Minireview

The *Neurospora crassa* genome opens up the world of filamentous fungi

Michael J Hynes

Address: Department of Genetics, University of Melbourne, Parkville, Victoria 3010, Australia. E-mail: mjhynes@unimelb.edu.au

Published: 28 May 2003

Genome ***Biology*** 2003, **4:**217

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/6/217

Abstract

The filamentous fungus *Neurospora crassa*, which has played an important role in the development of modern genetics, has several unique genome-defense mechanisms, including a process called repeat-induced point mutation. The draft genome sequence has revealed several unusual features, which suggest that the evolution of *N. crassa* has been greatly influenced by these defense mechanisms.

Neurospora history

There are not many organisms that can be described as "changing the genetic landscape for all time" [1], but *Neurospora crassa* is one such organism. In 1941 Beadle and Tatum established the field of modern microbial genetics and clearly demonstrated the relationship between genes and proteins using *N. crassa*; they formulated what became known as the 'one gene, one enzyme' hypothesis. This was possible because *Neurospora* is a haploid organism with well-understood genetic characteristics and it grows on defined media, which allowed the isolation of the first auxotrophic mutants. Unique biological features, such as unusual genome-defense mechanisms, circadian rhythms, regulation of meiotic recombination, vegetative incompatibility reactions, and responses to light, have enabled *N. crassa* to lead the way in several important areas of fundamental eukaryotic biology [2]. It is therefore appropriate that the *N. crassa* sequence is the first genome of a fungus that can grow as a mycelium with a network of hyphae to be described in detail [3,4].

The first eukaryotic genome to be analyzed was a fungal genome, that of *Saccharomyces cerevisiae*, and it has provided a basis for studies of central eukaryotic cell functions. *S. cerevisiae* has limitations as a model even for fungi, however, because it is strictly unicellular and has a rather specialized streamlined genome with most genes lacking introns and a metabolism in which glucose fermentation is strongly favored. Comparative genomic studies of yeast species, including the evolutionarily divergent *Schizosaccharomyces pombe*, are now possible [5]. Filamentous fungi have much to offer both comparatively and in terms of understanding their unique basic and applied biology [6]. For example, species such as *Neurospora*, with a history of classical genetics, are being analyzed at a molecular genetic level and serve as models for important pathogens and for fungi used in industrial processes for the production of enzymes and secondary metabolites (so-called because they are not required for growth and maintenance of the cell), such as antibiotics. The evolutionary diversity of fungi, which is reflected in different life histories, developmental processes and ecological niches, provides compelling reasons for sequencing many fungal genomes in the near future.

The diversity of *Neurospora* genes

A high-quality draft of the *N. crassa* genome has now been reported [3], following a preliminary report of the annotation of 12 megabases (Mb) of sequence [4]. The genomic data [3] provide more than 20-fold sequence coverage of the genome, which has a total length of 39.9 Mb. A total of 10,082 protein-coding genes have been predicted with, on average, one gene per 3.7 kilobases (kb) and an average of 1.7 short introns (134 bp on average) per gene. Of the predicted proteins 41% have no significant matches to known proteins, indicating the riches waiting to be discovered. In addition, of 1,421 genes with highest matches to either plant

or animal proteins, a significant number (584) have no high-scoring protein matches in either *S. cerevisiae* or *S. pombe*. Many of these proteins may be involved in determining hyphal growth and multicellular developmental structures in *Neurospora*, as these characteristics are not found in yeasts. It has also been noted that for many *Neurospora* genes the only known homologs are found in prokaryotes [4], indicating that occupation of similar ecological niches has resulted in conservation of genes for substrate degradation and secondary metabolism. This may be a major reason for the relatively large number of *N. crassa* genes with predicted products likely to be involved in catabolism, chemical detoxification and stress-defense mechanisms [3]. Functional characterization of all of these genes will be challenging.

Signaling pathways

The sequences encoding proteins that act on well-studied signaling pathways, including mitogen-activated protein (MAP) kinases and cyclic AMP-dependent protein kinase, as well as small GTPases of the Ras family, are readily recognized in the *N. crassa* genome [3]. In addition, ten distinct transmembrane G-protein-coupled receptors are found, although there are only three different G_{α} proteins. Similarities with the slime mould *Dictyostelium discoideum* suggest that there are uncharacterized extracellular cyclic AMP-signaling pathways in *N. crassa*. Two-component histidine kinases and calcium-signaling components are more abundant in *N. crassa* than in *S. cerevisiae*. *N. crassa* has proven one of the best organisms for studying the blue-light sensing that affects circadian rhythms as well as asexual and sexual development, and new players in these pathways have been revealed. A surprise is that two putative red-light sensors have also been found [3], although these have no known biological function in *N. crassa*.

The signaling pathways that have been identified determine the ability of *N. crassa* to respond to environmental cues, such as light, that influence growth and development, and correlate with many previous observations of *N. crassa* in the wild (see [2]). The best studies on the balance between growth, reproduction and secondary metabolism of filamentous fungi have been carried out in *Aspergillus nidulans* [7,8]. A comparison between the genome of *N. crassa* and the draft sequence of *A. nidulans* [9] shows that *N. crassa* lacks important genes involved in the control of asexual spore production that are present in *A. nidulans*, but it does have an apparent homolog of the *velvet* (*veA*) gene, which is central to these processes of growth and development, so we need some comparative developmental biology.

Growth and pathogenicity

Genomic analysis will make it possible to address two fascinating questions about fungal biology. First, is there a single common mechanism that determines the 'true' hyphal growth form of filamentous fungi? *Candida albicans*, like *S. cerevisiae*, has budding and pseudohyphal growth (with elongated budded cells that are constricted), but *C. albicans* also has true hyphal growth. Genes common to *C. albicans* and *N. crassa* but not *S. cerevisiae* will therefore be of interest. Analysis of genome sequences of dimorphic fungi that switch between distinct unicellular (yeast) and hyphal phases will be another approach.

The second question - what makes a fungus a pathogen - is an old one. Although *N. crassa* grows only on decaying plant material, the genome sequence [3] has revealed many proteins and systems with functions that have previously been found to be important in plant pathogens, for example, proteins such as cytochrome P450 and efflux systems involved in detoxification of plant antifungal compounds. Clearly, for many plant pathogens the ability to elaborate infection structures and evade plant resistance mechanisms is important. For those lacking specialized infection structures the distinction may be less clear. For many human opportunistic pathogens, for example, an ability to grow in the stressful foreign host-cell environment may be all that is required.

Genome defense

Studies of *N. crassa* have made great contributions to the field of epigenetic phenomena, all of which relate to so-called 'genome defense' directed against invasion by foreign nucleic acids, such as transposable elements and viruses. Two post-transcriptional gene-silencing mechanisms have been found in *Neurospora*: quelling and meiotic silencing by unpaired DNA. During quelling, which is related to RNA interference (RNAi) in plants, nematodes and mammals, duplicated sequences are recognized and silenced during vegetative growth. Meiotic silencing by unpaired DNA, which is so far unique to *N. crassa*, involves the recognition of unpaired DNA during meiosis, resulting in the silencing of all genes encoded by that DNA. This results in an RNA-mediated silencing process that affects the expression of all genes included in the unpaired region as well as the corresponding paired homologous sequences. Because a linear and ordered array of eight ascospores is produced during meiosis (Figure 1), silencing of genes that are required during meiosis can be easily detected by looking for the formation of nonviable or abnormal ascospores. The genome sequence has now revealed new genes involved in post-translational gene silencing through quelling and meiotic silencing by unpaired DNA in addition to those previously found [3]. The two pathways are distinct, and each involves an RNA-dependent RNA polymerase, a homolog of the Argonaute protein, a Dicer-like protein and a RecQ-like helicase (see Table 4 in [3]). Both pathways are apparent in *Aspergillus fumigatus*, which lacks a known sexual cycle, whereas homologs of only the meiotic pathway are present in *S. pombe*.

Repeat-induced point mutation and genome evolution

The study of the extraordinary phenomenon of repeat-induced point mutation in *N. crassa* has made seminal con-

Figure 1
A rosette of *Neurospora* asci. Four of the eight ascospores in each ascus are highlighted with green fluorescent protein (visible as bright spots). Image courtesy of Namboori Raju (Stanford University, USA).

tributions to the understanding of the methylation of both DNA and, more recently, histone proteins [10]. In the process of repeat-induced point mutation, duplicated sequences that are longer than 400 base-pairs (bp) and have more than 80% identity are recognized during the pre-meiotic dikaryotic stage and are mutated. A high number of C:G to T:A mutations (affecting up to 30% of C:G pairs) can occur within a single sexual cycle; there is a preference for mutation at CpA dinucleotides, resulting in a recognizable skewing of dinucleotide frequencies that allows the detection of sequences that have undergone repeat-induced point mutation. This process results in a high probability of non-sense codons and commonly methylation of the mutated sequences, leading to gene inactivation.

The process of repeat-induced point mutation has had a major impact on the *N. crassa* genome, and analysis of the whole genome has now provided an overview of the consequences of repeat-induced point mutation for genome evolution [3]. There are significantly fewer multigene families whose members have similar sequences than would be expected from the genome size, and only eight gene pairs with high identity (more than 80%) were found. This is in contrast to *S. pombe*, which has many such gene pairs, and to the fungal plant pathogen *Magnaporthe grisea*, which does not have an abnormally low number of gene pairs (James Galagan, personal communication). It might therefore be expected that *N. crassa* has been restricted in the evolution of families of genes with similar functions, but similar numbers of predicted sugar transporters, for example, are found in *N. crassa* and *S. cerevisiae* (with 32 and 30 sugar transporters, respectively). These are, however, far more divergent in *N. crassa* than in *S. cerevisiae*, in which the family of sugar transporters has arisen via relatively recent duplications. In the *N. crassa* genome sequence [3], predicted occurrences of repeat-induced point mutation were found in a total of only 59 genes, and only eight of these are predicted to be duplicated, indicating that few genes survive the mutation process. The evolution of new functions by gene duplication must therefore have occurred prior to the evolution of repeat-induced point mutation. Analysis of the genomes of a number of organisms has indicated that ancient polyploidy followed by reversion to a diploid state has been a common mechanism that has provided duplicated genes for evolutionary adaptations [11]. In *S. cerevisiae*, gene pairs are common in which each gene of the pair has a related but distinct function; the finding of only few highly related pairs of genes in *N. crassa* is in stark contrast to this.

The effectiveness of repeat-induced point mutation is apparent when repeat sequences are analyzed, as the majority have undergone the process. Consistent with the fact that the process operates on sequences that are longer than 400 bp, more than 90% of repeats longer than 400 bp have an average identity of less than 85% [3]. A large proportion (46%) of the sequences that have been subjected to repeat-induced point mutation are identifiable as relics of mobile elements, providing graphic evidence for the effectiveness of the mutation mechanism in defending the genome against invading sequences. We can only speculate about the evolutionary cost of this process - but *N. crassa* is clearly successful, with a genome encoding a wide variety of functions and a gene number 20-30% that of man. Mutation in a single gene (*RID-1*) encoding a cytosine methyl transferase [12] can eliminate repeat-induced point mutation, and the 'experiment' of losing the process must therefore occur naturally, but nevertheless the function has been maintained through evolution.

Comparisons of the genome sequences of many fungi are now required if we are to explore further the issues of pathogenicity and genome defense that have been raised by knowledge of the *N. crassa* sequence. At the recent Fungal Genetics Conference [13] experiments on more than 120 species of fungi were reported. High-quality draft sequences of *M. griseae* [14], *A. nidulans* [9] and *A. fumigatus* [15] are already in the public domain, and others, including the important basidiomycete pathogens *Cryptococcus neoformans* and *Ustilago maydis* are likely to be available soon. Careful choice of the appropriate fungal species for investigating biological problems will obviously bear rich rewards in the near future.

References

1. Horowitz NH: **Fifty years ago: the *Neurospora* revolution.** *Genetics* 1991,**127:**631-635.
2. Perkins DD: ***Neurospora*: the organism behind the molecular revolution.** *Genetics* 1992,**130:**687-701.

3. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, *et al.*: **The genome sequence of the filamentous fungus *Neurospora crassa*.** *Nature* 2003, **422:**859-868.
4. Mannhaupt G, Montrone C, Haase D, Mewes HW, Aign V, Hoheisel JD, Fartmann B, Nyakatura G, Kempken F, Maier J, Schulte U: **What's in the genome of a filamentous fungus? Analysis of the *Neurospora* genome sequence.** *Nucleic Acids Res* 2003, **31:**1944-1954.
5. ***Hemiascomycetes* - RST from 13 hemiascomycetes yeasts** [http://www.genoscope.cns.fr/externe/English/Projets/Projet_AR/AR.html]
6. **Fungal Genome Initiative** [http://www-genome.wi.mit.edu/seq/fgi/index.html]
7. Adams T, Wieser JK, Yu J-H: **Asexual sporulation in *Aspergillus nidulans*.** *Microbiol Mol Biol Rev* 1998, **62:**35-54.
8. Calvo AM, Wilson RA, Bok JW, Keller NP: **Relationship between secondary metabolism and fungal development.** *Microbiol Mol Biol Rev* 2002, **66:**447-459.
9. ***Aspergillus nidulans* database** [http://www-genome.wi.mit.edu/annotation/fungi/aspergillus/index.html]
10. Tamaru H, Selker EU: **A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*.** *Nature* 2001, **414:**277-283.
11. Kellogg EA: **What happens to genes in duplicated genomes.** *Proc Natl Acad Sci USA* 2003, **100:**4369-4371.
12. Freitag M, Williams RL, Kothe GO, Selker EU: **A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*.** *Proc Natl Acad Sci USA* 2002, **99:**8802-8807.
13. **22nd Fungal Genetics Conference** [http://www.fgsc.net/asil2003/asil2003.htm]
14. ***Magnaporthe grisea* database** [http://www-genome.wi.mit.edu/annotation/fungi/magnaporthe/index.html]
15. **The *Aspergillus fumigatus* genome database** [http://www.tigr.org/tdb/e2k1/afu1/]