Applied Mycology and Biotechnology
Volume 3 and 4. Fungal Genomics pp. 295-313 (THIS PDF IS A PREPRINT)
D.K. Arora and G.G. Khachatourians (editors)

Genomics in *Neurospora crassa*: From One-Gene-One-Enzyme to 10,000 Genes

Edward L. Braun[a], Donald O. Natvig[b], Margaret Werner-Washburne[b] and Mary Anne Nelson[b]

[a]Department of Zoology, University of Florida, Gainesville, Florida 32611, USA
[b]Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

Corresponding author: Mary Anne Nelson, Department of Biology, MSC03 2020, University of New Mexico, Albuquerque, New Mexico 87131, USA, manelson@unm.edu

Neurospora crassa was the central organism in the development of biochemical genetics, providing a model system that established the relationship between genes and enzymes; it remains the best-studied filamentous fungus. This review focuses on the impact that the recent publication of a high-quality draft sequence of the *N. crassa* genome will have upon efforts to understand the biology of the filamentous fungi. Although several fungal genomes have been sequenced, annotated, and published, the organisms that have been examined are yeasts with relatively small genomes. In sharp contrast, *N. crassa* contains about 10,000 protein-coding genes, approximately twice as many genes as the yeasts and only slightly fewer than the invertebrate animals. Analysis of this gene set suggests that several different processes have led to the differences in gene content between *N. crassa* and the yeasts. Evidence for the loss of genes in the yeasts and the acquisition of novel genes in *Neurospora* lineage is described, as well as details regarding the biological processes that have led to these changes. Analyses of the *N. crassa* genome sequence revealed the widest array of genome defense mechanisms known for any organism, and one of these defense mechanisms (RIP) appears to have blocked the productive duplication of genes. Since gene duplication is the most common pathway for the origin of novel genes, it seems likely that *N. crassa* will provide an excellent model system for understanding alternative ways in which novel genes arise. A number of unexpected genes were identified when the complete genome sequence was analyzed, indicating that *N. crassa* produces secondary metabolites, shares apparent "pathogenicity" genes with plant pathogens, and responds to environmental cues such as light in novel ways. The genome sequence for *N. crassa* is the first exciting step toward a detailed understanding of the biology of filamentous fungi, and it will allow fungal biologists to establish which features of the filamentous fungi are shared with non-fungal organisms and which features are unique.

1. INTRODUCTION

N. crassa has a long and distinguished history in classical and biochemical genetics. C. L. Shear and B. O. Dodge discovered the sexual cycle of *Neurospora* and named the genus nearly eighty years ago (Shear and Dodge 1927). Less than ten years later, C. C. Lindegren identified by mutation and mapped six loci on linkage group (LG) I, which he referred to as the "sex

chromosome" of *Neurospora* (Lindegren *et al*. 1939). The first biochemical mutants in any organism were identified in *Neurospora* by G. W. Beadle and E. L. Tatum; their work initiated the science of biochemical genetics (Beadle and Tatum 1941). In 1945, Beadle proposed the one-gene-one-enzyme hypothesis (Beadle 1945). Beadle and Tatum were awarded the Nobel Prize for this body of work.

Also in 1945, B. McClintock studied the meiotic cytology of *Neurospora* and showed that there are seven chromosomes, the same as the number of linkage groups (McClintock 1945). A few years later, the first temperature-sensitive mutants were isolated in *Neurospora* (Houlahan *et al*. 1949). In 1955, M. B. Mitchell was the first to demonstrate the process of gene conversion (Mitchell 1955).

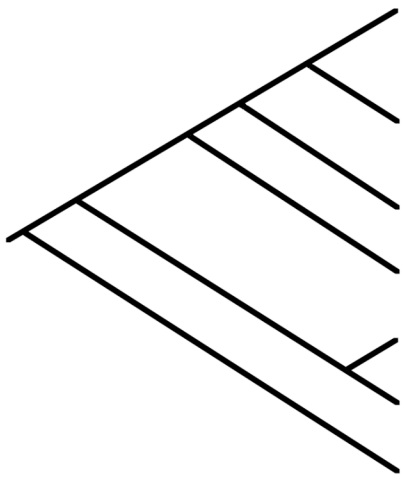
Since these early studies, a detailed genetic map of the seven linkage groups has been developed, including over 1,000 mapped genes; a fairly detailed restriction fragment length polymorphism (RFLP) map is also available (Perkins *et al*. 2001). Efficient DNA-mediated transformation is possible in *Neurospora* (Case *et al*. 1979), so that cloned genes can be introduced and their properties analyzed. The *Neurospora* genome sequence was completed and analyzed recently; this represents the first available sequence of a filamentous fungus (Galagan *et al*. 2003). Coordination of the physical (sequence) and genetic maps is greatly speeding gene discovery and characterization.

2. FUNGAL GENOMICS

In the era of genomics, one of the major questions is how many and which complete genome sequences will be necessary (Miller 2000). This is an especially important question for fungal biologists because of the diversity of fungi and the potential for these eukaryotic organisms to contribute to understanding fundamental, questions in biology. Some of the debate regarding the appropriate organisms to sequence has focused on how many genomes to obtain from representatives of the ascomycete fungi. Complete or nearly complete genome sequences now exist for three ascomycetes that exhibit predominantly yeast-like growth, *Saccharomyces cerevisiae* (Goffeau *et al*. 1996), *Candida albicans* (Berman and Sudbery 2002), and *Schizosaccharomyces pombe* (Wood *et al*. 2002). All three organisms have served as important models in cell biology and genetics.

The community of scientists who work on filamentous fungi have argued that additional complete genomes, especially genomes of filamentous fungi, are necessary because yeasts such as *S. cerevisiae* and *S. pombe* lack the complex life cycles typical of many filamentous fungi (Bennett 1997). The genome of *Ashbya gossypii*, a close relative of *S. cerevisiae* that possesses a type of filamentous growth, has also been sequenced (Brachat *et al*. 2003). However, *A. gossypii*

Fig. 1. Fungal phylogeny adapted from various sources (Bruns *et al*. 1992; Liu *et al*. 1999; Lutzoni *et al*. 2001; Keeling 2003). Although this phylogeny shows a zygomycete – microsporidia clade in the interest of simplicity, the zygomycetes are likely to be paraphyletic, representing multiple independent lineages branching between the Chytridiomycota and Basidiomycota (Keeling 2003). With the exception of the gene number from *N. crassa*, which reflects the annotation by Galagan *et al*. (2003), the numbers of genes present have been rounded and are based upon a variety of estimates. If gene numbers were unavailable, they were estimated using Kupfer *et al*. (1997) as a guide. There is little information available on genome size or gene number in the Chytridiomycota, but many chytridiomycetes exhibit substantial developmental complexity and exhibit features (e.g., flagella) absent in other fungi. For these reasons, at least some chytridiomycetes are likely to have large genomes and gene numbers similar to other developmentally complex fungi. The presence or absence of photobiology (responses to light and circadian rhythms) is shown as an example of a complex trait that has been lost in a variety of lineages. Although neither of the basidiomycetes shown here has a described photobiology, many basidiomycetes are known to exhibit responses to light.



Exemplar species:	Genome Size	Gene Number	Multicellular?	Photobiology Present?
Neurospora crassa	40	10,082	yes	yes
Magnaporthe grisea	40	10,000	yes	yes
Aspergillus nidulans	31	8,100	yes	yes
Saccharomyces cerevisiae	14	5,800	-	-
Candida albicans	16	7,000	-	-
Ashbya gossypii	8.5	4,000	yes	-
Schizosaccharomyces pombe	14	4,800	-	-
Phanerochaete chrysosporium	30	8,000	yes	-
Cryptococcus neoformans	18	7,000	-	-
Encephalitozoon cuniculi	2.5	2,000	-	-
Phycomyces blakesleeanus	30	8,000	yes	yes
Allomyces reticulatus	?	?	yes	yes

is a plant parasite that has fewer genes and a smaller genome than does *S. cerevisiae* (Fig. 1). Thus, the genome sizes and total gene numbers of these model ascomycetes are much smaller than those typical of most filamentous fungi (Fig. 1), suggesting that they lack certain gene functions important to more developmentally complex filamentous fungi and, therefore, may not be "models" for many critical processes.

It is clear that the absence of filamentous growth in the yeasts results in major differences, in terms of life history and ecology, between yeasts and filamentous fungi. The filamentous ascomycetes exhibit spectacular diversity in terms of morphology and ecology and include strict saprobes, facultative parasites, obligate parasites of plants and animals, mutualists of plants, and the fungal components of lichens (Alexopoulos *et al*. 1996). The yeasts also lack some specific responses to the environment, such as responses to light (Roenneberg and Merrow 2001), that are characteristic of many other fungi. Thus, there are important physiological questions to be answered by studying a variety of genomes within the ascomycetes.

A number of phylogenetic analyses (Bruns *et al*. 1992; Berbee and Taylor 1993; Liu *et al*. 1999) have suggested a model in which the yeasts arose from more complex filamentous ancestors, and some extremely simplified unicellular organisms are clearly derived from more complex fungal ancestors (Keeling 2003). The hypothesis that yeasts arose from multicellular ancestors by simplification suggests that developmentally complex filamentous fungi are likely to share genes that have been lost in the yeasts. In fact, analyses of *N. crassa* expressed sequence tags (ESTs) indicate that *S. cerevisiae* has undergone substantial gene loss (Braun *et al*. 2000), consistent with this hypothesis. However, rigorous tests of the hypothesis of genome simplification will require comparison of genomes from yeasts and diverse filamentous fungi.

The release of the *N. crassa* genome sequence (Galagan *et al*. 2003), representing the first complete genome sequence for a developmentally complex filamentous fungus, has revealed the presence of a number of surprising genes and opened the door to genomics of the filamentous fungi. In this chapter we review the impact of adding *N. crassa*, a model filamentous ascomycete, to the list of completely sequenced genomes and discuss the prospects for using this sequence information to investigate fungi and other groups of organisms.

3. *NEUROSPORA CRASSA* – A MODEL FILAMENTOUS FUNGUS

One of the major morphological differences between filamentous fungi like *N. crassa* and typical yeasts is the constitutive nature of hyphal growth. The fundamental cellular structure of filamentous fungi is the hypha (pl. hyphae), a multinucleate tube that grows by elongation from the apex (Heath and Steinberg 1999). This type of tip growth has only been found in the fungi and specific plant cells (root hairs and pollen tubes). Fungal hyphae typically have crosswalls called septa that divide the filament into a linear array of cells. Hyphal networks can remain relatively undifferentiated morphologically, as is typical of vegetative stages, or they can form very complex structures with highly specialized tissues, most notably in fungal fruiting bodies. Although the fruiting bodies (perithecia) of *N. crassa* are much smaller than the fruiting bodies of typical mushrooms or puffballs, they are complex multicellular structures that exhibit a characteristic shape (Nelson 1996).

Fungal hyphae allow a number of specialized functions that are not available to yeasts by providing a robust and intricate structure that supports continuous, directed growth. Hyphae form a network called the mycelium that gives filamentous fungi their structure. The mycelium of an individual fungus is a complex system in which the hyphae can transport nutrients and organelles, and these networks can exhibit trophic responses to a variety of signals such as

nutrients and light. Hyphae also have a great facility for penetrating substrates, reflecting the combination of physical forces and the release of digestive enzymes at the growing hyphal tip. In the case of pathogenic fungi, penetration through healthy tissues such as leaf surfaces often plays an important role in the initiation of infection (Gow *et al*. 2002). For saprobes like *N. crassa*, hyphal penetration is important in spreading a mycelial network through a substrate.

The mycelial network that results from the growth of hyphae is a true individual that can occupy a large space and live for extremely long periods of time. The most impressive examples of such individuals are found among basidiomycetes growing below ground, sometimes observed as fairy rings representing growth over hundreds of years starting from a single point. It has been suggested that the largest known living organism, covering 2,200 acres and with an estimated age of 2,400 years, is an individual of the basidiomycete *Armillaria bulbosa* (Smith *et al*. 1992). Mycelia also allow fungi to spread in and on living hosts; examples of this type of growth include plant pathogens, such as *Magnaporthe grisea* (rice blast fungus), and animal pathogens that can infect humans, such as *Trichophyton* spp. (ringworm).

3.1 *Neurospora* in the Environment

The life histories of *Neurospora* species are an interesting mix of specialization and generalization. A specialized aspect of *Neurospora* species is their adaptation to colonize dead or dying plants after fires, and most *Neurospora* isolates have been obtained from burned vegetation. Consistent with this observation, ascospores are routinely treated with either heat or furfuryl alcohol, a product of burned vegetation, in order to elicit germination in the laboratory. However, the generalist nature of *Neurospora* species is reflected in their lack of host specificity and their broad geographic ranges, which extend from the tropics to northern conifer forests (Perkins and Turner 1988; Jacobson *et al*. 2003).

Although *Neurospora* species are not pathogenic, they are likely to be excellent models for filamentous fungal pathogens, since closely related groups of perithecial ascomycetes include some of the most important plant pathogens (e.g., *Magnaporthe grisea*, the rice blast fungus). It is clear that *Neurospora* species have evolved in close association with plants and have life histories dependent on plants. The availability of the *N. crassa* genome presents an opportunity to assess whether plant pathogenicity and specialization for a narrow host range depend primarily on the acquisition of new genes or the modification of gene function and expression during the course of evolution.

3.2 The *Neurospora* Genome and its Impact on Fungal Genomics

In the genomic era, there are two approaches to the study of model organisms. One is to examine aspects of biology that are conserved in large portions of the "tree of life". Many experimental studies with members of the genus *Neurospora* have focused on fundamental aspects of cell biology and genetics likely to have such relevance (Davis and Perkins 2002), as have studies using yeast (*S. cerevisiae* and *S. pombe*) model systems. A second approach is to examine processes that are unique to specific groups of organisms, toward the goal of understanding processes that are unique to fungi or specific groups of fungi. Such studies will be extremely important to the understanding and control of fungal pathogens, for example. In this context, studies of hyphal growth and sexual development in *Neurospora* species have been responsible for a number of important contributions with special relevance to fungal biology. The complete sequence of the *N. crassa* genome will accelerate both types of research, but the

unique features of *N. crassa* as a model filamentous fungus are expected to have an especially significant impact on the second research approach.

As a model filamentous fungus, *N. crassa* also brings with it a considerable body of information regarding the population biology and evolution of the genus, and progress has been made in understanding the ecology of certain species (Jacobson *et al*. 2003). When the information on the natural history of *Neurospora* species is combined with past laboratory studies (reviewed by Davis and Perkins 2002) and the new genome sequence (Galagan *et al*. 2003), it should be possible to build a solid foundation for understanding how filamentous fungi respond to the environment at a molecular level. The *N. crassa* genome will provide a reference for comparison with other filamentous fungi exhibiting different lifestyles, including pathogens. Overall, the availability of the genome sequence from *N. crassa* should greatly accelerate the expansion of experimental biology into realms with specific relevance to fungal biology.

4. COMPARATIVE GENOMICS USING *NEUROSPORA CRASSA*

Analyses of the complete genome sequence for *N. crassa* revealed 10,082 protein-coding genes (Galagan *et al*. 2003), in agreement with previous predictions of gene number based upon more limited samples of genomic DNA that ranged from 9,200 to 13,000 (Kupfer *et al*. 1997; Nelson *et al*. 1997; Bean *et al*. 2001; Kelkar *et al*. 2001; Schulte *et al*. 2002; Mannhaupt *et al*. 2003). Thus, it was clear before the completion of the *N. crassa* genome that this filamentous fungus would have a substantially larger number of genes than the 4500 to 6000 genes that are typical of the yeasts (Goffeau *et al*. 1996; Wood *et al*. 2001; Wood *et al*. 2002; Kellis *et al*. 2003; also see Fig. 1).

There are three distinct processes that can account for the differences in gene numbers between the yeasts and the filamentous fungi (Fig. 2). One process, gene loss in the yeast lineages, would reduce the sizes of the yeast genomes. In contrast, two alternative processes, the creation of novel genes and lateral gene transfer (LGT) into the *Neurospora* lineage, would lead to an increase in the number of genes present in the *Neurospora* lineage. Prior to the completion of the *N. crassa* genome sequence, large-scale sequence comparisons had provided evidence that all of these processes have contributed to the differences in gene content between the *N. crassa* and *S. cerevisiae* genomes (detailed by Braun *et al*. 2000). The availability of the complete *N. crassa* genome sequence has allowed these basic conclusions to be expanded in several different ways.

4.1 Genetic Innovation in the *Neurospora* Lineage

Genetic innovation can be broadly defined as the origin of novel genes in specific lineages. Gene duplication is thought to represent the primary explanation for the process of genetic innovation in most lineages (Ohno 1970), and a small number of gene families are evident in the complete *N. crassa* genome sequence (Galagan *et al*. 2003). However, the proportion of *N. crassa* genes that are members of gene families is substantially lower than expected for an organism with approximately 10,000 genes (Table 1), and most of the gene duplications observed in *N. crassa* are relatively ancient. Although there was existing evidence that *N. crassa* had few gene families (Nelson *et al*. 1997), the striking near absence of gene families with closely-related members in the *N. crassa* genome only became clear upon analysis of the complete genome sequence (Galagan *et al*. 2003).

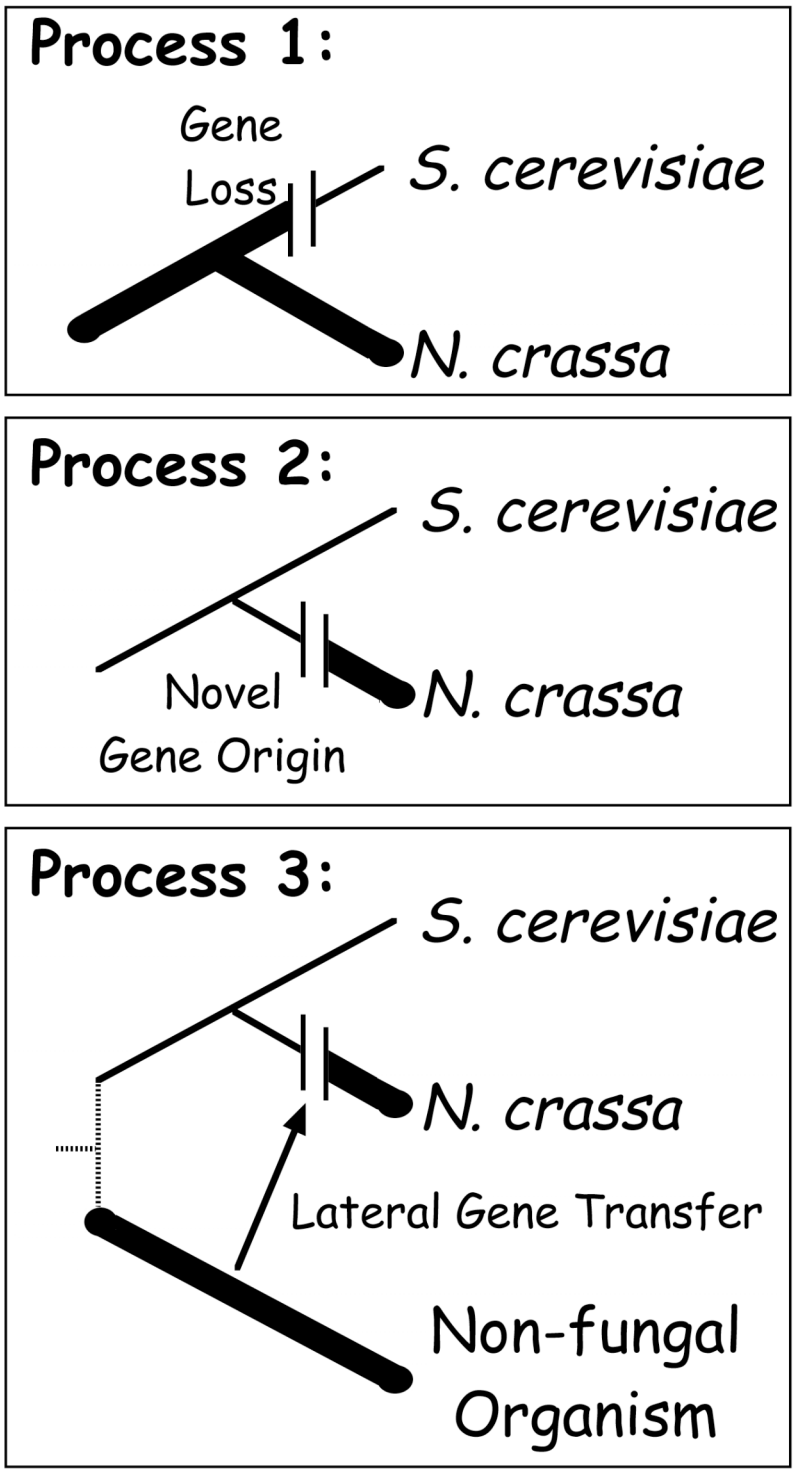


Fig. 2. Schematic of three processes that could give rise to the higher gene numbers observed in *N. crassa* compared to *S. cerevisiae*. The presence or absence of a specific gene is indicated using the thickness of the line in the evolutionary tree. Similar processes can also be invoked to explain the differences in gene numbers between *N. crassa* and fungi like *S. pombe* or *E. cuniculi*. Evidence for the action of all three processes has been obtained (see text).

Table 1. Percentage of genes in various organisms that are members of gene families

Organism	Number of Genes[a]	% Genes in Families[b]	Expected % Genes in Families[c]
Mycoplasma genitalium	500	16%	—
Haemophlius influenzae	1,700	17%	18%
Pyrococcus abyssi	1,800	28%	20%
Encephalitozoon cuniculi	2,000	21%	22%
Escherichia coli	4,400	33%	35%
Schizosaccharomyces pombe	4,800	26%	36%
Pseudomonas aeruginosa	5,600	44%	38%
Saccharomyces cerevisiae	5,800	33%	38%
Neurospora crassa	10,000	17%	41%
Drosophila melanogaster	14,500	39%	43%
Caenorhabditis elegans	18,000	46%	43%
Arabidopsis thaliana	25,000	71%	44%

[a] Due to the problems associated with gene prediction (Wood *et al*. 2001; Braun 2003; Kellis *et al*. 2003), these values have been rounded to the nearest hundred.

[b] These data are from analyses conducted for Galagan *et al*. (2003) and presented in Fig. 1 and S6.1 from that paper.

[c] The expected percentage of genes in families was calculated using expected number of paralogs (n_P) given a specific number of genes (n_G) using the equation $n_P = 0.46n_G - 471$ (from Hooper *et al*. 2003) and converting this value to a percentage. This equation does not yield an appropriate prediction for organisms with fewer than 1,000 genes, so the value for *M. genitalium* is not presented.

The absence of genes that arose by recent duplications probably reflects the activity of a process in *Neurospora* called RIP (repeat-induced point mutation), which causes multiple C to T transition mutations on both strands of duplicated sequences (Cambareri *et al*. 1989). RIP is one of the set of processes that operate as genome defense mechanisms in *Neurospora* species; additional defense mechanisms include the vegetative quelling of repeated sequences (Cogoni 2001) and the meiotic silencing of unpaired DNA (MSUD; Shiu *et al*. 2001). These genomic defense mechanisms have left a clear signature on the complete *N. crassa* genome sequence, with the extreme paucity of gene families. Analysis of the genome sequence has revealed that RIP has affected all transposable elements and the majority of other duplicated sequences present in the genome (Galagan *et al*. 2003). The notable exceptions are the ribosomal DNA repeat units present at the nucleolus organizer, which for reasons that remain unknown have not been affected by RIP. However, the virtual absence of recent protein-coding gene duplications in the *N. crassa* lineage is undoubtedly another consequence of RIP.

The complete *Neurospora* genome sequence also allowed examination of alternative hypotheses regarding processes involved in genetic innovation. Although it seems clear that the fixation of duplicated genes in the *Neurospora* lineage stopped after the process of RIP arose, detailed understanding of the RIP process may provide clues about types of duplication that can occur despite the existence of RIP. For example, duplications of relatively short genes or gene segments might occur, since RIP does not act on duplicated sequences unless they exceed a certain minimum size (~400 bp; Watters *et al*. 1999). Likewise, duplication coupled with loss of a segment of the gene (e.g., due to domain loss; Braun and Grotewold 2001) might allow

survival of both duplicates if the region exhibiting a high degree of nucleotide sequence identity is shorter than 400 bp and thus not susceptible to the process of RIP. The fixation of duplicates by any of these alternative pathways should still result in the duplication of individual protein domains. Examination of the complete proteome inferred using the *N. crassa* genome sequence revealed few closely related domains (Galagan *et al*. 2003), suggesting that these pathways to gene duplication are uncommon and that RIP has stopped the productive duplication of genes rather than altering the pathways that result in duplicated genes.

This leaves unanswered the major question of how *Neurospora* accumulated so many genes (prior to the establishment of the RIP process). The majority of *N. crassa* sequences that lack homologs in the yeast genomes (57% of predicted *N. crassa* proteins) correspond to "orphan" genes, defined as genes that lack identifiable homologs in available databases (Nelson *et al*. 1997; Braun *et al*. 2000; Schulte *et al*. 2002; Galagan *et al*. 2003; Mannhaupt *et al*. 2003). The number of true orphan genes will likely be reduced substantially when full comparisons are made between the genomes of *N. crassa* and closely-related ascomycetes. In fact, comparisons of *N. crassa* genes on linkage groups 2 and 5 to an assembly of the *Aspergillus fumigatus* genome (available at http://www.tigr.org/tdb/e2k1/afu1) greatly reduced the percentage of genes classified as orphans (Mannhaupt *et al*. 2003). However, it is quite clear that *N. crassa* possesses many genes with no identifiable homologs in other groups of organisms.

What processes could generate such a large proportion of orphan genes? In principle, orphan genes can be generated by duplication followed by rapid divergence of one member of the duplicated pair. However, the active genomic defense mechanisms of *Neurospora* species have blocked the productive duplication of genes. This raises the question of when the orphan genes present in the *N. crassa* genome arose. If many orphans arose prior to the origin of RIP, then they might reflect duplication followed by rapid divergence. Alternatively, if the orphan genes arose after the origin of RIP and they reflect rapid evolutionary divergences, they would have to be divergent orthologs of genes present in other organisms. In this model, the divergence would not be coupled with duplication, so *N. crassa* would lack less divergent copies of the genes. Since the extreme sequence divergence necessary to obscure evolutionary relationships among protein-coding genes is likely to be correlated with functional changes (Aravind *et al*. 2000; Braun 2003), this pathway seem implausible as a major pathway of innovation.

Assuming a limited number of genes have undergone divergence in the absence of duplication, only two pathways remain for genetic innovation after the evolution of RIP: 1) lateral (horizontal) gene transfer (LGT); and 2) "overprinting," a process defined as the generation of novel genes from noncoding sequences (Fig. 2; for details see Ohno 1984; Keese and Gibbs 1992). LGT would not tend to produce orphans unless there were no sequences available from the donor organism, and LGT into the *Neurospora* lineage appears relatively limited (Braun *et al*. 2000; also see Section 4.2 below), suggesting that it does not represent a major source of orphan genes. Overprinting, therefore, appears to be a possible explanation for the high proportion of orphan genes if many of the orphan genes arose after the evolution of RIP. Although the source of the requisite unexpressed ORFs that represent the raw material for overprinting remains obscure, the existence of specific processes that block the fixation of duplicated genes (e.g., RIP) should make *Neurospora* species excellent model systems to examine this process. Alternatively, if the orphans largely reflect duplication prior to the origin of RIP followed by divergence, then other ascomycetes that diverged after the origin of RIP should have the a set of genes very similar to that found in *N. crassa*. Since the rice blast fungus *M. grisea* is known to RIP duplicated sequences (Ikeda *et al*. 2002) and a *M. grisea* sequence

assembly is virtually complete (available from http://www-genome.wi.mit.edu/annotation/fungi/magnaporthe/), comparison of *N. crassa* sequences to this organism will be especially informative. Regardless of the evolutionary origin of the orphan genes in *Neurospora*, the observation that ESTs from specific developmental stages in *N. crassa* differ in the proportion of orphan genes (Nelson *et al*. 1997) is likely to provide a means to connect these unknown sequences with specific aspects of fungal biology.

4.2 Lateral Transfer into the *Neurospora* Genome

The explosion of genome sequence data has revealed surprising numbers of genes that are shared among genomes by the process of LGT (Doolittle 1999). As described above, the absence of productive gene duplication in the *Neurospora* lineage suggests that the origin of novel genes in this lineage after the origin of RIP has been limited largely to the poorly understood process of overprinting and to LGT. Although the impact of LGT on the evolution of prokaryotic genomes is fairly clear (Doolittle 1999), the number of genes that originate by LGT in eukaryotic lineages has been more controversial (Doolittle 1998; Salzberg *et al*. 2001; Braun 2003; Gogarten 2003). There have been suggestions that large numbers of eukaryotic genes originated by lateral transfers from prokaryotes (Doolittle 1998) and that the clustering of functionally related genes evident in some fungi reflects LGT (Prade *et al*. 1997). However, there are also reasons to suspect that the probability of prokaryote-to-eukaryote LGT is substantially lower than the probability of LGT among prokaryotes (Braun 2003). Likewise, the observed clustering of functionally related genes in the fungi could reflect processes other than LGT.

The existence of barriers to prokaryote-to-eukaryote LGT have been supported by phylogenetic analyses of many candidates for prokaryote-to-vertebrate LGT (Roelofs and Van Haastert 2001; Salzberg *et al*. 2001; Stanhope *et al*. 2001). These candidates for LGT from various prokaryotes to vertebrates were identified using the initial annotation of the draft sequence of the human genome (The International Human Genome Sequencing Consortium 2001) as those genes with possible orthologs in prokaryotic genomes (Salzberg *et al*. 2001). However, genes with this phylogenetic distribution could also reflect loss in multiple non-human eukaryotic lineages. In fact, the observed number of genes with this distribution is similar to that expected under a simple model of gene loss (Salzberg *et al*. 2001), so gene loss in several eukaryotic lineages represents a simpler explanation for these genes than does prokaryote-to-vertebrate LGT. Despite these results, the barriers to prokaryote-to-eukaryote LGT that exist are probably not absolute. Instead, they probably act to reduce the probability of prokaryote-to-eukaryote LGT. Thus, establishing an upper limit for the impact of prokaryote-to-eukaryote LGT upon the *N. crassa* lineage represents a useful exercise. Using a criterion to identify candidates for prokaryote-to-eukaryote LGT comparable to that used by the International Human Genome Sequencing Consortium (presence of a homolog only in the focal eukaryote and prokaryotes), we found 95 candidates for LGT from a prokaryotic organism to *N. crassa* (0.9% of the *N. crassa* proteome). If we relax this criterion slightly to include genes with substantially better BLASTP hits to prokaryotes than to eukaryotes (using a 10 log difference in *E*-values as a cut-off), then an additional 297 genes would be included (2.9% of the *N. crassa* proteome).

Estimates of the numbers of *N. crassa* genes that arose by LGT from prokaryotes using the sets of protein-coding genes that are present in prokaryotes and absent in other eukaryotes should exclude instances of fungus-to-prokaryote LGT, which would result in the same distribution of homologs. Unfortunately, accurate estimates of the proportions of genes reflecting each

direction of LGT are difficult to obtain without conducting detailed phylogenetic analyses. It may be necessary to include sequences from additional filamentous fungi in these analyses, and they would have to involve rooted phylogenetic trees despite the difficulties associated with establishing the position of the root in phylogenetic trees (Bieszke *et al*. 1999). Given these limitations, the values presented in this chapter should be viewed as upper limits for the number of *N. crassa* genes originating by LGT from prokaryotes.

The current lack of genome sequences from additional filamentous fungi makes it impossible to determine whether LGT among fungi has made a substantial contribution to the *N. crassa* genome. However, some aspects of the potential contribution of LGT within the fungi to the *N. crassa* genome can be examined by searching for clusters of functionally related genes, since it has been suggested that LGT may drive the clustering of certain genes in the fungi (Prade *et al*. 1997). By analogy with the "selfish operon" model (Lawrence and Roth 1996), sets of genes that carry out similar biological functions could be transferred together in tightly-defined clusters, while LGT involving individual genes would not result in transfer of the complete pathway. This phenomenon has been suggested to largely involve nonessential genes, such as those involved in secondary metabolism (Prade *et al*. 1997). Some clusters of functionally-related genes are evident in *N. crassa*, including the well-studied *qa* cluster (Geever *et al*. 1989) and a cluster of putative laccase and melanin biosynthetic genes on linkage group 5 (Mannhaupt *et al*. 2003). However, various analyses of the complete *N. crassa* genome did not reveal large numbers of clusters characterized by the presence of functionally-related genes (Galagan *et al*. 2003).

There are several additional reasons for the clustering of functionally-related genes in the fungi and other eukaryotes. These include the presence of shared regulatory elements (Rosewich and Kistler 2000), the existence of chromosomal domains containing multiple genes with similar patterns of expression (Cohen *et al*. 2000), and the possibility that the clustering may be an ancestral feature of the genes in question. For example, the *snz* and *sno* genes responsible for pyridoxine synthesis in some organisms are present in a cluster in *N. crassa*, *S. cerevisiae*, and the marine sponge *Suberites domuncula* (Padilla *et al*. 1998; Bean *et al*. 2001; Seack *et al*. 2001), and the divergence observed among the *snz* and *sno* homologs in these organisms is consistent with divergence due to speciation. Although it is likely that the selective pressure maintaining the *snz – sno* cluster in these diverse organisms is the presence of regulatory elements that drive the expression of these genes under the same conditions (Padilla *et al*. 1998), the evolutionary origin of this clustering appears to be ancient and unrelated to LGT. Thus, these analyses only serve to place an upper limit on LGT within the fungi. Nonetheless, they serve to emphasize the relatively limited contribution of LGT to the *N. crassa* genome, especially when they are combined with the limited number of genes potentially reflecting prokaryote-to-eukaryote LGT in *N. crassa*.

4.3 Gene Loss in the Yeasts

The initial analyses of the complete *N. crassa* genome (Galagan *et al*. 2003) revealed a total of 584 *N. crassa* genes that have matches in other groups of eukaryotes but lack a clear homolog in either of the completely sequenced yeasts (*S. cerevisiae* and *S. pombe*). These genes are likely to have been lost in both of the yeasts, suggesting that about 5% of the ancestral proteome was lost in both yeasts, assuming the ancestral proteome for the ascomycetes was similar in size to the *N. crassa* proteome. Although the small number of complete genome sequences currently available for fungi limits our ability to reconstruct the presence or absence of specific gene

products in the ancestral fungal proteome, the hypothesis that the yeasts arose from more complex filamentous ancestors (Bruns *et al*. 1992; Berbee and Taylor 1993; Liu *et al*. 1999) and the existence of developmentally complex fungi in all groups of fungi (Fig. 1) are consistent with the common ancestor of the ascomycetes having a relatively large proteome.

To extend the initial comparative analyses of the *N. crassa* genome, we examined the numbers of genes with distributions suggesting loss in the yeasts. Using strict criteria outlined in previous studies of gene loss (a significant hit in the non-fungal database [*E*-value $\leq 10^{-5}$] and no potential homolog in the yeast databases [*E*-value > 0.1]; Braun *et al*. 2000; Braun 2003), we found a slightly smaller number of candidates for gene loss (301 *N. crassa* queries). This smaller number of genes identified may reflect the loss of some sequences by restricting our consideration to eukaryotes for which complete and extensively annotated genome sequences are available, but it is more likely to reflect the restrictive criterion used to score loss (sequences must have no potential homolog in BLASTP searches of either yeast). If we consider *N. crassa* genes with homologs in the yeasts but non-fungal homologs that are substantially more closely related (the non-fungal homolog is at least 10-logs better than the yeast homolog; Braun *et al*. 2000; Braun 2003), then 438 additional candidates for loss can be identified. These sequences are expected to include genes that have undergone accelerated evolution in the yeasts, so the set of protein-coding genes present in *N. crassa* that were lost in both of the completely-sequenced yeasts falls somewhere between 301 and 739 genes (approximately 3% – 8% of the proteome).

Although detailed analyses of the *N. crassa* genes that have been lost in yeasts will be presented elsewhere (manuscript in preparation), several patterns emerge from a consideration of the most conservative set of 301 protein-coding genes. Using the same criteria to score loss revealed the potential loss of 535 genes in *S. cerevisiae* and 499 genes in *S. pombe*. Thus, more than half of the genes that have been lost in a single yeast species have been lost in both yeast species. The loss of these genes would have occurred independently in each yeast lineage if the fungal phylogeny presented in Fig. 1 is correct, as suggested by a variety of analyses (Bruns *et al*. 1992; Berbee and Taylor 1993; Liu *et al*. 1999; Lutzoni *et al*. 2001; Vivarès *et al*. 2002). However, since some phylogenetic analyses support a *S. pombe – S. cerevisiae* clade (Bullerwell *et al*. 2003), it remains possible that these genes were lost in the common ancestor of *S. cerevisiae* and *S. pombe*. If the loss in each yeast lineage did occur independently, the large size for the set of genes lost in both lineages relative to the sets of genes lost in each lineage suggests that certain genes may be more likely to undergo loss in the yeasts. As complete genome sequences become available from additional yeasts that are distantly related to *S. cerevisiae* and *S. pombe*, such as *Pneumocystis carinii* (a basal ascomycete) and *Cryptococcus neoformans* (a basidiomycete yeast), this question may be resolved.

In addition to the ascomycete yeasts, the complete genome sequence of one other fungus is currently available. The microsporidian *Encephelitozoon cuniculi* is an obligate intracellular parasite with a very small genome (2.9×10^6 bp) originally thought to be an ancient eukaryote but now recognized as a highly reduced fungus (Keeling and Fast 2002; Vivarès *et al*. 2002). More recent phylogenetic analyses suggest a specific relationship between the microsporidia and certain zygomycetes (Keeling 2003), indicating that the extreme reduction of the microsporidia occurred independently of the more modest reduction that occurred in the ascomycete yeasts. Not surprisingly, the set of *N. crassa* genes that have non-fungal homologs but lack *E. cuniculi* homologs is very large (2334 protein-coding genes). However, the set of *N. crassa* genes with non-fungal homologs that are clearly absent in the genomes of *S. cerevisiae*, *S. pombe*, and *E. cuniculi* is only 284 genes. Thus, there are 17 genes that were likely to have been lost in both of

the ascomycete yeasts that were retained in the genomes of the highly reduced microsporidian *E. cuniculi*. The degree to which these genes share similar functions is unclear, although it is surprising that genes retained in a highly-reduced genome would be dispensable in the yeasts. It is possible that the functions of these genes are mediated by non-orthologous genes in the yeasts (non-orthologous displacement; Koonin *et al*. 1996). Alternatively, these genes may provide information about the differences in the selective pressures that have acted to reduce the gene number and genome size in the free-living yeasts and parasitic microsporidia.

5. UNEXPECTED GENES IN THE *NEUROSPORA* GENOME SEQUENCE

One of the most exciting aspects of genomic sequencing is the potential to find genes that were completely unanticipated in the organism. The *Neurospora* genes that were lost in *S. cerevisiae* (Braun *et al*. 1998; Braun *et al*. 2000) might fall into this category, since it is possible to find many papers in molecular biology journals that assert specific features are universal to the eukaryotes because they are "conserved from yeast to man." However, the availability of multiple eukaryotic genome sequences has made the contribution of gene loss to eukaryotic genome evolution clear (Salzberg *et al*. 2001; Braun 2003), so these genes should not be considered surprising. Nonetheless, the completion of the *N. crassa* genome sequence did reveal a number of genes that were completely unanticipated despite the long history of experimental biology using *Neurospora* species.

5.1 Genes Associated with Secondary Metabolism

Despite the absence of characterized secondary metabolites synthesized by *Neurospora* species, the earliest large-scale sequence data sets revealed the existence of specific genes involved in the synthesis of secondary compounds (Nelson *et al*. 1997). Subsequent analyses revealed a number of genes involved in the synthesis and transport of compounds ranging from trichothecene, lovastatin, aflatoxin, and penicillin in other fungi (Mannhaupt *et al*. 2003). Thus, the diversity of genes involved in secondary metabolism identified using the complete genome sequence (Galagan *et al*. 2003) was consistent with previous large-scale sequence data sets, but they had been unanticipated based upon the biology of *Neurospora* species. A possible resolution of the apparent paradox of finding secondary metabolism genes in an organism may be the production of relatively small amounts of these compounds, possibly as signalling molecules.

One particularly interesting class of secondary metabolism genes found in the complete *N. crassa* genome includes those associated with the biosynthesis of diterpenes and, in particular, gibberellins in other organisms. Gibberellins are best known as regulators of stem elongation and other developmental processes in plants (Graebe 1987; Yamaguchi and Kamiya 2000). However, gibberellic acid (GA) was first identified as a metabolic product of the plant pathogen *Gibberella* (*Fusarium*) *fujikuroi*, a pyrenomycete relative of *N. crassa* that causes "foolish seedling" disease in rice. GA was shown to be responsible for the over elongation of shoots (preceding seedling death), a phenomenon from which the disease received its name (Kurosawa 1926; Yabuta and Sumiki 1938). GA was only later shown to be a normal growth regulator in plants.

Plants and *G. fujikuroi* share certain steps in GA synthesis but differ in others (Hedden *et al*. 2001). The proteins predicted for *N. crassa* include at least one member of each of three enzyme classes required for gibberellin biosynthesis in plants (Yamaguchi and Kamiya 2000), as well as homologs of all enzymes required for GA biosynthesis in *G. fujikuroi*. The enzymes encoded by

the *N. crassa* genes include an apparent terpene synthase (aka terpene cyclase; NCU09272.1), several members of the cytochrome P450 monooxygenase family closely related to enzymes involved in gibberellin biosynthesis in both plants (NCU02852.1) and fungi (NCU05376.1, NCU09274.1, NCU05967.1), three apparent homologs of GA4 desaturase from *G. fujikuroi* (NCU00751.1, NCU01598.1, NCU00847.1), and others. (Numbers in parentheses are annotation numbers associated with the *N. crassa* genome sequence reported by Galagan *et al*. 2003.)

Although these proteins have unknown functions in *N. crassa*, their presence has important implications for the evolution of pathogenicity in *G. fujikuroi* and other fungi, because it indicates that the components necessary for GA production were present in the fungal ancestors of pathogens and non-pathogens alike. Quite clearly, *G. fujikuroi* has evolved to use these genes in a special manner in the context of pathogenicity. This is evident in that in *G. fujikuroi* the genes encoding enzymes for GA synthesis reside in a cluster (Tudzynski and Holter 1998), whereas the *N. crassa* homologs do not. It is possible that *N. crassa* and its non-pathogen relatives synthesize gibberellins or related compounds for reasons unrelated to plant pathogenesis, possibly as signals during development. In fact, there is a report of GA3 from *N. crassa* (Kawanabe *et al*. 1983), as well as evidence that GA has effects on hyphal growth in this organism (Tomita *et al*. 1984). In this context, the growing availability of sequence data from the fungi may allow us to determine whether the clustering of GA biosynthetic enzymes in *G. fujikuroi* reflects LGT within the fungi, selection for a novel pattern of expression for a set of genes common to many pyrenomycetes, or a distinct selective pressure.

Alternatively, it is entirely possible that in *N. crassa* these genes have functions unrelated to GA biosynthesis. Throughout the fungal and plant kingdoms, the enzyme families in question participate in the synthesis of large numbers of secondary compounds, including certain antibiotics. In conifers, for example, the terpene synthase family includes a diverse group of enzymes, some with roles in GA synthesis and others that participate in the synthesis of terpene components of resins (Phillips and Croteau 1999; Trapp and Croteau 2001). In addition, specific enzymes can be both multifunctional (catalyzing multiple steps in a pathway) and promiscuous in terms of substrate (Yamaguchi and Kamiya 2000). Therefore, it is not possible to conclude precisely what roles the *N. crassa* enzymes have in diterpene-related secondary metabolism. Regardless of the specific functions of these enzymes, the discovery of genes related to terpene metabolism in *N. crassa* provides strong support for additional studies in this organism, focused on the possible roles of secondary compounds in defense and signaling.

5.2 Genes Associated with Pathogenicity in other Fungi

Neurospora species are not plant or animal pathogens, and they are not known to produce mycotoxins (Perkins and Davis 2000). However, analysis of the complete genome sequence revealed 12 genes associated with pathogenicity in other fungi (Galagan *et al*. 2003). Surprisingly, these potential pathogenicity-related genes include one of the few gene families present in *N. crassa*, with a total of five genes showing homology to the ECP2 gene of the tomato pathogen *Cladosporium fulvum*. ECP2 is a secreted protein that elicits the hypersensitive response and necrosis in some plants (Wubben *et al*. 1994; Takken *et al*. 2000), but its function in *Neurospora* is currently unclear. Some of the genes involved in secondary metabolism (see above, section 5.1) also play a role in the pathogenic phenotype of certain fungi, and their presence in *N. crassa* was unanticipated. Establishing the functions of these pathogenicity-

related genes in a non-pathogenic group like *Neurospora* presents both challenges and exciting opportunities to learn about the ways in which pathogenic and non-pathogenic fungi differ.

5.3 Genes Associated with Light Sensing

Although the existence of circadian rhythms and blue-light responses in members of the genus Neurospora has been known for some time (Loros and Dunlap 2001), large-scale sequencing in *N. crassa* has revealed several surprising genes involved in light sensing. All the known light-induced phenotypes in *Neurospora* are regulated by blue light, and all of these phenotypes can be blocked in mutants of either *white collar-1* (*wc-1*) or *white collar-2* (*wc-2*; (Lee *et al*. 2003). The products of the *wc* genes form a flavin mononucleotide-containing complex and act as a photoreceptor (Froehlich *et al*. 2002; He *et al*. 2002). Since the product of *wc-1* is a photoreceptor, and all known responses to light in *Neurospora* are abolished by the *wc* mutations, the potential roles of additional photoreceptors is unclear and their existence was unexpected.

Additional photoreceptors encoded in the *N. crassa* genome include homologs of the archaeal rhodopsins, cryptochromes, and phytochromes (Bieszke *et al*. 1999; Galagan *et al*. 2003). The archaeal rhodopsin homolog (*nop-1*) was identified using EST data before completion of the *N. crassa* genome sequence (Bieszke *et al*. 1999), but it also represents an unexpected gene revealed by large-scale sequencing. Exposure of *nop-1* mutants to the mitochondrial ATPase inhibitor oligomycin produces a light-dependent change in growth pattern (Bieszke *et al*. 1999), although it is unclear whether the change in growth and conidiation reflects an indirect effect of oligomycin or redundancy between the product of *nop-1* and the mitochondrial H+-ATPase for a light-regulated conidiation function. Regardless, these data suggest the existence of light-regulated processes in *Neurospora* that were not anticipated based upon the results of genetic screens.

The presence of phytochromes in *Neurospora*, revealed by analyses of the complete genome sequence (Galagan *et al*. 2003), is also unexpected because all known responses to light in *Neurospora* involve blue-light. However, the bacteriophytochromes related to the *N. crassa* phytochrome homologs and the *Aspergillus nidulans velvet* gene are involved in responses to red light. In fact, *N. crassa* also has a clear ortholog of the *A. nidulans fluG* gene (NCU04264.1), which is known to functionally interact with the *velvet* gene (Yagera *et al*. 1998). Establishing whether the genes that are involved in responses to red light in other organisms that are present in *Neurospora* regulate additional light regulated processes (that were not detected by genetic screens), or whether they interact functionally with the products of the *wc* genes should prove interesting.

6. CONCLUSIONS

The completion of the *N. crassa* genome sequence represents a landmark in fungal genetics and genomics. This event constitutes a fundamental step in the remarkable journey that has taken the biology of *Neurospora* from the experiments of Beadle and Tatum that established a direct relationship between genes and proteins to a preliminary description of *Neurospora*'s 10,000 genes. A number of challenges remain for the future, most notably that of establishing the ways in which these 10,000 genes interact to product the developmental complexity evident in *Neurospora*. However, the availability of the complete genome sequence of *N. crassa* has allowed us to understand more precisely the ways in which this organism differs from the yeasts.

The first fungi for which complete genome sequences were available (the yeasts *S. cerevisiae* and *S. pombe* and the microsporidian *E. cuniculi*) are organisms that have undergone substantial reduction by gene loss during evolution. Thus, it is not surprising that gene loss in the yeast lineages has made a substantial contribution to the differences between the genomes of these organisms and the genome of *N. crassa*. However, loss in the yeast lineages cannot explain all of the differences between the genomes of *Neurospora* species and the yeasts. Instead, some type of genetic innovation must be invoked to explain the large number of genes present in *N. crassa*. Gene duplication, the most common pathway of genetic innovation, has been blocked in the *Neurospora* lineage, and other pathways of genetic innovation such as LGT and partial duplications do not appear to be especially active in the *Neurospora* lineage. Thus, either much of the innovation occurred prior to the origin of the processes that block gene duplication (e.g., RIP), or pathways of innovation that are poorly understood at present are responsible for many of the differences between *N. crassa* and the yeasts.

The difficulties associated with the analysis of the *N. crassa* genome raise the question of to what extent comparative genomics will ultimately help unravel the nature of fungal diversity. Our analysis of the *N. crassa* genome suggests that much, if not most, of the genetic framework required for this diversity is reflected in the genome of this one organism. The genome sequence of *N. crassa* has revealed genes thought to have functions that *Neurospora* was not known to possess, particularly in the realms of secondary metabolism, interactions with plants, and the transduction of signals from the environment. The *N. crassa* genome may well have a nearly complete complement of the gene families present among the filamentous fungi. This presents both opportunities and challenges for experimental biologists, because it implies that major differences in biology may result from small changes in gene structure and expression.

In the eighty years since Shear and Dodge named the genus *Neurospora*, experiments using this organism established the fundamental relationship between genetics and biochemistry. With the publication of the *N. crassa* genome we now have the tools to elucidate the nature of the subtle changes in genetic function that have led to the diversity of phenotypes evident in the filamentous fungi. We believe this will ultimately lead to another exciting eighty years of *Neurospora* research.

Acknowledgements. This research was supported by the NSF Grant MCB-9874488 to M.A.N.

REFERENCES

Alexopoulos CJ, Mims CW, and Blackwell M (1996). Introductory Mycology. New York: John Wiley & Sons.

Aravind L, Watanabe H, Lipman DJ, and Koonin E (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci USA 97:11319-11324.

Beadle GW (1945). Biochemical genetics. Chem Revs 37:15-96.

Beadle GW and Tatum EL (1941). Genetic control of biochemical reactions in *Neurospora*. Proc Natl Acad Sci USA 27:499-506.

Bean LE, Dvorachek WH, Braun EL, Errett A, Saenz GS, Giles MD, Werner-Washburne M, Nelson MA, and Natvig DO (2001). Analysis of the *pdx-1* (*snz-1/sno-1*) region of the *Neurospora crassa* genome: Correlation of pyridoxine-requiring phenotypes with mutations in two structural genes. Genetics 157:1067-1075.

Bennett JW (1997). White paper: Genomics for filamentous fungi. Fungal Genet Biol 21:3-7.

Berbee ML and Taylor JW (1993). Dating the evolutionary radiations of the true fungi. Can J Bot 71:1114-1127.

Berman J and Sudbery PE (2002). *Candida albicans*: A molecular revolution built on lessons from budding yeast. Nature Rev Genet 3:918-930.

Bieszke JA, Braun EL, Bean LE, Kang S, Natvig DO, and Borkovich KA (1999). The *nop-1* gene of *Neurospora crassa* encodes a seven transmembrane helix retinal-binding protein homologous to archaeal rhodopsins. Proc Natl Acad Sci USA 96:8034-8039.

Brachat S, Dietrich FS, Voegeli S, Lerch A, Gaffney T, and Philippsen P (2003). The *Ashbya gossypii* genome: lessons learned by comparison to the *Saccharomyces cerevisiae* genome. Fungal Genet Newsl 50 (Suppl.):92.

Braun EL (2003). Innovation from reduction: Gene loss, domain loss, and sequence divergence in genome evolution. Appl Bioinform, in press.

Braun EL and Grotewold E (2001). Fungal Zuotin proteins evolved from MIDA1-like factors by lineage-specific loss of MYB domains. Mol Biol Evol 18:1401-1412.

Braun EL, Halpern AL, Nelson MA, and Natvig DO (2000). Large-scale comparison of fungal sequence information: Mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. Genome Res 10:416-430.

Braun EL, Kang S, Nelson MA, and Natvig DO (1998). Identification of the first fungal annexin: Analysis of annexin gene duplications and implications for eukaryotic evolution. J Mol Evol 47:531-543.

Bruns TD, Vilgalys R, Barns SM, Gonzalez D, Hibbett DS, Lane DJ, Simon L, Stickel S, Szaro TM, Weisburg WG, and Sogin ML (1992). Evolutionary relationships within the fungi: Analyses of nuclear small subunit rRNA sequences. Mol Phylogenet Evol 1:231-241.

Bullerwell CE, Leigh J, Forget L, and Lang BF (2003). A comparison of three fission yeast mitochondrial genomes. Nucl Acids Res 31:759-768.

Cambareri EB, Jensen BC, Schabtach E, and Selker EU (1989). Repeat-induced G-C to A-T mutations in *Neurospora*. Science 244:1571-1575.

Case ME, Schweizer M, Kushner SR, and Giles NH (1979). Efficient transformation of *Neurospora crassa* by utilizing hybrid plasmid DNA. Proc Natl Acad Sci USA 76:5259-5263.

Cogoni C (2001). Homology-dependent gene silencing mechanisms in fungi. Annu Rev Microbiol 55:381-406.

Cohen BA, Mitra RD, Hughes JD, and Church GM (2000). A computational analysis of whole genome expression data reveals chromosomal domains of gene expression. Nature Genet 26:183 186.

Davis RH and Perkins DD (2002). *Neurospora*: A model of model microbes. Nature Rev Genet 3:397-403.

Doolittle WF (1998). You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307-311.

Doolittle WF (1999). Lateral genomics. Trends Genet 15:5-8.

Froehlich AC, Liu Y, Loros JJ, and Dunlap JC (2002). White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. Science 297:815-819.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen C, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D,

Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, and Birren B (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859-868.

Geever RF, Huiet L, Baum JA, Tyler BM, Patel VB, Rutledge BJ, Case ME, and Giles NH (1989). DNA sequence, organization and regulation of the *qa* gene cluster of *Neurospora crassa*. J Mol Biol 207:15-34.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, and Oliver SG (1996). Life with 6000 genes. Science 274:546-567.

Gogarten JP (2003). Gene transfer: gene swapping craze reaches eukaryotes. Curr Biol 13:R53 R54.

Gow NAR, Brown AJP, and Odds FC (2002). Fungal morphogenesis and host invasion. Curr Opin Microbiol 5:366-371.

Graebe JE (1987). Gibberellin biosynthesis and control. Annu Rev Plant Physiol Plant Mol Biol 38:419-465.

He Q, Cheng P, Yang Y, Wang L, Gardner KH, and Liu Y (2002). White Collar-1, a DNA binding transcription factor and a light sensor. Science 297:840-843.

Heath IB and Steinberg G (1999). Mechanisms of hyphal tip growth: tube dwelling amebae revisited. Fungal Genet Biol 28:79-93.

Hedden P, Phillips AL, Rojas MC, Carrera E, and Tudzynski B (2001). Gibberellin biosynthesis in plants and fungi: A case of convergent evolution? J Plant Growth Regul 20:319-331.

Hooper SD and Berg OG (2003). On the nature of gene innovation: Duplication patterns in microbial genomes. Mol Biol Evol 20:945-954.

Houlahan MB, Beadle GW, and Calhoun HG (1949). Linkage studies with biochemical mutants of *Neurospora crassa*. Genetics 34:493-507.

Ikeda K, Nakayashiki H, Kataoka T, Tamba H, Hashimoto Y, Tosa Y, and Mayama S (2002). Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. Mol Microbiol 45:1355-1364.

Jacobson DJ, Powell AJ, Dettman JR, Saenz GS, Barton MM, Hiltz MD, Dvorachek WH, Glass NL, Taylor JW, and Natvig DO (2003). *Neurospora* in temperate forests of western North America. Mycologia, in press.

Kawanabe Y, Yamane H, Murayama T, Takahashi N, and Nakamura T (1983). Identification of gibberellin A3 in mycelia of *Neurospora crassa*. Agric. Biol. Chem. 47:1693-1694.

Keeling PJ (2003). Congruent evidence from alpha-tubulin and beta-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal Genet Biol 38:298-309.

Keeling PJ and Fast NM (2002). Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 56:93-116.

Keese PK and Gibbs A (1992). Origins of genes: "Big bang" or continuous creation? Proc Natl Acad Sci USA 89:9489-9493.

Kelkar HS, Griffith J, Case ME, Covert SF, Hall RD, Keith CH, Oliver JS, Orbach MJ, Sachs MS, Wagner JR, Weise MJ, Wunderlich JK, and Arnold J (2001). The *Neurospora crassa* genome: Cosmid libraries sorted by chromosome. Genetics 157:979-990.

Kellis M, Patterson N, Endrizzi M, Birren B, and Lander ES (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241-254.

Koonin EV, Mushegian AR, and Bork P (1996). Non-orthologous gene displacement. Trends Genet 12:334-336.

Kupfer DM, Reece CA, Clifton SW, Roe BA, and Prade RA (1997). Multicellular ascomycetous fungal genomes contain more than 8000 genes. Fungal Genet Biol 21:364-372.

Kurosawa E (1926). Experimental studies on the nature of the substance excreted by the "bakanae" fungus (in Japanese). Trans Nat His Soc Formosa 16:213-227.

Lawrence JG and Roth JR (1996). Selfish operons: Horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843-1860.

Lee K, Dunlap JC, and Loros JJ (2003). Roles for WHITE COLLAR-1 in circadian and general photoperception in *Neurospora crassa*. Genetics 163:103-114.

Lindegren CC, Beanfield V, and Barber R (1939). Increasing the fertility of *Neurospora* by selective inbreeding. Bot Gaz 100:592-599.

Liu YJ, Whelen S, and Hall BD (1999). Phylogenetic relationships among ascomycetes: Evidence from an RNA polymerase II subunit. Mol Biol Evol 16:1799-1808.

Loros JJ and Dunlap JC (2001). Genetic and molecular analysis of circadian rhythms in *Neurospora*. Annu Rev Physiol 63:757-794.

Lutzoni F, Pagel M, and Reeb V (2001). Major fungal lineages are derived from lichen symbiotic ancestors. Nature 411:937–940.

Mannhaupt G, Montrone C, Haase D, Mewes HW, Aign V, Hoheisel JD, Fartmann B, Nyakatura G, Kempken F, Maier J, and Schulte U (2003). What's in the genome of a filamentous fungus? Analysis of the *Neurospora* genome sequence. Nucl Acids Res 31:1944-1954.

McClintock B (1945). *Neurospora*. I. Preliminary observations of the chromosomes of *Neurospora crassa*. Am J Bot 32:671-678.

Miller W (2000). So many genomes, so little time. Nature Biotechnol 18:148-149.

Mitchell MB (1955). Aberrant recombination of pyridoxine mutants of *Neurospora*. Proc Natl Acad Sci USA 41:215-220.

Nelson MA (1996). Mating systems in ascomycetes: A romp in the sac. Trends Genet 12:69-74.

Nelson MA, Kang S, Braun EL, Crawford ME, Dolan PL, Leonard PM, Mitchell J, Armijo AM, Bean L, Blueyes E, Cushing T, Errett A, Fleharty M, Gorman M, Judson K, Miller R, Ortega J, Pavlova I, Perea J, Todisco S, Trujillo R, Valentine J, Wells A, Werner-Washburne M, and Natvig DO (1997). Expressed sequences from conidial, mycelial, and sexual stages of *Neurospora crassa*. Fungal Genet Biol 21:348-363.

Ohno S (1970). Evolution by Gene Duplication. New York: Springer.

Ohno S (1984). Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence. Proc Natl Acad Sci USA 81:2421-2425.

Padilla PA, Fuge EK, Crawford ME, Errett A, and Werner-Washburne M (1998). The highly conserved, coregulated *SNO* and *SNZ* gene families in *Saccharomyces cerevisiae* respond to nutrient limitation. J Bacteriol 180:5718-5726.

Perkins DD and Davis RH (2000). Evidence for safety of *Neurospora* species for academic and commercial uses. Appl Environ Microbiol 66:5107-5109.

Perkins DD, Radford A, and Sachs MS (2001). The *Neurospora* Compendium: Chromosomal Loci. San Diego: Academic Press.

Perkins DD and Turner BC (1988). *Neurospora* from natural populations: toward population biology of a haploid eukaryote. Exp Mycol 12:91-131.

Phillips MA and Croteau RB (1999). Resin-based defenses in conifers. Trends Plant Sci 4:184 190.

Prade RA, Griffith J, Kochut K, Arnold J, and Timberlake WE (1997). *In vitro* reconstruction of the *Aspergillus* (= *Emericella*) *nidulans* genome. Proc Natl Acad Sci USA 94:4564-14569.

Roelofs J and Van Haastert PJ (2001). Genes lost during evolution. Nature 411:1013-1014.

Roenneberg T and Merrow M (2001). Seasonality and photoperiodism in fungi. J Biol Rhythms 16:403-414.

Rosewich UL and Kistler HC (2000). Role of horizontal gene transfer in the evolution of fungi. Annu Rev Phytopathol 38:325-363.

Salzberg SL, White O, Peterson J, and Eisen JA (2001). Microbial genes in the human genome: Lateral transfer or gene loss? Science 292:1903-1906.

Schulte U, Becker I, Mewes HW, and Mannhaupt G (2002). Large scale analysis of sequences from *Neurospora crassa*. J Biotechnol 94:3-13.

Seack J, Perovic S, Gamulin V, Schröder HC, Beutelman P, Müller IM, and Müller WEG (2001). Identification of highly conserved genes: SNZ and SNO in the marine sponge *Suberites domuncula*: their gene structure and promoter activity in mammalian cells. Biochim Biophys Acta 1520:21-34.

Shear CL and Dodge BO (1927). Life histories and heterothallism of the red bread-mold fungi of the *Monilia sitophila* group. J Agric Res 34:1019-1042.

Shiu PKT, Raju NB, Zickler D, and Metzenberg RL (2001). Meiotic silencing by unpaired DNA. Cell 107:905-916.

Smith ML, Bruhn JN, and Anderson JB (1992). The fungus *Armillaria bulbosa* is among the largest and oldest living organisms. Nature 356:428-431.

Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, and Brown JR (2001). Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature 411:940 944.

Takken FL, Luderer R, Gabriels SH, Westerink N, Lu R, deWit PJ, and Joosten MH (2000). A functional cloning strategy, based on a binary PVX-expression vector, to isolate HR-inducing cDNAs of plant pathogens. Plant J 24:275-283.

The International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature 409:860-921.

Tomita K, Murayama T, and Nakamura T (1984). Effects of auxin and gibberellin on elongation of young hyphae in *Neurospora crassa*. Plant Cell Physiol 25:355-358.

Trapp S and Croteau R (2001). Defensive resin biosynthesis in conifers. Annu Rev Plant Physiol Plant Mol Biol 52:689-724.

Tudzynski B and Holter K (1998). Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: Evidence for a gene cluster. Fungal Genet Biol 25:157-170.

Vivarès CP, Gouy M, Thomarat F, and Méténier G (2002). Functional and evolutionary analysis of a eukaryotic parasitic genome. Curr Opin Microbiol 5:499-505.

Watters MK, Randall TA, Margolin BS, Selker EU, and Stadler DR (1999). Action of repeat induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. Genetics 153:705-714.

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, McDonald S, McLean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, O'Neil S, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor R, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Fritzc C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, Eger P, Zimmermann W, Wedler H, Wambutt R, Purnelle B, Goffeau A, Cadieu E, Dreano S, Gloux S, Lelaure V, Mottier S, Galibert F, Aves SJ, Xiang Z, Hunt C, Moore K, Hurst SM, Lucas M, Rochet M, Gaillardin C, Tallada VA, Garzon A, Thode G, Daga RR, Cruzado L, Jimenez J, Sanchez M, delRey F, Benito J, Dominguez A, Revuelta JL, Moreno S, Armstrong J, Forsburg SL, Cerrutti L, Lowe T, McCombie WR, Paulsen I, Potashkin J, Shpakovski GV, Ussery D, Barrell BG and Nurse P (2002). The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871-880.

Wood V, Rutherford KM, Ivens A, Rajandream M-A, and Barrell BG (2001). A re-annotation of the *Saccharomyces cerevisiae* genome. Comp Funct Genom 2:143-154.

Wubben JP, Joosten MH, and deWit PJ (1994). Expression and localization of two *in planta* induced extracellular proteins of the fungal tomato pathogen *Cladosporium fulvum*. Mol Plant Microbe Interact 7:516-524.

Yabuta T and Sumiki Y (1938). On the crystal of gibberellin, a substance to promote plant growth. J Agric Chem Soc Japan 14:1526.

Yagera LN, Lee H-O, Nagle DL, and Zimmerman JE (1998). Analysis of *fluG* mutations that affect light-dependent conidiation in *Aspergillus nidulans*. Genetics 149:1777-1786.

Yamaguchi S and Kamiya Y (2000). Gibberellin biosynthesis: its regulation by endogenous and environmental signals. Plant Cell Physiol 41:251-257.