:216–232.

Weisman CM, Murray AW, Eddy SR. 2020. Many but not all lineage-specific genes can be explained by homology detection failure. *PLoS Biology* 18:e3000862.

Wu C, Kim Y-S, Smith KM, Li W, Hood HM, Staben C, Selker EU, Sachs MS, Farman ML. 2009. Characterization of chromosome ends in the filamentous fungus Neurospora crassa. *Genetics* 181:1129–1145.

Wu C, Yang F, Smith KM, Peterson M, Dekhang R, Zhang Y, Zucker J, Bredeweg EL, Mallappa C, Zhou X, et al. 2014. Genome-wide characterization of light-regulated genes in Neurospora crassa. *G3* 4:1731–1745.

Xiang Q, Louise Glass N. 2004. The control of mating type heterokaryon incompatibility by vib-1, a locus involved in het-c heterokaryon incompatibility in Neurospora crassa. *Fungal Genetics and Biology* [Internet] 41:1063–1076. Available from: http://dx.doi.org/10.1016/j.fgb.2004.07.006

Xiong Y, Wu VW, Lubbe A, Qin L, Deng S, Kennedy M, Bauer D, Singan VR, Barry K, Northen TR, et al. 2017. A fungal transcription factor essential for starch degradation affects integration of carbon and nitrogen metabolism. *PLoS Genet.* 13:e1006737.

Zámocký M, Tafer H, Chovanová K, Lopandic K, Kamlárová A, Obinger C. 2016. Genome sequence of the filamentous soil fungus Chaetomium cochliodes reveals abundance of genes for heme enzymes from all peroxidase and catalase superfamilies. *BMC Genomics* 17:763.

Zhang L, Ren Y, Yang T, Li G, Chen J, Gschwend AR, Yu Y, Hou G, Zi J, Zhou R, et al. 2019.

Rapid evolution of protein diversity by de novo origination in Oryza. *Nature Ecology & Evolution* [Internet] 3:679–690. Available from: http://dx.doi.org/10.1038/s41559-019-0822-5

Zhang N, Wang Z. 2015. 3 Pezizomycotina: Sordariomycetes and Leotiomycetes. *Systematics and Evolution* [Internet]:57–88. Available from: http://dx.doi.org/10.1007/978-3-662-46011-5_3

Zhang, Zhang Z, Townsend J. 2009. Maximum-Likelihood Model Averaging To Profile Clustering of Site Types across Discrete Linear Sequences. *SciVee* [Internet]. Available from: http://dx.doi.org/10.4016/12250.01

Zhang Z, López-Giráldez F, Townsend JP. 2010. LOX: inferring Level Of eXpression from diverse methods of census sequencing. *Bioinformatics* 26:1918–1919.

Zhao J, Gladieux P, Hutchison E, Bueche J, Hall C, Perraudeau F, Glass NL. 2015. Identification of Allorecognition Loci in Neurospora crassa by Genomics and Evolutionary Approaches. *Mol. Biol. Evol.* 32:2417–2432.

Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, Cate JHD, Glass NL. 2012. Induction of lignocellulose-degrading enzymes in Neurospora crassa by cellodextrins. *Proc. Natl. Acad. Sci. U. S. A.* 109:6012–6017.