The Genome Sequence of the Filamentous Fungus *Neurospora crassa*

James E. Galagan[1], Sarah E. Calvo[1], Katherine A. Borkovich[2], Eric U. Selker[3], Nick D. Read[4], David Jaffe[1], William FitzHugh[5], Li-Jun Ma[1], Serge Smirnov[1], Seth Purcell[1], Bushra Rehman[1], Timothy Elkins[1], Reinhard Engels[1], Shunguang Wang[1], Cydney B. Nielsen[1], Jonathan Butler[1], Matthew Endrizzi[1], Dayong Qui[1], Peter Ianakiev[1], Deborah Bell-Pedersen[6], Mary Anne Nelson[7], Margaret Werner-Washburne[7], Claude P. Selitrennikoff[8], John A. Kinsey[10], Edward L. Braun[11], Alex Zelter[4,26], Ulrich Schulte[12], Gregory O. Kothe[3], Gregory Jedd[13], Werner Mewes[9,17], Chuck Staben[14], Edward Marcotte[15], David Greenberg[16], Alice Roy[1], Karen Foley[1], Jerome Naylor[1], Nicole Stange-Thomann[1], Robert Barrett[1], Sante Gnerre[1], Michael Kamal[1], Manolis Kamvysselis[1], Cord Bielke[9], Stephen Rudd[17], Dmitrij Frishman[17], Svetlana Krystofova[2], Carolyn Rasmussen[18], Robert L. Metzenberg[19], David D. Perkins[19], Scott Kroken[18], David Catcheside[21], Weixi Li [14], Robert J. Pratt[6], Stephen A. Osmani[23], Colin P.C. DeSouza[24], Louise Glass[18], Marc J. Orbach[25], J. Andrew Berglund[3], Rodger Voelker[3], Oded Yarden[26], Mike Plamann[27], Stephan Seiler[27], Jay Dunlap[22], Alan Radford[28], Rodolfo Aramayo[6], Donald O. Natvig[7], Lisa A. Alex[29], Gertrud Mannhaupt[9], Daniel J. Ebbole[30], Michael Freitag[3], Ian Paulsen[16], Matthew S. Sachs[31], Eric S. Lander[1,32], Chad Nusbaum[1] & Bruce Birren[1]

[1]*Whitehead Institute Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA*
[2]*Department of Plant Pathology, University of California, Riverside, CA 92521 USA*
[3]*Institute of Molecular Biology, University of Oregon, Eugene, OR 97403 USA*
[4]*Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH UK*
[5]*Celera Genomics, Rockville, MD 20850 USA*
[6]*Department of Biology, Texas A & M University, College Station, TX 77843, USA*
[7]*Department of Biology, University of New Mexico, Albuquerque, NM 87131 USA*
[8]*Department of Cellular and Structural Biology, University of Colorado Health Sciences Center, Denver, CO 80262, USA*
[9]*Department of Genome Oriented Bioinformatics, Technical University of Munich, Wissenschaftzentrum Weihenstephan, 85350 Freising-Weihenstephan, Germany*
[10]*Department of Microbiology, University of Kansas Medical School, Kansas City, KS 66160 USA*
[11]*Department of Zoology, University of Florida, Gainesville, FL 32611-8525, USA*
[12]*Institute of Biochemistry, Heinrich Heine University, 40225 Duesseldorf Germany*
[13]*Lab of Plant Molecular Biology, Rockefeller University, New York, NY 10021 US*
[14]*T.H. Morgan School of Biological Sciences, University of Kentucky, Lexington, KY 40506 USA*
[15]*The Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712 USA*
[16]*The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20878, USA*
[17]*Institute for Bioinformatics (MIPS), GSF-National Research Center for Environment and Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany*
[18]*Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720 USA*
[19]*Department of Biological Sciences, Stanford University, Stanford, CA 94305*
[21]*School of Biological Sciences, Flinders University, P.O. Box 2100, Adelaide, South Australia 5001*

[22]*Department of Genetics, Dartmouth Medical School, Hanover, NH 03755 USA*
[23]*Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210*
[24]*Department of Molecular Genetics, The Ohio State University, Columbus, OH 43211*
[25]*Department of Plant Pathology, University of Arizona, Tucson, AZ 85721 USA*
[26]*Dept. of Plant Pathology and Microbiology, Faculty of Agricultural, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem, Rehovot 76100 Israel*
[27]*School of Biological Sciences, University of Missouri-Kansas City, Kansas City, MO 64110 USA*
[28]*School of Biology, Leeds University, Leeds LS2 9JT UK*
[29]*Department of Chemistry, California State Polytechnic University Pomona, Pomona, CA 91768*
[30]*Department of Plant Pathology and Microbiology, Texas A & M University, College Station, TX 77843, USA*
[31]*Department of Biochemistry and Molecular Biology, OGI School of Science and Engineering, Oregon Health & Science University, Beaverton, OR 97006 USA*
[32]*Department of Biology, MIT, Cambridge, Massachusetts 02139, USA*

***Neurospora crassa* is a central organism in the history of 20th century genetics, biochemistry, and molecular biology. Here, we report a high-quality draft sequence of the *Neurospora* genome. The ~40 Mb genome encodes ~10,000 protein-coding genes — more than twice as many as the fission yeast *S. pombe* and only ~25% fewer than the fruit fly *D. melanogaster*. Analysis of the gene set yields insights into unexpected aspects of *Neurospora* biology including the identification of genes potentially associated with red light photobiology, genes implicated in secondary metabolism, and important differences in Ca^{2+} signalling as compared to plants and animals. *Neurospora* possesses the widest array of genome defence mechanisms known for any organism, including a process unique to fungi called Repeat Induced Point Mutation (RIP). Genome analysis suggests that RIP has had a profound impact on genome evolution, greatly slowing the creation of new genes through genomic duplication and resulting in a genome with an unusually low proportion of closely related genes.**

Research on *Neurospora* in the early part of the 20th century paved the way for modern genetics and molecular biology. First documented in 1843 as a contaminant of bakeries in Paris[1], *Neurospora* was developed as an experimental organism in the 1920s[2,3]. Subsequent work on *Neurospora* by Beadle and Tatum[4] in the 1940s established the relationship between genes and proteins, summarized in the "one-gene-one-enzyme" hypothesis. In the latter half of the century, *Neurospora* played a key role as a model organism, contributing to the fundamental understanding of genome defence systems, DNA methylation, mitochondrial protein import, circadian rhythms, post-transcriptional gene silencing, and DNA repair[5]. Because *Neurospora* is a multicellular filamentous fungus, it has also provided a system to study cellular differentiation and development as well as other aspects of eukaryotic biology[6].

The legacy of over 70 years of research[7], coupled with the availability of molecular and genetic tools, offers enormous potential for continued discovery. The sequencing of the *Neurospora crassa* genome was undertaken to maximize this potential. Here, we report an initial sequence and analysis of the *Neurospora* genome.

Neurospora Genome Sequence

The *Neurospora* genome is much larger (>40 Mb) than that of *S. pombe* and *S. cerevisiae* (both ~12 Mb). Accordingly, we sought to first produce and analyse a high-quality draft sequence, en route to a finished sequence.

The genome sequence was assembled from deep whole-genome shotgun (WGS) coverage obtained by sequencing both ends of a variety of clone types, including plasmids, Fosmids, bacterial artificial chromosomes and jumping clones (Table 1 and Methods). In all, the data provided an average of >20-fold sequence coverage and >98-fold physical coverage of the genome. The Arachne package[8] was used to assemble the draft genome sequence. The resulting assembly consists of 958 sequence contigs with a total length of 38.6 Mb (Table 2) and an N50 length of 114.5 kb (that is, 50% of all bases are contained in contigs of at least 114.5 kb). The sequence contigs are assembled into scaffolds (or supercontigs) when they are joined by at least two forward-reverse links arising from the paired-end reads. The

assembly consists of 163 scaffolds with a total length of 39.9 Mb (including gaps between contigs) and an N50 length of 1.56 Mb.

The vast majority of the assembly (97%) is contained in the 44 largest scaffolds. At the opposite end of the spectrum, there are 38 tiny scaffolds with lengths <4 kb. Forty-two of these large scaffolds (and one of the smaller ones) could be readily anchored to the *Neurospora* genetic map[7] by virtue of their containing known genetic markers with sequence, allowing the assembly to be positioned along the seven *Neurospora* chromosomes.

The genome assembly is of very high quality in terms of continuity and representation. The assembly has long-range continuity, with the N50 scaffold size being nearly 1000-fold larger than the average gene size (see below). The assembly represents the vast majority of the genome, as assessed by comparison with available finished sequence and genetic markers. It contains 99.13% of available finished sequence (17 Mb from linkage groups II and V[9]) and all of the 252 genetic markers with sequence. This estimate, however, does not account for unusual genomic regions such as the ribosomal DNA repeats; such regions may contain ~1.7 Mb of additional sequence[10] corresponding to 2-3% of the genome that cannot be readily assembled with available techniques.

The assembly also has high accuracy, with 99.5% of the sequence having Arachne quality scores ≥30. Comparison with finished sequence confirms the sequence accuracy, with the discrepancy rate for this subset being less than 10^{-5}. The comparison also largely confirms the assembly. Twelve minor discrepancies were identified in 17 Mb. Five lie at the end of contigs and are most likely caused by a misaligned or low-quality terminal read. Four are short insertions or deletions, ranging from 9-559 bp. The remaining three discrepancies appear to be instances in which the finished sequence does not correctly represent the genome owing to chimerism in the BACs and in which the draft sequence is correct.

Efforts are currently underway to produce a complete sequence of the genome (apart from problematic regions such as the rDNA repeats). The existing high quality draft sequence, however, provides the substrate for genomic analysis.

Genes

Gene Count and Basic Characteristics. A total of 10,082 protein-coding genes (9,200 longer than 100 amino acids) were predicted (Table 2). This constitutes nearly twice as many genes as in *S. pombe* (~4800) and *S. cerevisiae (*~6300*)* and nearly as many as in [is 4300 fewer really 'nearly as many'*D. melanogaster* (~14,300). Genes span at least 44% of the genome sequence (38% in coding exons and an additional 6% in introns between such exons), with an average gene density of one gene per 3.7 kb. The average gene length of 1.67 kb, is slightly longer than the 1.4 kb average gene length for both *S. cerevisiae* and *S. pombe*. The difference in gene length is due to the greater number of introns in *Neurospora* genes -- an average of 1.7 introns per gene with an average intron size of 134 nucleotides.

Interestingly, most predicted *Neurospora* introns lack a polypyrimidine (PY) tract, which are common in other eukaryotic introns, but do contain a strong branchpoint (BP) sequence. The position of the BP sequence in *Neurospora* introns is unusual, in comparison to well-characterized human and *S. cerevisiae* introns, in being close to the 3' acceptor site. The distance to the acceptor site in *Neurospora* averages 14 nt, in contrast to an average of 31-40 nucleotides in yeast[11] and 24 nucleotides in humans[12]. The strong BP sequence and its proximity to the 3' splice site (consensus of AYAG) suggest that in *Neurospora* these elements play an important role in defining the 3' end of the intron in the absence of a PY tract.

Comparative Analysis. Comparison of the *Neurospora* predicted protein set to those of selected sequenced eukaryotes (see Methods) highlights the biological complexity of the filamentous fungi. Fully 4,140 (41%) *Neurospora* proteins lack significant matches to known proteins from public databases (Table 2), reflecting the early stage of fungal genome exploration and the diversity of fungal genes remaining to be described. Furthermore, 5,805 (57%) *Neurospora* proteins do not have significant matches to genes in either of the sequenced yeasts (Figure 1a), underscoring the greater complexity of this filamentous fungus as compared to the yeast models.

When compared to sequenced eukaryotes, a total of 1,421 (14%) *Neurospora* genes display best BLASTP matches to proteins in either plants or animals (Figure 1b). Of these, 584 lack high scoring hits to either sequenced yeast. These data reflect the biology shared between filamentous fungi and other higher eukaryotes, which in a number of cases is absent in the yeasts. These results, and others discussed below, emphasize the importance of *Neurospora* as a model organism for the study of biology in eukaryotic organisms.

Protein Domain Architectures. To characterize the functional organization of the proteome, we analysed the domain structure of the *Neurospora* protein set. Predicted proteins were searched against the Interpro database of protein families and domains using InterproScan[13]. The domain architecture for each protein was then defined as the ordered list (from N terminus to C terminus) of corresponding Interpro domains. An identical analysis was performed on selected sequenced eukaryotes for comparison.

As shown in Table 3, larger genome size is generally correlated with a greater number of distinct domains. Consistent with this, *Neurospora* possesses a number of protein domain families that display expansions in number relative to yeast (Table 4). However, the relative increase in domains between organisms is modest, with *Neurospora* displaying an 8% increase in the number of domains compared with the yeasts. In contrast, the number of unique permutations of domains increases substantially with genome size. In particular, *Neurospora* has 26% more identifiable architectures than the yeasts, with a 19% increase in the average number of architectures per domain. Finally, organisms with larger genomes also possess a larger proportion of architectures represented by more than one gene, with the single notable exception of *Neurospora* (Table 3). In contrast to the other eukaryotes considered, *Neurospora* possesses a smaller proportion of duplicated architectures than expected given its genome size. This difference may be due to the phenomenon of RIP (Repeat Induced Point Mutation, see below).

Epigenetics, Genome Defence, and Genome Evolution

Epigenetic phenomena, heritable changes in gene expression that do not involve changes to the DNA code, are widespread among organisms. *Neurospora* is an important model for the study of epigenetic phenomena, possessing a wide variety of epigenetic mechanisms and related genome defence mechanisms. The most remarkable of these phenomena is Repeat Induced Point Mutation (RIP), a process unique to fungi. The availability of a *Neurospora* genome sequence allows an analysis of RIP and its dramatic impact on genome evolution.

Repeat Induced Point Mutation. First discovered in *Neurospora*[14,15], RIP is a genome defence system that efficiently detects and mutates both copies of a sequence duplication. RIP acts during the haploid dikaryotic stage of the *Neurospora* sexual reproductive cycle, causing numerous C:G to T:A mutations within duplicated sequences. In a single passage through the sexual cycle up to 30% of the C:G pairs in duplicated sequences can be mutated, with a strong preference for C to T mutations occurring at CpA dinucleotides[16]. The pattern of mutations produces a characteristic skewing of dinucleotide frequencies that allows RIP-mutated sequences to be accurately detected[17]. RIP requires a minimal duplicated sequence length of about 400bp[18] and greater than roughly 80% sequence identity between duplicates[19], consistent with the proposal that RIP involves homologous pairing[20,21]. In addition to suffering mutations, RIP-mutated sequences are frequent targets for DNA methylation. As with mammals, DNA methylation has been shown to cause gene silencing in *Neurospora*[22]. RIP thus mutates and epigenetically silences repetitive DNA.

RIP has been proposed to act as a defence against selfish or mobile DNA[15,20]. Consistent with this hypothesis, examples of mobile elements inactivated by RIP have been reported for *Neurospora crassa* and related species[17,23,24]. However, because RIP mutation and methylation can extend beyond the bounds of duplicated sequences[25], RIP can have both mutational and epigenetic effects on neighbouring unique sequences. Furthermore, RIP acts on all duplicated sequences, including those arising from large-scale chromosomal duplications as well as gene duplications[26]. The presence of RIP thus has profound consequences for the evolution of the *Neurospora* genome. Indeed, it has been proposed that RIP might prevent gene innovation through gene duplication[15]. With the availability of the *Neurospora* genome sequence, we were able to address this hypothesis.

Multigene Families. To investigate the impact of RIP on protein families in *Neurospora,* genes were clustered into "multigene families" based on an all-vs.-all comparison of protein sequences (see Methods). As shown in Figure 2, the percentage of genes in multigene families in selected sequenced eukaryotes is correlated with genome size. However, in striking contrast to the other analysed organisms, *Neurospora* possesses many fewer genes in multigene families than expected given its size. When the analysis is expanded to include an additional 17 sequenced prokaryotes, only *Mycoplasma genitalium*, *Mycoplasma pulminus*, *Ureaplasma urealyticum* and *Vibrio cholerae* display a correspondingly small proportion of genes in families. This is noteworthy considering that the Mycoplasmas are thought to have undergone reductive evolution and represent minimal life forms[27].

The analysis reveals another characteristic of *Neurospora* gene families. Unlike other sequenced eukaryotes, *Neurospora* possesses only a handful of highly similar gene pairs. Figure 3 displays histograms of amino acid and nucleotide similarities between each gene in the six organisms analysed and the best matching gene in that organism. A significant proportion of genes have best matches with greater than 80% amino acid and nucleotide identity in all the organisms considered except *Neurospora*. *Neurospora* contains only 8 genes with top matches of greater than 80% amino acid or coding sequence identity. This value is significant because, as described above, RIP mutates duplicated sequences that display greater than 80% nucleotide similarity. Thus, the small proportion of genes in multigene families and the near absence of highly similar genes are consistent with the actions of RIP.

A striking example of the lack of highly similar genes in multigene families is revealed in an analysis of predicted major facilitator superfamily (MFS) sugar transporters (Figure 4). *Neurospora* has a similar number of predicted MFS sugar transporters as *S. cerevisiae*. However, a phylogenetic analysis of fungal sugar transporters indicates that the *Neurospora* proteins are substantially more divergent than those of *S. cerevisiae* as well as those of *S. pombe*. Furthermore, the *Neurospora* transporters contain no apparent instances of recent gene duplication. In contrast, the majority of *S. cerevisiae* HXT hexose and *S. pombe* GHT transporters represent two relatively recent and independent expansions and include very recently duplicated genes such as HXT15 and HXT16. Thus, despite a diversity of MFS sugar transporters, *Neurospora* appears to lack close paralogs in this gene family, consistent with the results of the genome-wide multigene family analysis. Analyses of other gene families yielded similar results (data not shown).

The paucity of closely related sequences is evident not only at the level of complete genes, but even at the level of individual exons and protein domains. Analysis of the 27,200 predicted exons in *Neurospora* revealed only 36 pairs of exons (0.1%) with greater than 80% nucleotide identity. Furthermore, an analysis of 4,047 PFAM protein domains predicted by the HMMER program[28] revealed only 37 pairs (0.9% of predicted domains) with greater than 80% amino acid identity over more than 50% the length of the predicted domain. This compares with 16% of predicted domains in *S. cerevisiae* and 10% in *S. pombe*.

Gene Evolution through Gene Duplication. The above results suggest that RIP has had a powerful impact on suppressing the creation of new genes or partial genes through genomic duplication. This is consistent with the large number of mutations induced in duplicated sequences by RIP. Indeed, computer simulation (see Methods) indicates that, following a gene duplication, each copy has an 80% probability of acquiring an in-frame stop codon after only a single round of RIP and a 99.5% probability by the point that RIP has mutated the copies to less than 85% nucleotide similarity. The high frequency of stop codons reflects the preference of RIP for mutating CpA to TpA, increasing the prevalence of the stop codons TAA and TAG.

These results raise the critical question of whether any significant gene duplication has occurred in *Neurospora* subsequent to the acquisition of RIP. We searched for empirical evidence of duplicated genes that have survived RIP by analysing the set of *Neurospora* coding sequences using two different previously developed measures[17] for detecting RIP-mutated sequences (see Methods). These measures utilize the characteristic skewing of dinucleotides produced by RIP to detect mutated sequences. According to these measures, only 59 of the 9,200 predicted genes encoding proteins ≥ 100 amino acids show evidence of having been mutated by RIP. Of these, only 8 consist of pairs of predicted duplicated genes (genes in the same multigene family) in which both copies are predicted to be RIP-mutated. Thus, few pairs of duplicated genes display evidence of having both survived RIP. The remaining genes predicted to be RIP-mutated include 3 genes that are in multigene families but lack a matching RIP-mutated gene, and 48 single-copy genes. Of the latter, 18 are situated in or adjacent to RIP-mutated relics of mobile elements and were thus potentially mutated as a consequence of this association.

Gene duplication is thought to play a primary role in the innovation of new genes[29-31]. However, taken together, these data support the conclusion that most, if not all, paralogous genes in *Neurospora* duplicated and diverged prior to the emergence of RIP, and since that point the evolution of new genes through gene duplication has been virtually arrested. This remarkable conclusion raises the question of whether and how *Neurospora* is able to evolve new genes. A number of mechanisms not involving gene duplication are conceivable, including gene sharing and lateral transfer. Ultimately, a conclusive analysis of these and other potential mechanisms of evolution may only be possible by comparing *Neurospora* to the genomes of closely related species to illuminate recent evolutionary history. Nonetheless, our results make clear that the cost to *Neurospora* for increased genome security through RIP is a significant impact on the evolution of new gene functions through gene duplication.

Repetitive DNA. The lack of genic redundancy in the *Neurospora* predicted proteome is mirrored by a relative paucity of repetitive DNA in the genome. An analysis of repeat sequences longer than 200 bp and with greater than 65% similarity (see Methods), revealed that only 10% of the *Neurospora* assembly consists of repeat sequences, consistent with previously reported estimates[32,33].

The repeat content of *Neurospora* provides a testament to the efficiency of RIP. Applying the measures of RIP mentioned above to the *Neurospora* genome revealed that the vast majority of repetitive sequences (81%) in *Neurospora* have been mutated by RIP. Conversely, only 18% of predicted RIP-mutated sequence is non-repetitive, likely reflecting the loss of the corresponding duplicated sequence. As described above, duplications greater than about 400 bp are susceptible to RIP[18]. In keeping with this, we observe that over 97% of genomic repeats greater than 400 bp in length are RIP-mutated. Moreover, repeats longer than 400 bp clustered based on sequence similarity display an average sequence identity within clusters of 78%, with 93% of repeat clusters displaying an average identity of less than 85%. This corresponds to previous estimates indicating that RIP requires greater than approximately 80% sequence identity to detect duplicated sequences.

Consistent with the hypothesis that RIP acts as a defence mechanism against selfish DNA[34], no intact mobile elements were identified. Furthermore, a significant proportion (46% of repetitive nucleotides) of the *Neurospora* RIP-mutated repetitive sequence can be identified as relics of mobile elements (Figure 5). The largest proportion of identified repeat elements consists of Tad relics, followed by relics of Punt, Pogo, gypsy, and copia-like elements, all previously described for *Neurospora*[17,23,24,35-37]. Strikingly, nearly half of all identified elements appear to be relics of mobile element families not previously known for *Neurospora*. Over 25 copies of a Tc1/mariner superfamily member were identified, similar to members of the fungal impala family that have also been characterized in *Magnaporthe grisea* and *Fusarium oxysporum*[38-40]. Twenty-five RIP-mutated copies of the Pot2 transposon, previously characterized in *M. grisea*[41], and 37 retroposon relics with similarity to the gypsy-like MAGGY elements in *M. grisea*[42] were also found.

Ribosomal RNA. The only repetitive sequences known to have survived RIP in *Neurospora* are the roughly 175-200 copies[10] of the large rDNA tandem repeat containing the 17S, 5.8S, and 25S rRNA genes. As in higher eukaryotes, these tandem repeats occur within the nucleolar organizer region (NOR), and their resistance to RIP appears to stem from this localization[15]. Within the genome sequence we found several copies of the rDNA repeat in regions outside the NOR. In every case they display evidence of mutation by RIP, consistent with previous observations[15]. Thus, the sequence of the rDNA repeat does not in itself appear to confer resistance to RIP.

The 5S rRNA genes in *Neurospora* have survived RIP in a different manner. In contrast to most higher eukaryotes in which the 5S rRNA genes form tandem repeats[43-47], the 5S genes are dispersed through the genome in *Neurospora*[48]. A total of 74 copies comprising several different subtypes of 5S rDNA are dispersed through all seven chromosomes. This dispersed distribution coupled with the small size (~120 nt) of the 5S rRNA genes ensures they are not recognized by RIP.

DNA Methylation. *Neurospora* has been extensively used as a model for studying DNA methylation in eukaryotes. As with animals and plants, methylation has been shown to cause epigenetic silencing in *Neurospora*[22,49]. In contrast to mammals, methylation in *Neurospora* is not restricted to symmetric CpG sites[50,51]. In *Neurospora*, an estimated 1.5% of cytosines are methylated[52,53], and it has been suggested that nearly all DNA methylation in the *Neurospora* genome is a result of RIP[51,54-56]. The *Neurospora* genome includes two potential cytosine DNA methyltransferase genes. One, called DIM-2, is required for all known DNA methylation[57]. The other, called RID, is required for RIP and is a member of a family found thus far only in filamentous fungi[58]. Importantly, the genome sequence confirms that *Neurospora* does not have proteins bearing canonical methyl-binding domains (MBD), suggesting the existence of alternative motifs involved in recognition and maintenance of methylated DNA[59].

Plasmid reads for *Neurospora* were sequenced from libraries cloned separately in methylation-tolerant and methylation-intolerant strains of *E. coli* (Table 1). Although not intended for this purpose, the two libraries provided a basis for predicting regions of methylation in the genome by comparing the representation of regions in sequence obtained

from each library (see Methods). Testing the accuracy of such predictions, we found that 8 of 10 regions predicted to be methylated were experimentally confirmed as such. The predictions thus have good specificity. (The data does not provide high sensitivity, however. The predictions include only a minority of actually methylated regions, probably because the methylation intolerance is incomplete (see Methods). This could be overcome by sequencing of bisulfite-treated genomic DNA[60].)

The specificity of the predictions provides valuable insight into the pattern of methylation in the *Neurospora* genome. Regions predicted to be methylated show a remarkable correspondence to regions predicted to be repetitive and RIP-mutated (Figure 6). Fully 85% correspond to predicted RIP-mutated sequences. However, a small proportion (10%) of predicted methylated regions correspond to predicted non-repetitive and non-RIP-mutated sequence. In two out of ten such cases, both the methylation and the non-repetitive nature of these sequences were experimentally verified. This raises the possibility that methylation in *Neurospora* may also play non-defence roles, as has been proposed for methylation in higher organisms.

RNA Silencing. Post-transcriptional gene silencing (PTGS), or RNA silencing, is widespread among organisms[61] and is increasingly being recognized as a key switch for controlling eukaryotic gene expression[62]. For example, co-suppression in plants[63] and RNA interference (RNAi) in animals[64-66] both control gene expression through the specific degradation of homologous RNAs. RNA silencing pathways are thought to be derived from ancestral natural defence systems directed against invading nucleic acids[67]. Consistent with this, all known PTGS mechanisms share similar components including RNA-dependent RNA polymerases (RdRPs), Argonautes (related to translation initiation factors), Dicers (double stranded-specific endonucleases), as well as many other ATP-dependent DNA and ATP-dependent RNA helicases and dsRNA-binding proteins[68].

Neurospora possesses two RNA silencing pathways. The first, called quelling, silences expression of transgenes during vegetative growth. This pathway was uncovered through the study of three genes, *qde-1, qde-2* and *qde-3,* coding respectively for an RdRP, an Argonaute, and a RecQ helicase[69-74]. The second pathway, called meiotic silencing by unpaired DNA, acts during sexual reproduction[75,76]. A gene called *sad-1*, encoding an RdRP, has been identified for this pathway[76,77].

Our analysis of the *Neurospora* genome sequence uncovered several additional genes implicated in RNA silencing (Table 5). These include one RdRP, one Argonaute-like protein, and one RecQ-like helicase, as well as two Dicer-like ribonucleases (also known as CAF proteins). A phylogenetic analysis of the predicted RdRPs, Argonaute-like proteins, and Dicer-like proteins (see Methods) indicates that the *Neurospora* genes comprise two paralogous sets. One set includes the three *qde* genes and is thus predicted to correspond to the quelling pathway. The other set includes *sad-1*, and in phylogenetic trees these genes branch consistently with those of the single pathway observed in *S. pombe*[76,78]. Based on this analysis, we predict that one of the identified Dicers, *Sms-3,* is part of the meiotic silencing pathway, while the other, *dcl-2*, is part of the quelling pathway (Table 5). In addition, we

predict that the identified Argonaute, *Sms-2*, is also part of the meiotic silencing pathway. Experimental work has confirmed the roles for *Sms-2*[79] and *Sms-3*[80]. Taken together, these results support the conclusion that meiotic silencing and quelling represent two separate RNA-dependent silencing pathways. We further hypothesize that both pathways might have evolved from a single ancestral RNA silencing pathway.

Fungal Biology and Evolution

The availability of the *Neurospora* genome sequence provides an opportunity to study the genetic basis underlying the extraordinary biochemical and metabolic diversity exhibited by a filamentous fungus. Our analysis of the genome sequence has resulted in a number of surprising insights into the biology and evolution of *Neurospora* and other filamentous fungi.

Cell Signalling and Environmental Responses

Unexpected discovery of putative red light sensing genes. Blue light is an important regulator of *Neurospora* growth and development, impacting the circadian rhythm of conidiation, carotenogenesis of hyphae, and numerous facets of sexual development[81]. Although *Neurospora* photobiology has been intensively studied for more than two decades, the genome sequence has nonetheless revealed a number of previously uncharacterized sequences with similarity to blue light sensing genes, including both a cryptochrome homolog and a gene whose product contains a single PAS/LOV-type domain associated with light sensing.

In addition, *Neurospora* possesses two putative homologs of phytochromes, genes known for their role in red light sensing in other organisms, and a putative homolog of the *Aspergillus nidulans velvet* gene implicated in the regulation of both red and blue light responses. The presence of these genes is unexpected given that no red light photobiology has been described for *Neurospora* to date. It has been recently shown that in addition to red light sensing, some *Arabidopsis* phytochromes associate with cryptochromes to play a role in blue light sensing and signaling[82]. Therefore, the two phytochromes and the *velvet* homolog may also regulate this aspect of *Neurospora* photobiology.

Expansion of histidine kinases reflects the importance of two-component signalling in filamentous fungi. Mitogen-activated protein kinase (MAPK) pathways[83] integrate signals from multiple receptor pathways including two-component signalling systems[84]. The basic two-component system consists of a histidine kinase and a cognate response regulator protein. The nine MAPK pathway proteins identified in the *Neurospora* genome sequence (Figure 7) correspond to those found in *S. pombe* and *S. cerevisiae*, indicating that the basic MAPK machinery is conserved between these species. In contrast, *Neurospora* has a significantly expanded complement of histidine kinases as compared to both sequenced yeasts. *Neurospora* possesses 11 histidine kinases, compared with one in *S. cerevisiae* and three in *S. pombe*. Two of the 11 genes have been previously characterized in *Neurospora*[85], while a third is similar to proteins in *A. fumigatus* and *A. nidulans* that affect conidiation[85,86].

Functions for the remaining genes are unknown, although 7 (including the two phytochromes discussed above) contain PAS/PAC domains, implicating them in oxygen and light responses. The expansion of histidine kinases in *Neurospora* relative to yeasts, coupled with the conservation of the core MAPK pathway, suggests a larger role than previously expected for two-component signalling in filamentous fungi. The large number of histidine kinases reveals filamentous fungi to be more similar in this regard to plants, where two component systems are abundant, than to animals, where these systems are absent.

A novel family G protein coupled receptors. Eukaryotic cells sense many environmental stimuli via seven-transmembrane helix, G protein-coupled receptors (GPCRs)[87]. Our analysis indicates that *Neurospora* possesses a total of 10 predicted seven-transmembrane helix proteins (Figure 7) three of which belong to a novel class not previously identified in any fungus. These genes encode proteins similar to cAMP GPCRs from the protists *D. discoideum*[88] and *Polysphodylium pallidum*, and also to predicted proteins from *A. thaliana*[89] and *C. elegans*. The *D. discoideum* proteins have been shown to sense cAMP levels during chemotaxis and multicellular development[90]. This suggests a possible analogous function in *Neurospora*. The existence of an extracellular cAMP signalling pathway has never been previously demonstrated in any fungal system.

In support of this hypothesis, along with the presence of putative cAMP receptors, *Neurospora* was found to possess the full complement of proteins required for the synthesis and degradation of cAMP. For example, one adenylyl cyclase[91], one cyclase-associated protein, and four genes encoding putative cAMP phosphodiesterases are present. Furthermore, *Neurospora* wild-type strains accumulate cAMP in the extracellular medium[92], although a role in extracellular signalling has not been established. These data, coupled with the presence of GPCRs similar to slime mould cAMP receptors, suggest the possibility that cAMP or a related molecule may serve as an extracellular signal in *Neurospora*.

Numerous calcium signalling proteins support the importance of Ca^{2+} sensory transduction in filamentous fungi. A considerable body of evidence, primarily from pharmacological studies, indicates that Ca^{2+}-signaling regulates numerous processes in filamentous fungi including secretion, cytoskeletal organization, hyphal tip growth, hyphal branching, sporulation, and circadian rhythmicity[93]. However, the identification of the main components of even one Ca^{2+}-mediated response pathway in filamentous fungi has remained elusive. The genome sequence of *Neurospora* has provided for the first time over 25 of the proteins likely to be necessary for Ca^{2+}-signaling in filamentous fungi (Figure 7). The presence of the genes encoding these proteins underscores the importance of Ca^{2+}-signalling for filamentous fungi.

A surprising difference between Ca^{2+}-signalling in *Neurospora* as compared with plants and animals was also revealed by the genome sequence. An important aspect of Ca^{2+}-signalling in plant and animal cells involves Ca^{2+} release from internal stores. This is commonly mediated by the second messengers inositol 1,4,5 trisphosphate ($InsP_3$) and cADP ribose, or by Ca^{2+}-induced Ca^{2+} release (although other second messenger systems also exist)[94]. $InsP_3$ is present within *Neurospora* hyphae[95] and physiological evidence, including intracellular membrane associated $InsP_3$-activated Ca^{2+} channel activity, supports a role in

Ca^{2+}-signaling[96,97]. In spite of this, *Neurospora* (and *S. cerevisiae*) lacks recognizable $InsP_3$ receptors. In addition, neither ADP ribosyl cyclase (which synthesizes cADP ribose) nor ryanodine receptor proteins, key components of Ca^{2+} release mechanisms in plant and animal cells, are found in *Neurospora*. These observations raise the question of whether other, perhaps novel, second messenger systems responsible for Ca^{2+} release from internal stores remain to be discovered in filamentous fungi.

Growth and Development

Hyphal growth. True hyphae produced by filamentous fungi are tubular structures consisting of cellular compartments usually delineated by incomplete septa[5]. In contrast, the pseudohyphae produced by yeasts such as *S. cerevisiae* consist of chains of uninucleate elongated cells[98] with no apparent cytoplasmic continuity between cellular compartments. The molecular mechanisms underlying these two modes of growth are not well understood.

The two parallel signalling pathways that regulate pseudohyphal growth in *S. cerevisiae* - the MAPK and cAMP modules - are both conserved in the *Neurospora* genome. In *C. albicans*, which is capable of pseudohyphal, true hyphal, and budding growth, both pathways are required for true hyphal production, suggesting a similar role in *Neurospora*[99]. However, at least three transcription factors, Tec1p, Flo8p and Sfl1, specifically required for regulating pseudohyphal development in *S. cerevisiae*[99] were not found in *Neurospora*. Conversely, *Neurospora* possesses a gene with similarity to a transcription factor shown to be critical for hyphal growth in *C. albicans*[75] (Efg1). This transcription factor is not required for pseudohyphal growth in *C. albicans*, nor is the homologous gene in *S. cerevisiae* (Phd1p) required for pseudohyphal growth[99]. More study of the complex pathways underlying these modes of growth is required. Nonetheless, these data clarify in part the distinctions and similarities between the signalling pathways and regulatory components of hyphal and pseudohyphal growth.

Macroconidiation pathway differs from that in *A. nidulans*. Macroconidia are asexual spores common to filamentous fungi, but absent from yeast. Macroconidiation, the development of macroconidia, begins with the differentiation of specialized aerial hyphae that rise perpendicular to the substrate[5,100]. The tips of aerial hyphae switch in *Neurospora* from an elongation to an apical budding program to form chains of proconidia in a structure termed the conidiophore. Maturation of the conidiophore culminates in the release of multinucleated conidia.

Components of the macroconidiation pathway have been identified in both *Neurospora* and the filamentous fungus *A. nidulans*, and known upstream signalling proteins appear to be conserved in both species[101]. In contrast, there is little conservation of downstream components between the two fungi. In *Neurospora*, the *acon-2*, *acon-3*, *fld*, and *fl* genes are essential for conidiation[5], whereas in *A. nidulans*, the FlbC, FlbD, BrlA, AbaA and WetA gene products are required. Our analysis of the genome sequence revealed that *Neurospora* possesses no FlbC, BrlA or AbaA homologues, and a protein with only very weak similarity to approximately 100 amino acids at the C-terminus of WetA. These data make clear that

the molecular machinery underlying macroconidiation in *Neurospora* differs significantly from that in *A. nidulans*. This may have important implications for the evolution of macroconidiation.

Secondary Metabolism

The fungal kingdom produces a spectacular array of small, bioactive compounds termed secondary metabolites best known for their roles as pigments, antibiotics, and mycotoxins. With the exception of the synthesis of carotenoid and melanin pigments, *Neurospora* has not been demonstrated to possess secondary metabolism. It was thus surprising that the *Neurospora* genome sequence revealed a number of putative genes for secondary metabolite production.

Non-ribosomal Peptide Synthetases. Three genes resembling non-ribosomal peptide synthetase (*NRPS*) genes and one *NRPS*-related gene were identified in the *Neurospora* genome sequence (Figure 8). One gene is orthologous to an *Aureobasidium pullulans NRPS*. The most closely related *NRPS* of known function is *sid2* of *Ustilago maydis* that is responsible for production of hydroxamate siderophores[102]. The remaining two *NRPS* genes are of unknown function, although one is orthologous to an *NRPS* in the genome of *Magnaporthe grisea* (Whitehead, Dean *et al*., unpublished data), and the other is orthologous to an *NRPS* found in the genomes of all filamentous ascomycetes sequenced to date, including four (*Cochliobolus heterostrophus, Botryotinia fuckeliana, Gibberella verticillioides* and *G. zeae*) sequenced by the Torrey Mesa Research Institute/Syngenta. The product of this gene is required for virulence of *C. heterostrophus* (B.-N. Lee *et al*, Torrey Mesa Research Institute, unpublished data) The *NRPS*-related gene is orthologous to the *C. heterostrophus CPS1* gene , a representative of a novel group of adenylate forming enzymes; the *CPS1* product contributes to virulence of *C. heterostrophus* on maize, *C. victoriae* on oats and *G zeae* on wheat (S.-W. Lu *et al*).

Polyketide Synthases. Seven polyketide synthase (*PKS*) genes were identified in the *Neurospora* genome, which could be classified into three groups based on domain structure (Figure 8). The first class contains genes similar to DHN-melanin PKS genes of *Exophila dermatitidis*[103], *Colletotrichum lagenarium*[104], and *Alternaria alternata*[105]. Sequence identity to a large number of EST sequences from sexual and perthecial libraries suggest a role for this PKS in melanin pigment synthesis during sexual development[106]. The genes in the second class of PKSs are similar in structure to several fungal genes including the *Aspergillus terreus lovF* gene required for lovastatin synthesis. The genes in the third class resemble other fungal genes including the *Aspergillus terreus* lovB gene that is also required for lovastatin synthesis.

Diterpene metabolism. Diterpenes comprise a diverse group of compounds, primarily in plants and fungi, with roles in defence, pathogenicity and regulation of plant growth. The genome sequence revealed several genes that are associated with diterpene biosynthesis in other organisms, including terpene synthase, several genes related to gibberellin oxidases, and a member of the cytochrome P450 monooxygenase gene family. These genes include at

least one member of each of the three enzyme classes required for the biosynthesis of gibberellic acid (GA). GA, a normal growth regulator in plants, was first identified as a metabolic product of the plant pathogen *Gibberella* (*Fusarium*) *fujikuroi*, a relative of *Neurospora* that causes "foolish seedling" disease in rice[107]. The presence of these genes in *Neurospora* suggests that many components necessary for GA production, an important factor in pathogenicity, were present in the ancestors of *Neurospora* and *G. fujikuroi*.

The discovery of genes associated with secondary metabolism provides new insight into the biology and evolution of *Neurospora*. We speculate that these genes may play roles in morphogenesis and chemotropism[108], interspecies communication, and possibly even chemical defence. Moreover, the identification of these genes in *Neurospora* suggests that apparent major differences in lifestyles among related fungi, such as pathogenicity, may derive in part from minor modifications of gene function and expression.

Plant Pathogenicity and *Neurospora*

The ability to parasitise living plants is widespread throughout the fungal kingdom. Although *Neurospora* is a saprotroph (that is, feeds on dead or decaying matter), the genome sequence contains numerous genes similar to those required for plant pathogenesis identified in fungal pathogens. A number of these genes have no other known function in other organisms except pathogenesis (Table 6). *Neurospora* possesses a wide range of extracellular enzymes capable of digesting plant cell wall polymers, although there is no clear cutinase homolog. Cutin is one of the main layers protecting the epidermis of the leaves of plants and many, but not all, plant pathogens have cutinase activity. *Neurospora* has a wide range of cytochrome P450 enzymes that are important in some host-pathogen systems for detoxification of plant anti-fungal compounds. In addition, a large number of identified ABC and MFS drug efflux systems could play a role in combating toxic plant compounds. The capability to form secondary metabolite members of the PKS, NRPS and terpenoid families, as described above, is present. Also, *Neurospora* contains all signal transduction components implicated in ascomycete pathogenesis that have been described to date (G protein, MAPK , calcium, and cAMP signalling). Finally, putative homologs were identified to several genes whose only defined roles are in infection efficiency and pathogen aggressiveness (Table 6). In sum, although *Neurospora* is not known to be a pathogen, the genome sequence has revealed many genes with similarity to those required for pathogenesis.

Conclusion

Although *Neurospora* has been a subject of intense study for over 70 years, the analysis of the genome sequence has provided many new insights into a variety of cellular processes including cell signalling, growth and differentiation, secondary metabolism, and genome defence. The analysis has also uncovered surprising similarities between the saprotrophic *Neurospora* and pathogenic fungi, providing a new perspective on the molecular underpinnings of these two lifestyles. Finally, the genome sequence has revealed the remarkable impact of RIP on the evolution of genes in *Neurospora*. Recent reports indicating the apparent presence of RIP in other fungi[109,110] broaden the implications of our findings.

The apparent lack of functional gene duplication in *Neurospora* provides a unique opportunity to study other modes of evolution in this experimentally tractable organism.

The genome sequence of *Neurospora* provides only a first glimpse into the genomic basis for the biological diversity of the filamentous fungi. Fungal genome sequences from the many ongoing[111] and planned[112] projects will expand this view as well as provide extraordinary opportunities for comparative analyses. This resulting revolution in fungal biology promises to yield insight into this important group of organisms, as well as to provide a deeper understanding of the fundamental cellular processes common to all eukaryotes.

Tables

Table 1 Sequencing Coverage by Clone Type

Clone Type	Methylation Tolerant Host	Average Insert Size (kb)	Assembled Reads	Sequence Coverage[1]	Physical Coverage[1]
Plasmid	No	4.6	424,269	6.0x	20.6x
Plasmid	Yes	4.0	921,589	13.2x	38.2x
Fosmid	Yes	41.2	30,234	0.33x	11.7x
Cosmid	Yes	40.3	12,958	0.18x	4.6x
Cosmid	Yes	38.5	6,938	0.10x	2.2x
BAC	Yes	66.9	16,873	0.17x	8.8x
Jumping Clones	Yes	47.0	18,442	0.26x	9.6x
Jumping Clones	Yes	28.7	892	0.01x	0.3x

[1]Based on estimated genome size of 41 Mb

Table 2 *Neurospora crassa* Genome Features

General	
Size (Assembly 5)	38,639,769 bp
Chromosomes	7
%G+C	50%
Protein-coding genes	10,082
Protein-coding genes > 100aa	9200
tRNA genes	424
5S rRNA genes	74
Percent coding	44%
Average gene size	1673 bp (481aa)
Average intergenic distance	1953 bp
Predicted protein-coding sequences	
Identified by similarity to known sequences	1336 (13%)
Conserved hypothetical proteins	4606 (46%)
Predicted proteins (no similarity to known sequences)	4140 (41%)

Table 3 Summary of Interpro Domain and Domain Architecture Analysis

Organism	Total Genes	Domains	%Genes w/Domain Hits	Distinct Architectures	Ave #Genes/ Architecture	Ave # Architectures per Domain	Ave # Domains per Architecture	Percent Duplicated Architectures
A. thaliana	26147	539	38%	1579	5.4	5.4	3.2	46%
C. elegans	20206	586	37%	1959	6.9	6.9	4.4	36%
D. melanogaster	14335	580	35%	1696	5.9	5.9	4.3	33%
N. crassa	9200[1]	477	31%	965	3.5	3.5	3.1	28%
S. cerevisiae	6306	443	34%	786	3.0	3.0	3.0	34%
S. pombe	4824	438	38%	744	2.9	2.9	3.0	31%
E. cuniculi	1996	209	35%	296	2.2	2.2	2.7	23%

[1]Only genes coding for proteins longer than 100 aa were included in the analysis

Table 4 Most Abundant Interpro Entries in *Neurospora*

Interpro Domain	*Neurospora*	*S. pombe*	*S. cerevisiae*	*A. thaliana*
Proline rich extensin	646 (6.4%)	55 (1.1%)	78 (1.2%)	461 (1.8%)
Proline rich region	575 (5.7%)	48 (1.0%)	90 (1.4%)	559 (2.2%)
Type I antifreeze protein	171 (1.7%)	18 (0.4%)	34 (0.5%)	70 (0.3%)
Zinc finger, C2H2 type	93 (0.9%)	35 (0.7%)	53 (0.8%)	162 (0.6%)
Eggshell protein	78 (0.8%)	3 (0.1%)	7 (0.1%)	116 (0.5%)
Esterase/lipase/thioesterase family active site	76 (0.8%)	25 (0.5%)	37 (0.6%)	211 (0.8%)
SAM binding motif	66 (0.7%)	31 (0.6%)	21 (0.3%)	95 (0.4%)
Tropomyosin	65 (0.6%)	19 (0.4%)	22 (0.3%)	136 (0.5%)
Allergen Poa p 9, N- terminal	61 (0.6%)	6 (0.1%)	12 (0.2%)	24 (0.1%)
Short-chain dehydrogenase/reductase (SDR) superfamily	61 (0.6%)	24 (0.5%)	16 (0.2%)	83 (0.3%)
Glucose/ribitol dehydrogenase	60 (0.6%)	26 (0.5%)	21 (0.3%)	103 (0.4%)
Ankyrin-repeat	51 (0.5%)	14 (0.3%)	18 (0.3%)	105 (0.4%)
FAD-dependent pyridine nucleotide-disulphide oxidoreductase	44 (0.4%)	17 (0.3%)	18 (0.3%)	66 (0.3%)
Cytochrome P450 enzyme	38 (0.4%)	3 (0.1%)	4 (0.1%)	231 (0.9%)
2,3-dihydro-2,3dihydroxybezoate dehydrogenase (EntA)	38 (0.4%)	18 (0.4%)	10 (0.2%)	67 (0.3%)
Pyridine nucleotide-disulphide oxidoreductase, class I	37 (0.4%)	12 (0.2%)	17 (0.3%)	56 (0.2%)
NAD binding site	37 (0.4%)	15 (0.3%)	10 (0.2%)	73 (0.3%)
E-class P450 group I	33 (0.4%)	2 (0.0%)	3 (0.0%)	216 (0.9%)
HMG-I and HMG-Y DNA-binding domain (A+T-hook)	35 (0.3%)	5 (0.1%)	7 (0.1%)	51 (0.2%)
Histone H1 and H5 family	34 (0.3%)	3 (0.1%)	7 (0.1%)	28 (0.1%)
E-class P450 group IV	33 (0.3%)	2 (0.0%)	3 (0.0%)	216 (0.8%)
Aromatic-ring hydroxylase	32 (0.3%)	2 (0.0%)	4 (0.1%)	20 (0.1%)
Molluscan rhodopsin C-terminal tail	30 (0.3%)	6 (0.1%)	4 (0.1%)	66 (0.3%)

Table 5 *Neurospora* has two RNA-silencing pathways

Predicted protein	*Neurospora*	*A. fumigatus*[1]	*S. pombe*[2]	Pathway[3]
RNA-directed RNA polymerases	*qde-1* (NCU07534.1)	*rrpA* (Contig158)		Quelling
	Sad-1 (NCU02178.1)	*rrpB* (Contig472)	*rdp1*$^+$ (SPAC6F12.09)	Meiotic Silencing
	rrp-3 (NCU08435.1)			Unknown
Argonaute-like, related to translation initiation factors	*qde-2* (NCU04730.1)	*ppdA* (Contig720)		Quelling
	Sms-2 (NCU09434.1)	*ppdB* (Contig196)	*ago1*$^+$ (SPCC736.11)	Meiotic Silencing
Dicer-like, related to SFII-RNAseIII ribonucleases of the carpel factory	*dcl-2* (NCU06766.1)	*dclB* (Contig618)		Quelling
	Sms-3 (NCU08270.1)	*dclA* (Contig310)	*dcr1*$^+$ (SPCC584.10C)	Meiotic Silencing
RecQ helicase-like, related to Bloom's and Werner syndrome helicases	*qde-3* (NCU08598.1)	*rqhA* (Contig443)		Quelling
	rqh-2 (NCU03337.1)[4]	*rqhB* (Contig58)[4]	*hus2*$^+$ (SPAC2G11.12)	Unknown

[1]Unfinished *Aspergillus fumigatus* Genome Project at: http://www.tigr.org. [2]*Schizosaccharomyces pombe* Genome Project at: http://www.genedb.org. [3]Pathway assigned based on either known experimental data for *qde-1*, *qde-2* and *qde-3* (quelling pathway); *Sad-1*, *Sms-2,* and *Sms-3* (meiotic silencing pathway); or predicted based on phylogenetic analysis[4]. *RecQ helicase-like* (*rqh*).

Table 6 Pathogenesis associated genes in *Neurospora*

Gene	Homolog[a]	Organism	Pathogenesis Function
NCU08038	GAS1	*M. grisea*	Infection efficiency
NCU06170	GAS2	*M. grisea*	Infection efficiency
NCU02903	PTH11	*M. grisea*	Putative membrane receptor for host sensing
NCU07432	PLS1	*M. grisea*	Tetraspanin, req. for pathogenic development
NCU05521	PATH531	*M. grisea*	Aggressiveness factor
NCU05730	PEP2	*N. haematococca*[b]	Aggressiveness factor
NCU03370	CAP20	*C. gloeosporioides*[c]	Aggressiveness factor
NCU06937 NCU06938 NCU02021 NCU04482 NCU09660	ECP2	*C. fulvum*[d]	Aggressiveness factor, avirulence factor. Gene family in *Neurospora*

[a]Only genes with no known role other than pathogenesis. [b]*Nectria haematococca*. [c]*Colletotrichum gloeosporioides*. [d]*Cladosporium fulvum*.

Figures

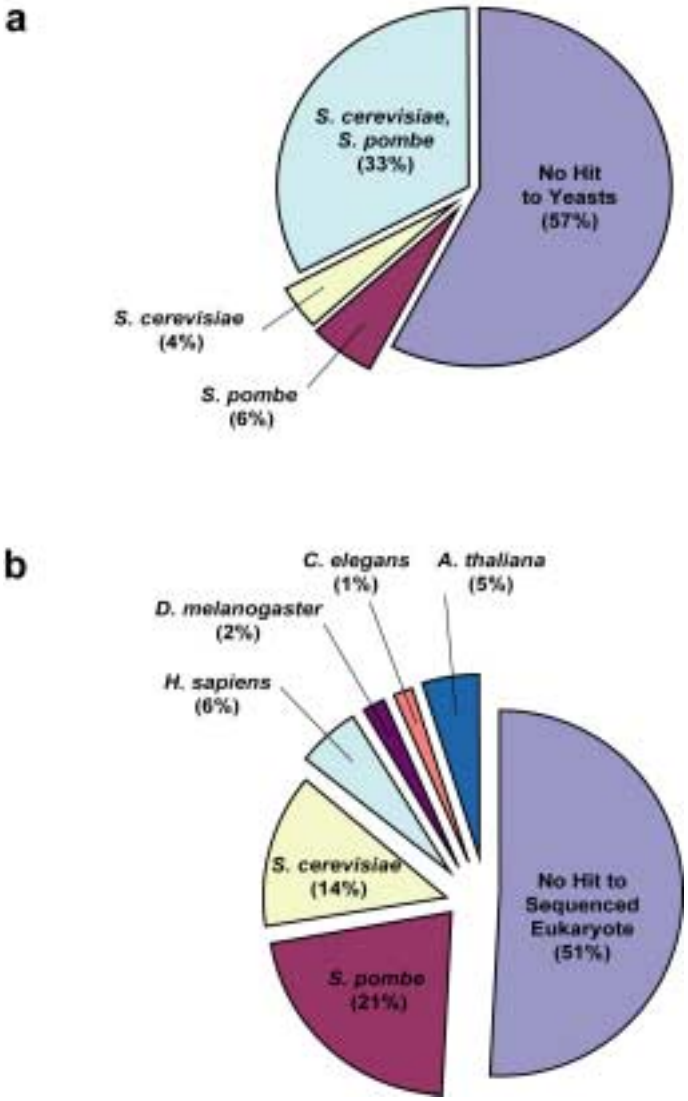

Figure 1: Summary of BLASTP Analysis of *Neurospora* proteins. (a) Proportion of *Neurospora* proteins with BLASTP hits (p<1e-5) to *S. cerevisiae* and *S. pombe*. (b) Proportion of *Neurospora* proteins with top BLASTP hits to selected sequenced eukaryotic organisms.

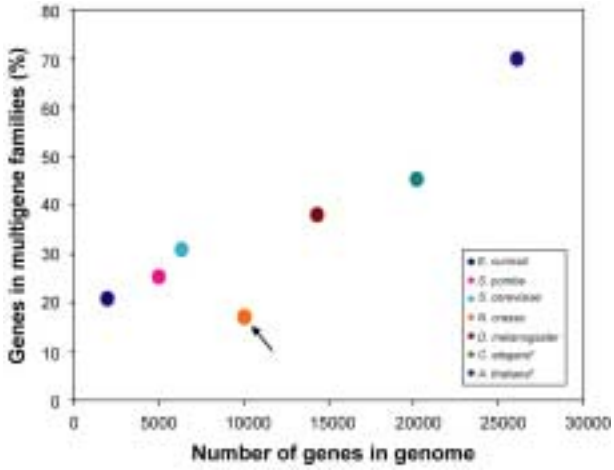

Figure 2: *Neurospora* has a low proportion of genes in multigene families. The graph displays the proportion of genes in multigene families (see Methods) as a function of the number of genes in the genomes of selected sequenced eukaryotic organisms. Arrow indicates *Neurospora*. See text for more details.

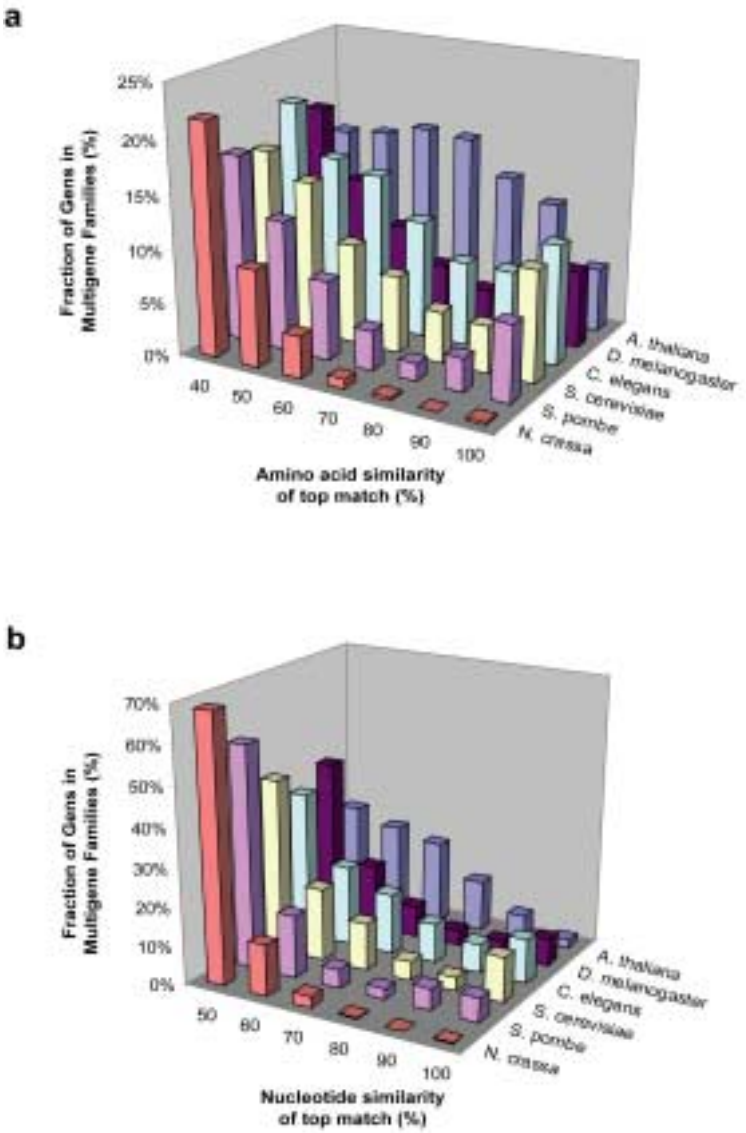

Figure 3: *Neurospora* possesses few highly similar genes. Histogram of (a) amino acid and (b) nucleotide percent identity of top scoring self-matches for genes in selected sequenced eukaryotic genomes. For each organism, the protein and coding regions for each gene (not including pseudogenes) were compared to those of every other gene in the same genome using BLASTX. Top scoring matches were aligned using ClustalW and percent identities calculated. In contrast to other eukaryotes, *Neurospora* possesses only 8 genes with a top match of greater than 80% amino acid or nucleotide identity.

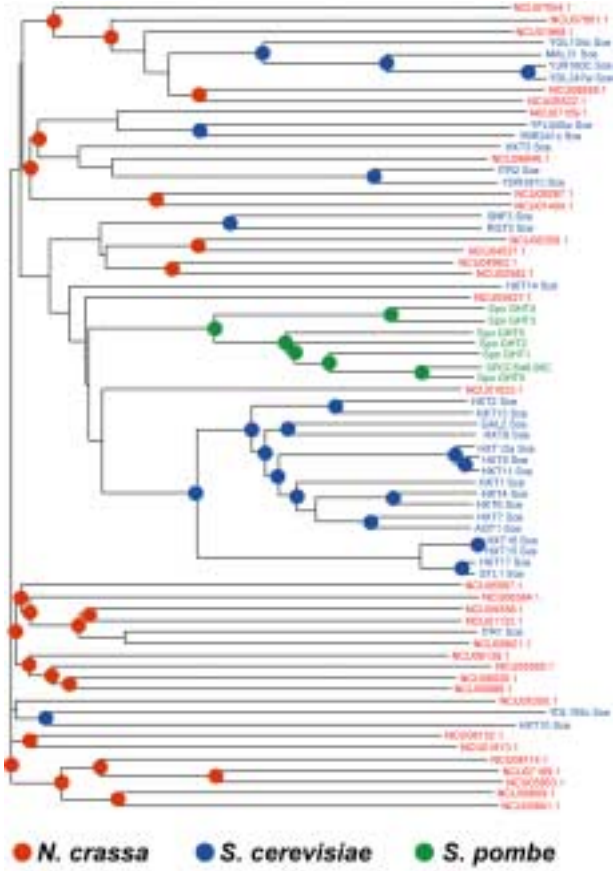

Figure 4: Example of lack of recent duplications in *Neurospora* gene family. Phylogenetic tree of major facilitator superfamily (MFS) sugar transporters from *S. cerevisiae*, *S. pombe*, and *Neurospora*. The amino acid sequences were aligned using CLUSTALW, manually trimmed to remove ambiguously aligned regions, and a neighbour-joining phylogenetic tree was generated from the alignment using PAUP*. Coloured dots represent branching points between predicted paralogous genes in *Neurospora* (orange), *S. cerevisiae* (blue), and *S. pombe* (green). In contrast to both yeasts, *Neurospora* transporters contain no predicted instances of recent paralogous duplication.

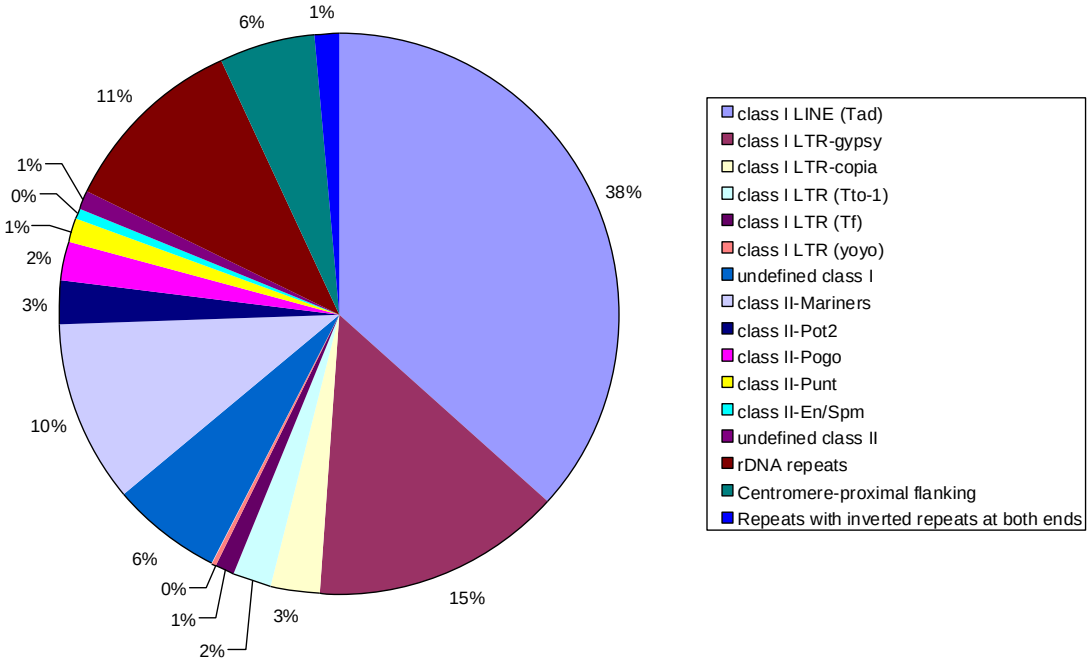

Figure 5: Distribution of identified RIP-mutated repeat elements. A total of 46% of repetitive sequence in the *Neurospora* assembly could be identified as mobile element relics. No intact mobile elements were identified.

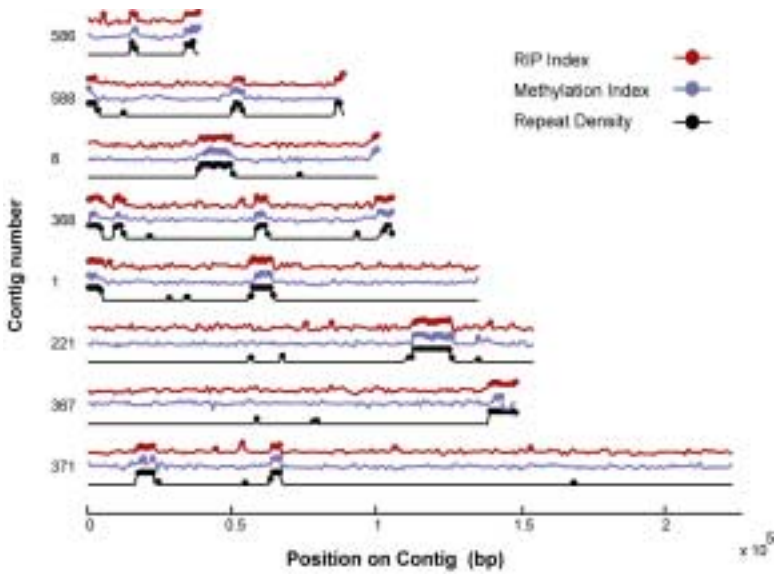

Figure 6: Correspondence between predicted RIP, methylation, and repetitive DNA. Prediction of RIP, methylation, and repeat sequence in 1kb windows for selected contigs. Red lines plot the TpA/ApT RIP Index (see Methods). Red dots indicate windows predicted to be RIP-mutated (TpA/ApT>1.2). Blue lines plot proportion of reads from methyl-tolerant library. Blue dots indicate windows predicted to be methylated based on >70% methyl-tolerant reads (see Methods). Black lines plot repeat content as fraction of nucleotides in each window that is in repetitive sequence. Black dots indicate windows with >50% repeat sequence. Contigs were selected to illustrate regions predicted as methylated.

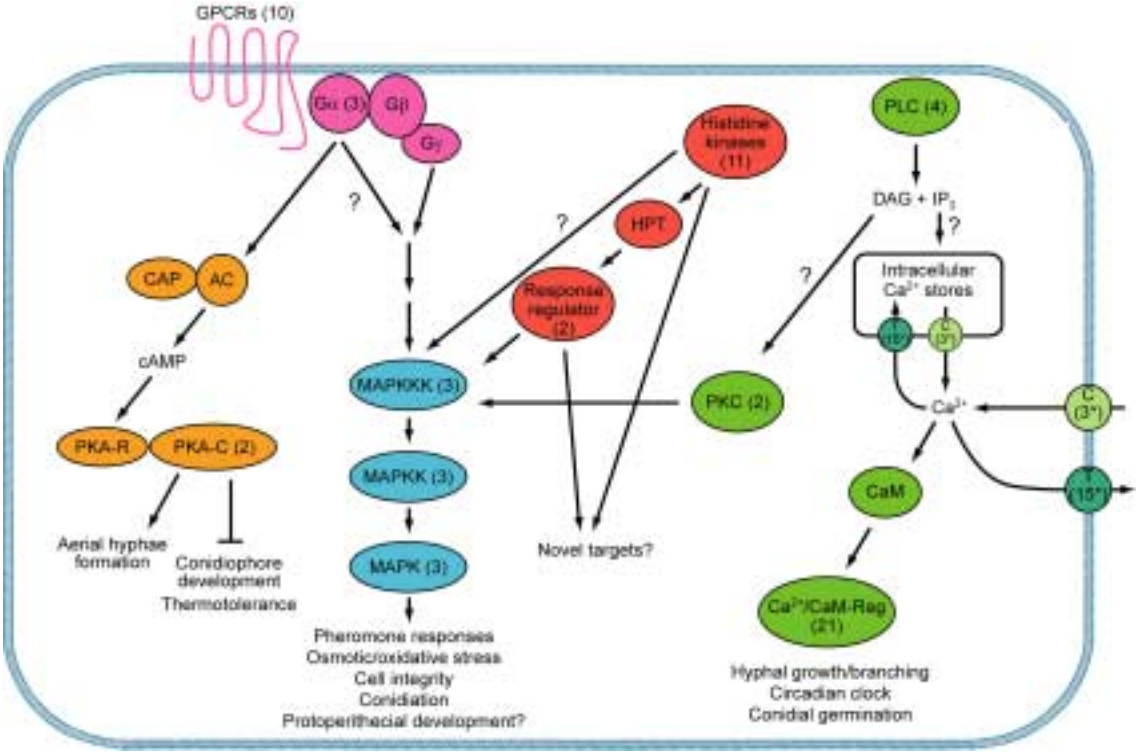

Figure 7: Overview of major intracellular signalling pathways in *Neurospora*. Abbreviations: AC, adenylyl cyclase; C, Ca^{2+}-channel protein; CaM calmodulin; Ca^{2+}/CaM-Reg, calcium and calmodulin regulated protein; CAP, cyclase-associated protein; GPCR, G-protein coupled receptor; Gα, G-protein α subunit; Gβ, G-protein β subunit; Gγ, G-protein γ subunit; HPT, histidine-containing phosphotransfer domain protein; MAPK, MAP kinase; MAPKK, MAPK kinase; MAPKKK, MAPKK kinase; PKA-R, protein kinase A regulatory subunit; PKA-C, protein kinase A catalytic subunit; PLC, phospholipase C; PKC, protein kinase C; T, Ca^{2+}-transport protein (P-type Ca^{2+}-ATPase, H^+/Ca^{2+}-exchanger, or Na^+/Ca^{2+}-exchanger);. *Location in plasma membrane and/or organelle membranes not determined.

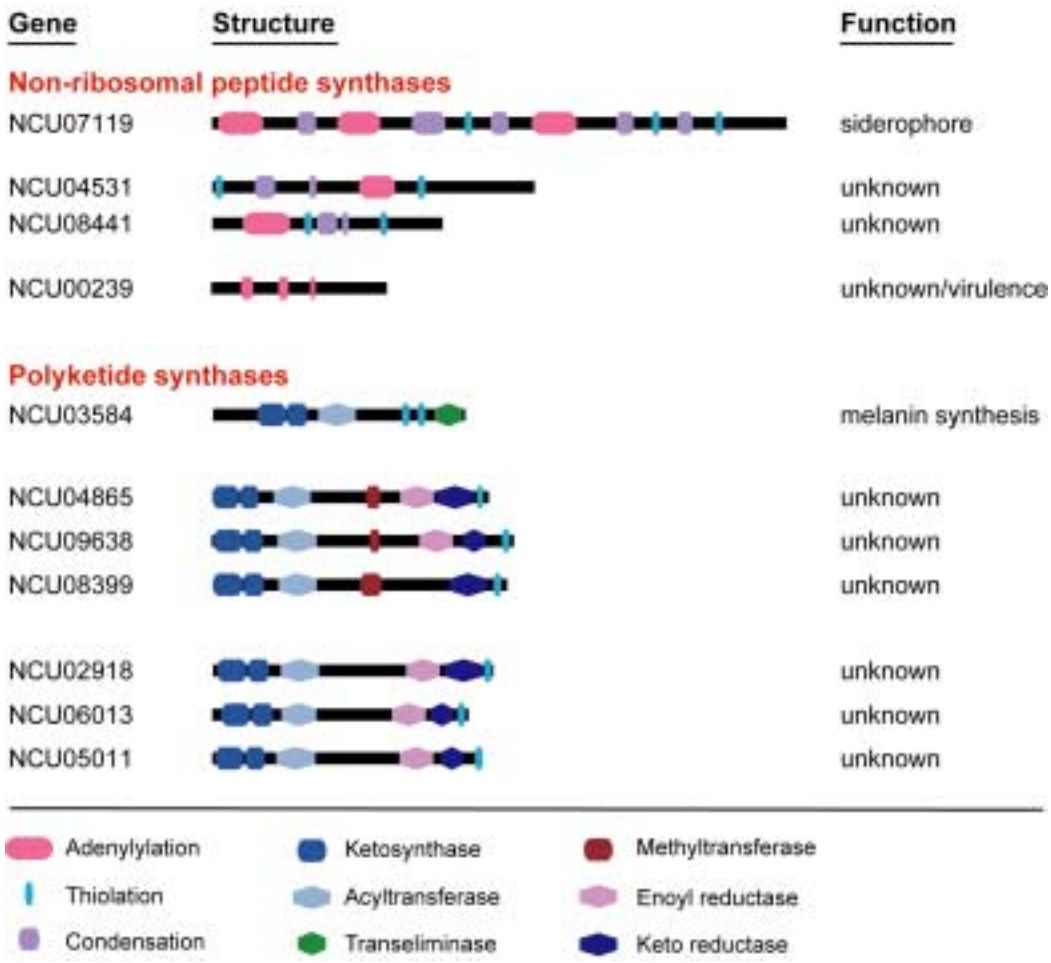

Figure 8: Domain structures of predicted *Neurospora* non-ribosomal peptide synthase (NRPS) and polyketide synthase (PKS) genes. Domains were predicted using a combination of PFAM searches using HMMER, protein alignments, and manual inspection.

Methods

Strain and Growth Conditions. Twenty 5 ml cultures of *N. crassa* wild-type strain N150 ("74-OR23-1VA"; FGSC#2489) were grown on a shaker in Vogel's minimal medium[5] for 3 days at 32° C. Tissues were harvested, freeze-dried overnight and DNA extracted as previously described[113]. DNA from the twenty samples was mixed and used for library construction.

Sequencing and assembly. The genome was sequenced by the whole-genome shotgun method. Plasmid (4-kb inserts) and Fosmid (40-kb inserts) libraries were generated as described at http://www-genome.wi.mit.edu/. Jumping clone (subclone) libraries with 50-kb inserts were generated as described elsewhere[114]. *Neurospora* cosmid and BAC clones were obtained from previously constructed libraries[115,116]. Sequencing methods for all clone types are described at http://www-genome.wi.mit.edu/. All inserts were sequenced from both ends to generate paired-reads. The sequence coverage generated is shown in Table 1. The sequence was assembled using Arachne[8]. Finished sequence from linkage groups II and V was provided by MIPS and is available at http://mips.gsf.de/proj/*Neurospora*/.

Annotation and Analysis. The *Neurospora* genome was annotated using the Calhoun annotation system. The genome sequence was searched against the public protein databases using BLASTX with threshold E≤1e-5. Genes were predicted using a combination of FGENESH, FGENESH+, and GENEWISE. The gene calling programs were validated against a test set of 191 previously characterized *Neurospora* proteins. Predicted genes were validated against ESTs aligned to the genome using SIM4. All predicted genes were searched against the PFAM set of hidden Markov models using the HMMER program and the public protein databases using BLASTP. Transfer RNAs were identified using the tRNAScan-SE program (version X). Multigene families were constructed by searching each annotated gene against every other gene using BLASTP, requiring matches with E ≤ 1e-5 over 60% of the longer gene length, and clustering genes based on single linkage transitive closure. Repeat sequences were detected by searching the genome sequence against itself using CrossMatch, filtering for alignments longer than 200bp in length, and clustering pairs based on region overlap.

RIP-mutated regions were detected by calculating one or both of two different dinucleotide ratios for sequence regions[17]. Regions with TpA/ApT>2 or (CpA+TpG)/(ApC+GpT)<0.7 were predicted as RIP-mutated. Prediction of RIP sequence across the genome utilized only the TpA/ApT ratio, while the analysis of coding sequences used both (with a positive prediction by either measure taken as a prediction as RIP). RIP simulations were based on parameters derived from Watters et al. (1999). DNA methylation was predicted by calculating the proportion of plasmid reads overlapping 1 kb windows from both the methylation-tolerant and methylation-intolerant libraries. Regions with greater than 70% reads derived from the methylation-tolerant library were predicted methylated. Specificity was estimated as described in the text. Methylation was experimentally assessed using Southern analyses as described elsewhere[51]. Sensitivity was estimated by testing 19 repetitive and RIP mutated 1 kb regions that were not predicted to be methylated. Of the 19 regions, 14 were in fact methylated.

Predicted RNA silencing genes were aligned with homologs from plants, animals and other fungi using T-Coffee v1.37. C-terminal and N-terminal regions of low homology were removed and the sequences realigned until alignments started and stopped at regions of

unambiguous similarity. Both joining-joining trees, using ClustalX, and maximum posterior probability trees, using MrBayes 2.01, were generated and analysed.

Additional details, analysis results, and the genome sequence are available at http://www-genome.wi.mit.edu/.

References

1. Payen, A. (rapporteur) Extrait d'un rapport adressé à M. Le Maréchal Duc de Dalmatie, Ministre de la Guerre, Président du Conseil, sur une altération extraordinaire du pain de munition. *Ann. Chim. Phys. 3e Ser.* **9**, 5-21 (+one plate) (1843).
2. Shear, C. L. & Dodge, B. O. Life histories and heterothallism of the red bread-mold fungi of the Monilia sitophila group. *J. Agr. Res.* **34**, 1019-1042 (1927).
3. Lindegren, C. C. A six-point map of the sex chromosome of Neurospora crassa. *J. Genet.* **32**, 243-256 (1936).
4. Beadle, G. W. & Tatum, E. L. Genetic control of biochemical reactions in Neurospora. *Proc Natl Acad Sci U S A* **27**, 499-506 (1941).
5. Davis, R. H. *Neurospora : contributions of a model organism* (Oxford University Press, New York, 2000).
6. Davis, R. H. & Perkins, D. D. Timeline: Neurospora: a model of model microbes. *Nat Rev Genet* **3**, 397-403 (2002).
7. Perkins, D. D., Radford, A. & Sachs, M. S. *The Neurospora Compendium: Chromosomal loci* (Academic Press, San Diego, CA, 2001).
8. Batzoglou, S. et al. ARACHNE: a whole-genome shotgun assembler. *Genome Res* **12**, 177-89 (2002).
9. Schulte, U., Becker, I., Mewes, H. W. & Mannhaupt, G. Large scale analysis of sequences from Neurospora crassa. *J Biotechnol* **94**, 3-13 (2002).
10. Free, S. J., Rice, P. W. & Metzenberg, R. L. Arrangement of the genes coding for ribosomal ribonucleic acids in Neurospora crassa. *J Bacteriol* **137**, 1219-26. (1979).
11. Spingola, M., Grate, L., Haussler, D. & Ares, M., Jr. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *Rna* **5**, 221-34 (1999).
12. Voelker, R. & Berglund, J. A. (Unpublished).
13. Zdobnov, E. M. & Apweiler, R. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-8. (2001).
14. Selker, E. U., Cambareri, E. B., Jensen, B. C. & Haack, K. R. Rearrangement of duplicated DNA in specialized cells of Neurospora. *Cell* **51**, 741-52. (1987).
15. Selker, E. U. Premeiotic instability of repeated sequences in Neurospora crassa. *Annu Rev Genet* **24**, 579-613 (1990).
16. Cambareri, E. B., Jensen, B. C., Schabtach, E. & Selker, E. U. Repeat-induced G-C to A-T mutations in Neurospora. *Science* **244**, 1571-5. (1989).
17. Margolin, B. S. et al. A methylated Neurospora 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* **149**, 1787-97. (1998).
18. Watters, M. K., Randall, T. A., Margolin, B. S., Selker, E. U. & Stadler, D. R. Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in Neurospora. *Genetics* **153**, 705-14. (1999).
19. Cambareri, E. B., Singer, M. J. & Selker, E. U. Recurrence of repeat-induced point mutation (RIP) in Neurospora crassa. *Genetics* **127**, 699-710. (1991).
20. Selker, E. U. Repeat-induced gene silencing in fungi. *Adv Genet* **46**, 439-50 (2002).
21. Selker, E. U. & Garrett, P. W. DNA sequence duplications trigger gene inactivation in Neurospora crassa. *Proc Natl Acad Sci U S A* **85**, 6870-4. (1988).
22. Rountree, M. R. & Selker, E. U. DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa. *Genes Dev* **11**, 2383-95. (1997).

23. Cambareri, E. B., Aisner, R. & Carbon, J. Structure of the chromosome VII centromere region in Neurospora crassa: degenerate transposons and simple repeats. *Mol Cell Biol* **18**, 5465-77. (1998).
24. Kinsey, J. A., Garrett-Engele, P. W., Cambareri, E. B. & Selker, E. U. The Neurospora transposon Tad is sensitive to repeat-induced point mutation (RIP). *Genetics* **138**, 657-64. (1994).
25. Irelan, J. T., Hagemann, A. T. & Selker, E. U. High frequency repeat-induced point mutation (RIP) is not associated with efficient recombination in Neurospora. *Genetics* **138**, 1093-103. (1994).
26. Perkins, D. D., Margolin, B. S., Selker, E. U. & Haedo, S. D. Occurrence of repeat induced point mutation in long segmental duplications of Neurospora. *Genetics* **147**, 125-36. (1997).
27. Razin, S., Yogev, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* **62**, 1094-156 (1998).
28. Apweiler, R. et al. InterPro--an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* **16**, 1145-50. (2000).
29. Ohno, S. *Evolution by Gene Duplication* (Springer-Verlag, 1970).
30. Walsh, J. B. How Often Do Duplicated Genes Evolve New Functions? *Genetics* **139**, 421-428 (1995).
31. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155 (2000).
32. Krumlauf, R. & Marzluf, G. A. Genome organization and characterization of the repetitive and inverted repeat DNA sequences in Neurospora crassa. *J Biol Chem* **255**, 1138-45. (1980).
33. Krumlauf, R. & Marzluf, G. A. Characterization of the sequence complexity and organization of the Neurospora crassa genome. *Biochemistry* **18**, 3705-13. (1979).
34. Selker, E. U. Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion. *Trends Genet* **13**, 296-301 (1997).
35. Centola, M. & Carbon, J. Cloning and characterization of centromeric DNA from Neurospora crassa. *Mol Cell Biol* **14**, 1510-9. (1994).
36. Bibbins, M., Cummings, N. J. & Connerton, I. F. DAB1: a degenerate retrotransposon-like element from Neurospora crassa. *Mol Gen Genet* **258**, 431-6. (1998).
37. Schechtman, M. G. Characterization of telomere DNA from Neurospora crassa. *Gene* **88**, 159-65. (1990).
38. Langin, T., Capy, P. & Daboussi, M. J. The transposable element impala, a fungal member of the Tc1-mariner superfamily. *Mol Gen Genet* **246**, 19-28. (1995).
39. Hua-Van, A., Hericourt, F., Capy, P., Daboussi, M. J. & Langin, T. Three highly divergent subfamilies of the impala transposable element coexist in the genome of the fungus Fusarium oxysporum. *Mol Gen Genet* **259**, 354-62. (1998).
40. Villalba, F., Lebrun, M. H., Hua-Van, A., Daboussi, M. J. & Grosjean-Cournoyer, M. C. Transposon impala, a novel tool for gene tagging in the rice blast fungus Magnaporthe grisea. *Mol Plant Microbe Interact* **14**, 308-15. (2001).
41. Kachroo, P., Leong, S. A. & Chattoo, B. B. Pot2, an inverted repeat transposon from the rice blast fungus Magnaporthe grisea. *Mol Gen Genet* **245**, 339-48. (1994).
42. Nakayashiki, H., Nishimoto, N., Ikeda, K., Tosa, Y. & Mayama, S. Degenerate MAGGY elements in a subgroup of Pyricularia grisea: a possible example of successful capture of a genetic invader by a fungal genome. *Mol Gen Genet* **261**, 958-66. (1999).

43. Campell, B. R., Song, Y., Posch, T. E., Cullis, C. A. & Town, C. D. Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. *Gene* **112**, 225-8 (1992).
44. Timofeeva, M. et al. [Organization of a 5S ribosomal RNA gene cluster in the human genome]. *Mol Biol (Mosk)* **27**, 861-8 (1993).
45. Belkhiri, A., Intengan, H. & Klassen, G. R. A tandem array of 5S ribosomal RNA genes in Pythium irregulare. *Gene* **186**, 155-9 (1997).
46. Martins, C. & Galetti, P. M., Jr. Two 5S rDNA arrays in neotropical fish species: is it a general rule for fishes? *Genetica* **111**, 439-46 (2001).
47. Valarik, M. et al. Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (Musa spp.). *Chromosome Res* **10**, 89-100 (2002).
48. Selker, E. U. et al. Dispersed 5S RNA genes in N. crassa: structure, expression and evolution. *Cell* **24**, 819-28. (1981).
49. Irelan, J. T. & Selker, E. U. Cytosine methylation associated with repeat-induced point mutation causes epigenetic gene silencing in Neurospora crassa. *Genetics* **146**, 509-23. (1997).
50. Selker, E. U. & Stevens, J. N. DNA methylation at asymmetric sites is associated with numerous transition mutations. *Proc Natl Acad Sci U S A* **82**, 8114-8. (1985).
51. Selker, E. U., Fritz, D. Y. & Singer, M. J. Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in Neurospora. *Science* **262**, 1724-8. (1993).
52. Russell, P. J., Rodland, K. D., Rachlin, E. M. & McCloskey, J. A. Differential DNA methylation during the vegetative life cycle of Neurospora crassa. *J Bacteriol* **169**, 2902-5. (1987).
53. Foss, H. M., Roberts, C. J., Claeys, K. M. & Selker, E. U. Abnormal chromosome behavior in Neurospora mutants defective in DNA methylation. *Science* **262**, 1737-41. (1993).
54. Selker, E. U., Jensen, B. C. & Richardson, G. A. A portable signal causing faithful DNA methylation de novo in Neurospora crassa. *Science* **238**, 48-53. (1987).
55. Singer, M. J., Marcotte, B. A. & Selker, E. U. DNA methylation associated with repeat-induced point mutation in Neurospora crassa. *Mol Cell Biol* **15**, 5586-97. (1995).
56. Miao, V. P., Freitag, M. & Selker, E. U. Short TpA-rich segments of the zeta-eta region induce DNA methylation in Neurospora crassa. *J Mol Biol* **300**, 249-73. (2000).
57. Kouzminova, E. & Selker, E. U. dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in Neurospora. *Embo J* **20**, 4309-23. (2001).
58. Freitag, M., Williams, R. L., Kothe, G. O. & Selker, E. U. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in Neurospora crassa. *Proc Natl Acad Sci U S A* **99**, 8802-7. (2002).
59. Selker, E. U. et al. Induction and maintenance of nonsymmetrical DNA methylation in Neurospora. *Proc Natl Acad Sci U S A* (2002).
60. Frommer, M. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89**, 1827-31 (1992).
61. Grant, S. R. Dissecting the mechanisms of posttranscriptional gene silencing: divide and conquer. *Cell* **96**, 303-6 (1999).
62. Zamore, P. D. Ancient pathways programmed by small RNAs. *Science* **296**, 1265-9 (2002).

63. Vance, V. & Vaucheret, H. RNA silencing in plants--defense and counterdefense. *Science* **292**, 2277-80 (2001).
64. Hammond, S. M., Caudy, A. A. & Hannon, G. J. Post-transcriptional gene silencing by double-stranded RNA. *Nat Rev Genet* **2**, 110-9 (2001).
65. Rougvie, A. E. Control of developmental timing in animals. *Nat Rev Genet* **2**, 690-701 (2001).
66. Hannon, G. J. RNA interference. *Nature* **418**, 244-51 (2002).
67. Wolffe, A. P. & Matzke, M. A. Epigenetics: regulation through repression. *Science* **286**, 481-6 (1999).
68. Hutvagner, G. & Zamore, P. D. RNAi: nature abhors a double-strand. *Curr Opin Genet Dev* **12**, 225-32 (2002).
69. Cogoni, C. & Macino, G. Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166-9 (1999).
70. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Involvement of small RNAs and role of the qde genes in the gene silencing pathway in Neurospora. *Genes Dev* **16**, 790-5 (2002).
71. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Gene silencing in worms and fungi. *Nature* **404**, 245 (2000).
72. Cogoni, C. et al. Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *Embo J* **15**, 3153-63 (1996).
73. Cogoni, C. & Macino, G. Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa. *Proc Natl Acad Sci U S A* **94**, 10233-8 (1997).
74. Cogoni, C. & Macino, G. Posttranscriptional gene silencing in Neurospora by a RecQ DNA helicase. *Science* **286**, 2342-4 (1999).
75. Aramayo, R. & Metzenberg, R. L. Meiotic transvection in fungi. *Cell* **86**, 103-13 (1996).
76. Shiu, P. K., Raju, N. B., Zickler, D. & Metzenberg, R. L. Meiotic silencing by unpaired DNA. *Cell* **107**, 905-16 (2001).
77. Shiu, P. K. & Metzenberg, R. L. Meiotic Silencing by Unpaired DNA. Properties, regulation and suppression. *Genetics* **161**, 1483-95 (2002).
78. Volpe, T. A. et al. Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* **297**, 1833-7 (2002).
79. Lee, D., Pratt, R. J., McLaughlin, M. & Aramayo, R. *Genetics* (In Press).
80. McLaughlin, M., Lee, D., Pratt, R. J., Dado, J. & Aramayo, R. (In Preparation).
81. Linden, H., Ballario, P., Arpaia, G. & Macino, G. Seeing the light: news in Neurospora blue light signal transduction. *Adv Genet* **41**, 35-54 (1999).
82. Devlin, P. F. & Kay, S. A. Circadian photoperception. *Annu Rev Physiol* **63**, 677-94 (2001).
83. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37-40 (2001).
84. West, A. H. & Stock, A. M. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369-76 (2001).
85. Alex, L. A. & Simon, M. I. (Unpublished data).
86. Pott, G. B., Miller, T. K., Bartlett, J. A., Palas, J. S. & Selitrennikoff, C. P. The isolation of FOS-1, a gene encoding a putative two-component histidine kinase from Aspergillus fumigatus. *Fungal Genet Biol* **31**, 55-67. (2000).

87. Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. Model systems for the study of seven-transmembrane-segment receptors. *Annu Rev Biochem* **60**, 653-88 (1991).
88. Klein, P. S. et al. A chemoattractant receptor controls development in Dictyostelium discoideum. *Science* **241**, 1467-72 (1988).
89. Josefsson, L. G. & Rask, L. Cloning of a putative G-protein-coupled receptor from Arabidopsis thaliana. *Eur J Biochem* **249**, 415-20 (1997).
90. Aubry, L. & Firtel, R. Integration of signaling networks that regulate Dictyostelium differentiation. *Annu Rev Cell Dev Biol* **15**, 469-517 (1999).
91. Kore-eda, S., Murayama, T. & Uno, I. Isolation and characterization of the adenylate cyclase structural gene of Neurospora crassa. *Jpn J Genet* **66**, 317-34 (1991).
92. Ivey, F. D., Yang, Q. & Borkovich, K. A. Positive regulation of adenylyl cyclase activity by a galphai homolog in Neurospora crassa. *Fungal Genet Biol* **26**, 48-61 (1999).
93. Gadd, G. M. in *The Growing Fungus* (eds. Gow, N. A. R. & Gadd, G. M.) 183-210 (Chapman & Hall, London, 1994).
94. Bootman, M. D. et al. Calcium signalling--an overview. *Semin Cell Dev Biol* **12**, 3-10 (2001).
95. Lakin-Thomas, P. L. Effects of inositol starvation on the levels of inositol phosphates and inositol lipids in Neurospora crassa. *Biochem J* **292 (Pt 3)**, 805-11 (1993).
96. Silverman-Gavrila, L. B. & Lew, R. R. Regulation of the tip-high [Ca2+] gradient in growing hyphae of the fungus Neurospora crassa. *Eur J Cell Biol* **80**, 379-90 (2001).
97. Cornelius, G., Gebauer, G. & Techel, D. Inositol trisphosphate induces calcium release from Neurospora crassa vacuoles. *Biochem Biophys Res Commun* **162**, 852-6 (1989).
98. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. & Fink, G. R. Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. *Cell* **68**, 1077-90 (1992).
99. Lengeler, K. B. et al. Signal transduction cascades regulating fungal development and virulence. *Microbiol Mol Biol Rev* **64**, 746-85 (2000).
100. Griffin, D. H. *Fungal Physiology* (Wiley-Liss, New York, 1994).
101. Ivey, F. D., Kays, A. M. & Borkovich, K. A. Shared and independent roles for a Galpha(i) protein and adenylyl cyclase in regulating development and stress responses in Neurospora crassa. *Eukaryot Cell* **1**, 634-42 (2002).
102. Yuan, W., Gentil, G., Budde, A. & Leong, S. Characterization of the Ustilago maydis sid2 gene encoding a multidomain peptide synthetase in the ferrichrome biosynthetic gene cluster. *J. Bacteriol.* **183**, 4040-4051 (2001).
103. Feng, B. et al. Molecular cloning and characterization of *WdPKS1*, a gene involved in dihydroxynaphthalene melanin biosynthesis and virulence in *Wangiella* (*Exophiala*) *dermatitidis*. *Infect. Immun.* **69**, 1781-1794 (2001).
104. Takano, Y. et al. Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium*. *Mol. Gen. Genet.* **249**, 162-167 (1995).
105. Takano, Y., Kubo, Y., Kawamura, C., Tsuge, T. & Furusawa, I. The *Alternaria alternata* melanin biosynthesis gene restores appressorial melanization and penetration of cellulose membranes in the melanin-deficient albino mutant of *Colletotrichum* lagenarium. **Fungal Genet. and Biol.**, 131-140 (1997).
106. Nelson, M. A. et al. Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa. *Fungal Genet Biol* **21**, 348-63. (1997).

107. Kurosawa, E. Experimental studies on the nature of the substance excreted by the "bakanae" fungus (in Japanese). *Trans. Nat. His. Soc. Formosa* **16**, 213-227 (1926).
108. Vigfusson, N. & Cano, R. Artificial induction of the sexual cycle of Neurospora crassa. *Nature* **249**, 383-385 (1974).
109. Graia, F. et al. Genome quality control: RIP (repeat-induced point mutation) comes to Podospora. *Mol Microbiol* **40**, 586-95. (2001).
110. Ikeda, K. et al. Repeat-induced point mutation (RIP) in Magnaporthe grisea: implications for its sexual cycle in the natural field context. *Mol Microbiol* **45**, 1355-64 (2002).
111. TIGR Microbial Database, http://www.tigr.org/tdb/mdb/mdbinprogress.html.
112. Fungal Genome Initiative, http://www-genome.wi.mit.edu/seq/fgi/. (2002).
113. Luo, Z., Freitag, M. & Sachs, M. S. Translational regulation in response to changes in amino acid availability in Neurospora crassa. *Mol Cell Biol* **15**, 5235-45 (1995).
114. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
115. Kelkar, H. S. et al. The Neurospora crassa genome: cosmid libraries sorted by chromosome. *Genetics* **157**, 979-90 (2001).
116. Aign, V., Schulte, U. & Hoheisel, J. D. Hybridization-based mapping of Neurospora crassa linkage groups II and V. *Genetics* **157**, 1015-20 (2001).