Clipboard

Blueprint of a red mould: Unusual and unexpected findings in the Neurospora genome sequence

Many genes whose presence in *Neurospora crassa* was unexpected have been revealed in the just published genome sequence (Galagan *et al* 2003). Among them were two genes that encoded putative homologues of phytochrome proteins known to be required for red light sensing in other organisms. The surprising thing is that no red light photobiology is known in Neurospora. Three genes appear to encode proteins similar to dictyostelid cAMP receptors. Genes were also found for all the proteins required for the synthesis and degradation of cAMP. The implication is that cAMP or a related molecule may serve as an extracellular signal in Neurospora: an extracellular cAMP signalling pathway has not been demonstrated in fungi. Three genes resemble genes for non-ribosomal peptide synthetases, seven for polyketide synthases, and several for diterpene biosynthesis, indicating a surprisingly broad repertoire of secondary metabolite production in Neurospora (the only secondary metabolites known to be made by Neurospora to date are carotenoids and melanin). Even though Neurospora is not known to be a plant pathogen, it was found to have a number of homologs of genes of plant pathogenic fungi that would appear to have no other function except in pathogenesis, thereby suggesting that saprophytic and parasitic fungi might not differ much in lifestyle. Neurospora carries eleven histidine kinase genes compared with only one in the yeast *Saccharomyces cerevisiae* and three in *Schizosaccharomyces pombe*, pointing to a greater role for two-component signalling in filamentous fungi. On the other hand, no Neurospora homologs were found of several genes known to be required for macroconidiation in *Aspergillus nidulans*; so that the molecular machinery underlying macroconidiation may differ significantly in these two filamentous fungi. In short, the sequence enabled Neurospora researchers to quickly update their ignorance about the system they had been studying for over 70 years.

The assembled sequence was ~ 40 Mb (38,639,769 bp) in size and predicted 10,082 protein-coding genes, of which 9200 encoded proteins of > 100 amino acid residues. This is almost twice as many genes as in the yeasts *S. pombe* (~ 4800) and *S. cerevisiae* (~ 6300), which have a much smaller genome (~ 12 Mb), and comparable with the fly *Drosophila* (~ 14,300). About 57% of Neurospora proteins lacked significant matches to any protein in the two yeasts, and of the sizeable fraction (14%) of genes showing best matches to plant or animal proteins, 584 did not have a yeast homologue. These figures underscore the importance of Neurospora as a model organism. About 41% of the genes lacked significant matches to known proteins from public databases. These would include genes that are important for filamentous fungi. All the 252 previously sequenced genetic markers on the seven linkage groups were identified which allowed the assembly to be anchored to the genetic map. Other genes found included the 424 tRNA genes and 74 5S rRNA genes. Only about 2–3% (~ 1·7 Mb) of the genome sequence (including sequences such as the ribosomal DNA repeats which cannot be assembled with available techniques) was not covered by the draft sequence.

Repeat sequences longer than 200 bp and sharing greater than 65% similarity represented only 10% of the assembly. Most of them showed evidence of having suffered the sexual phase-specific genome defense mechanism called RIP (repeat-induced point mutation). *N. crassa* is one of only four species in which RIP has been demonstrated. The other three are the fungi *Podospora anserina*, *Magnaporthe grisea* and *Leptosphaeria maculans*. RIP causes multiple G : C to A : T mutations and cytosine methylation in duplicated DNA sequences that are longer than ~ 400 bp and share greater than ~ 80–85% sequence identity (review: Selker 1990). RIP is a highly efficient process; a single passage through the sexual cycle can mutate up to 30% of the C : G pairs in the duplicated sequences. It has a strong preference for C to T mutations at CpA dinucleotides. Therefore RIP-mutated sequences show

a skewing of dinucleotide frequencies which allows them to be accurately detected. Only after the elements of the duplication have diverged sufficiently (< 85% similarity) do they become insensitive to further rounds of RIP. It is estimated that by this point there is a 99·5% probability that their open reading frames would have acquired an in-frame stop codon (via mutation of CAA and CAG to TAA and TAG). Most of the repeat clusters in the genomic sequence had an average identity of < 85% which was consistent with the idea that they had diverged to below the threshold required for detection by RIP. No intact mobile elements were found, and 46% of the RIP-mutated repetitive sequences were identified as relics of mobile elements. Most of the regions that were predicted to be methylated corresponded to regions that were predicted to be repeated and RIP-mutated. Only the 175–200 copies of large rDNA tandem repeats containing the 17S, 5·8S, and 25S rRNA genes are known to be resistant to RIP because of their localization within the nucleolar organizing region (NOR); rDNA repeats that had strayed to other locations in the genome displayed evidence of mutation by RIP. The 5S rRNA genes also have survived RIP but in this case it was because of their small size (~ 120 nt) and dispersal in the genome.

In all species similar genes can be clustered into multigene families and, in general, the percentage of genes in multigene families is correlated with genome size. The Neurospora genome however possessed many fewer genes in multigene families than expected for its genome size. Only eight genes shared matches of > 80% amino acid and nucleotide identity. The small proportion of genes in multigene families and the small number of highly similar genes are consistent with the action of RIP. The effects of RIP were evident even at the level of individual exons and protein domains. Only 0·1% exons shared > 80% nucleotide identity and only 0·9% of predicted protein domains showed > 80% amino acid identity compared with 10–16% of predicted protein domains in the yeasts.

Given that RIP suppresses the creation of new genes through genomic duplication, how might Neurospora have acquired new genes? It is possible that no paralogous genes evolved subsequent to the emergence of RIP and that new genes were acquired via lateral transfer. Another possibility is that new genes evolved in strains that rarely entered the sexual cycle. A third possiblity is that the evolution might have taken advantage of the fact that RIP is not 100% efficient and thus paralogous sequences escaped RIP during a few rounds of the sexual cycle. In this context, a significant role might have been played by strains that dominantly suppress RIP. Studies carried out in our laboratory show that about 1–2% of wild-isolated *N. crassa* strains are dominant suppressors of RIP (Bhat *et al* 2003).

References

Bhat A, Noubissi F K, Vyas M and Kasbekar D P 2003 Genetic analysis of wild-isolated *Neurospora crassa* strains identified as dominant suppressors of repeat-induced point mutation (RIP); *Genetics* (in press)

Galagan J E, Calvo S E, Borkovich K A, Selker E U *et al* 2003 The genome sequence of the filamentous fungus *Neurospora crassa*; *Nature* (*London*) **422** 859–868

Selker E U 1990 Premeiotic instability of repeated sequences in *Neurospora crassa*; *Annu. Rev. Genet.* **24** 579–613

DURGADAS P KASBEKAR
Centre for Cellular and Molecular Biology,
Hyderabad 500 007, India
(Email, kas@ccmb.res.in)