Current Biology, Vol. 13, R552–R553, July 15, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0960-9822(03)00470-6

Genome Sequencing: The Ripping Yarn of The Frozen Genome

Dispatch

J.F.Y. Brookfield

The completion of the genome sequence of the filamentous fungus *Neurospora crassa* reveals a gene number very much higher than those of yeasts. Of particular interest in this species are the effects of the repeat-induced point mutation (RIP) process, which appears to have prevented recent evolution through gene duplication in this lineage.

The recent publication of the *Neurospora crassa* genome sequence [1] has allowed the first insights into the complete genome of a fungus that is not one of the yeasts. This species, famous in the history of genetics as the model system in which the 'one gene–one enzyme' relationship was demonstrated [2], turns out to have an estimated 10,082 genes. This is considerably more than the approximately 6,300 seen in the genome of the budding yeast *Saccharomyces cerevisiae* or the approximately 4,800 in the genome of the fission yeast *Schizosaccharomyces pombe*. The *N. crassa* genome size, at around 40 megabase pairs, is also more than three times as large as those of these very distantly related yeast species. What are these extra genes doing? The answer to this is not clear — more than 4,000 of the apparent *N. crassa* genes lack any significant matches to the databases. Many may be involved in cell signaling processes which are involved in the true hyphal growth seen in *N. crassa* but absent in yeasts.

Perhaps the most interesting aspect of the *N. crassa* genome, however, is the signs of the process called repeat-induced point mutation (RIP). This was first identified through genetic experiments [3], in which it was found that, as *N. crassa* cells pass through their sexual cycle, any repeated DNA sequences of length greater than 400 base pairs that share greater than 80% sequence identity, become peppered with a large number of C:G to T:A mutations (up to 30 per cent of the C:G base pairs are changed). The events occur after the fertilization that creates a dikaryon, but before the premeiotic DNA synthesis that follows this event [3]. The result is that the probability of a newly duplicated open reading frame being preserved, even from a single round of 'RIPing', is only 20% for a gene encoding a polypeptide of typical length.

The *N. crassa* genome reveals the effects of the RIP process. Scanning along the genome sequence allows the detection of sequences that show the signs of the RIP phenomenon in the skewing of their base frequencies; many of these sequences indeed turn out to be repeats. These sequences are also hotspots for methylation.

Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK.

The other expectation from a RIP process that has existed over a considerable evolutionary period is that there will be a paucity of pairs of recently duplicated genes. This is also seen [1]. Of all the pairs of paralogous genes in the genome, almost none shows amino acid sequence identity above 80% or DNA sequence identity above 60%. This latter observation, in particular, when compared to the greater than 80% sequence identity required to trigger RIP, implies that the process has operated in *N. crassa* and its ancestors' genomes for a considerable evolutionary time. The RIP phenomenon will result in the selective elimination of gene duplicates soon after they are formed. If a local duplication arises, creating two linked copies of a gene, both will be RIPped, and rendered non-functional — selection will prevent this allele from spreading in the population. The absence of redundancy that this implies may indeed be of use to those trying to establish the functions of genes by knockout mutations [4].

The result of the RIP process will be that the genome will become effectively 'frozen' with respect to the acquisition of new genes (genes no longer required could, of course, still be lost). This potentially allows empirical investigation of one of the most interesting evolutionary questions raised by the availability of genome sequences. Prior to the *N. crassa* genome, all genome sequences studied had revealed the signs of an ongoing process of gene duplication. One view [5] is that major innovations in evolution are only brought about through gene duplications creating a stage of temporary redundancy, where organisms have a 'spare' gene with which to evolve new functions missing in the ancestor. The implication of this model is that, in the evolution of the phenotype, some genes are maintained, and they determine phenotypic features shared with ancestors, whilst new genes, created by duplication or, in the case of many bacteria, by horizontal transfer, are the architects of new phenotypes.

In its most extreme form, this view, applied particularly to the genomes of microorganisms, is that it is in differences between the lists of genes present in different genomes that the explanation for the phenotypic differences between organisms should be sought. This view is understandable from a pragmatic standpoint, in that the identification of new genes is easy from genome sequence comparisons, whereas the identification of adaptive changes among the very many sequence changes in orthologs is much more difficult.

The alternative view, however, is that it is the changes in amino-acid sequence of the protein product, or in regulatory sequences that control expression, in orthologs that underlie many, perhaps most, phenotypic changes. But as it appeared that all organisms evolve through a combination of potentially advantageous changes in orthologous sites and the creation of new genes following duplications, determining the relative contributions of these two sources of adaptive phenotypic change appeared to be impossible.

Now, however, it seems that the *N. crassa* genome offers the opportunity to investigate directly phenotypic evolution in a clade where gene duplication is effectively impossible. The finding that, in this species, the amino-acid sequence difference between gene pairs is almost never less than 20% allows the potential estimation, from molecular clocks, of the time since the RIP system evolved, and thus the identification of other fungal species sharing ancestry with *N. crassa* more recently than the time of origin of the RIP process. Indeed, recent results have identified the existence of RIP in the close *N. crassa* relative *Podospora anserina* [6,7] and in the more distantly related *Magnaporthe grisea* [8]. This indicates the existence of a large clade of organisms that have lost the capacity for evolution by gene duplication from their evolutionary repertoire. These could — notwithstanding the difficulties of producing an objective scale of phenotypic diversification that can be applied to fungi — be seen as a natural experiment, in which a genome frozen with respect to duplications is tested for its ability to create phenotypic evolution. This will give insight into how much phenotypic change can be generated by base changes in orthologs alone.

Why does the RIP process happen? It may have evolved to protect the genome from harmful transposable elements. Support for this view has come from studies of RIP and of methylation [9] — almost all methylated, RIPped sequences turn out to be relics of inactivated transposons. It is not obvious, however, how it is possible for the rate of new introduction of transposons to be sufficient to create strong enough selection to maintain the RIP machinery. Evolution does not, after all, allow foresight.

Given the apparent utility of the RIP process, however, one can ask the question why other organisms do not use it. Is this because of a loss of evolutionary potential resulting from the elimination of gene duplication? Clearly, any RIP process needs to protect the genome's functional repeated DNAs. In *N. crassa*, the ribosomal (r)DNA clusters, which are tandemly repeated, are protected by their position in the nucleolus. The tRNA and 5S rRNA genes are too small to be affected by RIP. There may, however, be a problem in the evolution of a RIP process — new mutant alleles causing the RIP phenomenon are only likely to be successful if they arise in a genome that does not possess pairs of recently duplicated but now functionally diverged genes. As most genomes possess such pairs, it may be that few lineages ever go through a window in evolutionary time when the initial creation of a RIP process would be permissible.

References

1 Galagan, J.A., Calvo, S.E., Borkovich, K. A., Selker, E.U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S., Purcell, S. *et al.* (2003). The genome sequence of the filamentous fungus Neurospora crassa. Nature *422*, 859-868.

2. Beadle, G.W. and Tatum, E.I., (1941). Genetic control of biochemical reactions in Neurospora. Proc. Natl. Acad. Sci. U.S.A. *27*, 499-506.
3. Selker, E. U. (1990). Premeiotic instability of repeated sequences in Neurospora crassa. Annu. Rev. Genetics *24*, 579.
4. Arnold, J. and Hilton, N. (2003). Revelations from a bread mould. Nature *422*, 821-822.
5. Ohno, S. (1970). Evolution by Gene and Genome Duplication (Berlin: Springer).
6. Hamann, A., Feller, F., and Osiewacz, H.D. (2000). The degenerate DNA transposon Pat and repeat-induced point mutation (RIP) in Podospora anserina. Mol. Gen. Genet. *263*, 1061-1069.
7. Graïa, F., Lespinet, O., Rimbault, B., Dequard-Chablat, M., Coppin, E., and Picard, M. (2001). Genome quality control: RIP (repeat-induced point mutation) comes to Podospora. Mol. Microbiol. *40*, 586-595.
8. Ikeda, K.-I., 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. Micro. *45*, 1355-1364.
9. Selker, E.U., Tountas, N. A., Cross, S. H., Margolin, B.S., Murphy, J. G., Bird, A. P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. Nature *422*, 893-897.