



A 21st century miniguide to fungal biotechnology

Una miniguía del siglo XXI para la biotecnología de hongos

Carmen Sánchez^{1*}, David Moore², Geoff Robson², Tony Trinci²

¹Research Centre for Biological Sciences, Universidad Autónoma de Tlaxcala, Tlaxcala, Mexico. ²School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom.

*Corresponding author

E-mail address: carmen.sanchezh@uatx.mx (C. Sánchez)

Article history:

Received: 13 November 2019 / Received in revised form: 15 December 2019 / Accepted: 16 December 2019 / Published online: 1 January 2020.

<https://doi.org/10.29267/mxjb.2020.5.1.11>

ABSTRACT

Realising the biotechnological potential of fungi requires full appreciation of the molecular biology and genetics of this kingdom. We review recent advances in our understanding of fungal genetic structure as it might influence biotechnology; including introns, alternative splicing of primary transcripts, transposons (transposable elements, or TEs), heterokaryosis, ploidy and genomic variation, sequencing, annotation and comparison of fungal genomes, and gene editing. We end by indicating under-researched, but unique, aspects of fungal cell biology that offer opportunities for developing new strategies to manage the activities of fungi to our benefit. As a closing example, we discuss the potential of bioengineering fungi specifically for bioremediation of plastic wastes.

Keywords: bioremediation, gene editing, genomics, molecular biology, molecular genetics, new strategies.

RESUMEN

Entender el potencial biotecnológico de los hongos requiere una apreciación completa de la biología molecular y la genética de este reino. En este documento revisamos los avances recientes de la estructura genética de hongos, ya que ésta podría influir en la biotecnología; incluidos intrones, empalme alternativo de transcripciones primarias, transposones (elementos

genética transponible o TEs), heterocariosis, ploidía y variación genómica, secuenciación, anotación y comparación de genomas fúngicos, y edición de genes. Terminamos indicando aspectos poco investigados, pero únicos, de la biología celular fúngica que ofrecen oportunidades para desarrollar nuevas estrategias en el manejo de las funciones de los hongos en nuestro beneficio. Como ejemplo final, discutimos el potencial de la bioingeniería de los hongos específicamente para la biorremediación de desechos plásticos.

Palabras clave: biología molecular, biorremediación, edición de genes, genética molecular, genómica, nuevas estrategias.

1. INTRODUCTION

We have a long heritage of using fungi and fungal products. Some of our current fungal biotechnology, such as baking, brewing and numerous fermented food products originated hundreds or even thousands of years ago, and largely by the chance association between natural fungi and one or more of the constituents of the food or beverage material.

Although the original discovery of penicillin was also a matter of chance, its industrial production in the middle of the 20th century was a much more directed process, as was the development of other products such as citric acid, fungal-modified steroids and, more recently, statin pharmaceuticals and strobilurin fungicides. Yet most improvements were at the organismal level. Techniques were found that enabled cultivation of specific organisms that had advantageous biological characteristics and improved strains were selected during development of the production process. It is certainly the case that the domestication of the many fungi now used for food or beverage production has been accomplished by unconscious (natural) selection. Also true is that in the second half of the 20th century the rapidly accumulating knowledge of fungal genetics was brought to bear, and many advances were made. But what humans have achieved by manipulating the classical segregational genetics of fungi over the past few thousand years is only a foretaste of what might be achieved in the next few decades if we can thoroughly exploit our growing understanding of the molecular genetics of fungi.

We have claimed elsewhere (Moore *et al.*, 2020) that fungi comprise the most crucial kingdom of eukaryotic organisms on this planet. To show that we are not alone in holding such a high opinion of fungi, we note the Editorial published on 25th July 2017 in the journal *Nature Microbiology* entitled 'Stop neglecting fungi', from which we quote:

'...over 300 million people suffer from serious fungal-related diseases, ... fungi collectively kill over 1.6 million people annually, which is more than malaria and similar to the tuberculosis death toll. Fungi and oomycetes destroy a third of all food crops each year, which would be sufficient to feed 600 million people. Furthermore, fungal infestation of amphibians has led to the largest disease-caused loss of biodiversity ever recorded, while fungi also cause mass mortality of bats, bees and other animals, and decimate fruit orchards, pine, elm and chestnut forests...'

[«... más de 300 millones de personas sufren de enfermedades graves relacionadas con los hongos, ... hongos matan colectivamente a más de 1,6 millones de personas al año, lo que es

más que el paludismo y similar a la cifra de muerte por tuberculosis. Los hongos y oomicetos destruyen un tercio de todos los cultivos alimentarios cada año, lo que sería suficiente para alimentar a 600 millones de personas. Además, la infestación por hongos de anfibios ha provocado la mayor pérdida de biodiversidad causada por enfermedades, mientras que los hongos también causan la mortalidad masiva de murciélagos, abejas y otros animales, y diezman los huertos frutales, los bosques de pinos, el almos y castaños...»] (Anonymous Editorial, 2017)

Similarly, the report published by the Royal Botanic Gardens, Kew, entitled State of the World's Fungi 2018, also coined a memorable phrase: '...when looking for nature-based solutions to some of our most critical global challenges, fungi could provide many of the answers'. [«... al buscar soluciones basadas en la naturaleza para algunos de nuestros desafíos globales más críticos, los hongos podrían proporcionar muchas de las respuestas. »] (Willis, 2018)

In this paper we use material extracted, with permission, from the **second edition** of the book *21st Century Guidebook to Fungi* [www.cambridge.org/9781108745680] (Moore *et al.*, 2020) to review the most recent developments in understanding the molecular genetics and molecular biology of fungi as they might be applied to biotechnological manipulation of fungi. We end by indicating some aspects of cell biology that are unique to fungi but are under-researched despite offering new ways to manipulate fungi to our benefit. Finally, we discuss the potential for bioengineering fungi specifically to remediate plastic wastes.

2. BASIC GENETIC STRUCTURE OF FUNGI

The basic genetic architecture of fungi is typical of eukaryotes in general. All the major principles of eukaryote genetics apply in fungi, including gene structure and organisation, Mendelian segregations, recombination, and the rest of the meiosis-dependent features. These are aspects of what might be called *Mendelian or segregational genetics*, which apply because of the chromosomal architecture and mechanisms of meiosis (Moore & Novak Frazer, 2002; Dyer *et al.*, 2017).

Eukaryotes and prokaryotes have quite different types of genome, but it is generally assumed that something like the prokaryotic grade of organisation is the primitive form from which the eukaryote organisation evolved. Modern prokaryotes and eukaryotes have a great deal in common, including that the DNA of a gene is transcribed into RNA, which is called a messenger RNA (mRNA) if it is a transcript of a protein-coding gene, and the mRNA is translated into protein by the ribosomes and other translation machinery. The part of a protein-coding gene sequence that is translated into protein is called the *open reading frame*, usually abbreviated to *ORF*.

As a genome sequence is assembled the functional genes in the sequence are recognised as open reading frames (ORFs); the process is called genome *annotation* and is discussed in more detail below. Not all the ORFs that are identified can be associated with a gene of identified function; an ORF specifying a product that does not resemble a known protein is called an *unidentified reading frame*, or *URF*. But comparative genomics does more than

identify the genes. It can show the evolutionary relationships between different organisms, and aids understanding of how the genotype relates to life-style and environment.

Characteristically, the ORF is read in the 5' to 3' direction along the mRNA, and it starts with an initiation codon and ends with a termination codon (Fig. 1). Nucleotide sequences that occur in the mRNA before the ORF make up the *leader* sequence, and sequences following the ORF make up the *trailer* segment.

Many eukaryotic genes are split into *exons* (meaningful segments) and *introns* (sequence segments that do not contribute to the protein-coding sequence). The introns are removed from the primary RNA transcript by the splicing machinery to form the functional mRNA (Fig. 1), and several different mRNAs may be produced from any one primary RNA transcript by a (very common) process known as *alternative splicing*.

And then there are the transposable elements (TEs), often called *transposons*, which are seemingly ubiquitous genetic elements that have been discovered in all prokaryotes and eukaryotes so far investigated. In remarkable contrast to all other genes, transposable elements are able to move to new locations within their host genomes in a process called transposition. Transposable elements contribute enormously to eukaryotic genome diversity (and evolution). Their ubiquitous presence affects the genomes of all species; mediating genome evolution by causing mutations, repetitions and chromosomal rearrangements and by modifying gene expression.

Indeed, introns and transposons seem to be extremely ancient genetic structures, which certainly existed long before the eukaryote grade of organisation emerged. So, over evolutionary time they have created new patterns of gene expression by 'shuffling' functional motifs together and then combining them with new control elements to produce differentiated cellular structures with new morphologies and/or new developmental possibilities. And not just within the one genome. Transposition is also responsible for *horizontal gene transfer* events, which is the transmission of genetic material between organisms of the same generation and usually across major taxonomic boundaries. There is a growing amount of evidence that many horizontal gene transfer events have occurred during the evolution of fungi (Richards *et al.*, 2011; Slot *et al.*, 2017; Steenkamp *et al.*, 2018). All of this is discussed further below.

The smallest eukaryotic genomes, of some yeasts, are in the region of 10 Mb (Mb is the usual abbreviation for a million base pairs or a '*megabase*'), and the largest (in vertebrates and plants) are over 100,000 Mb, so we can observe some surprising structural differences when we compare different eukaryotes. Generally speaking (but remember there are exceptions to all generalisations), it appears that space is saved in the genomes of less complex organisms by having the genes more closely packed together and by having much less repetition (Fig. 1).

The genome of *Saccharomyces cerevisiae* contains more genes per unit length of DNA than occur in human or maize DNA. On the other hand, up to 40% of the gene sequences of *S. cerevisiae* are *duplicated*. In most cases the duplicated sequences are so similar that their protein products are identical and, presumably, either functionally redundant, or (more likely) under very different *regulatory control*. The whole of chromosome XIV of *S. cerevisiae* is made up of regions duplicated on other chromosomes.

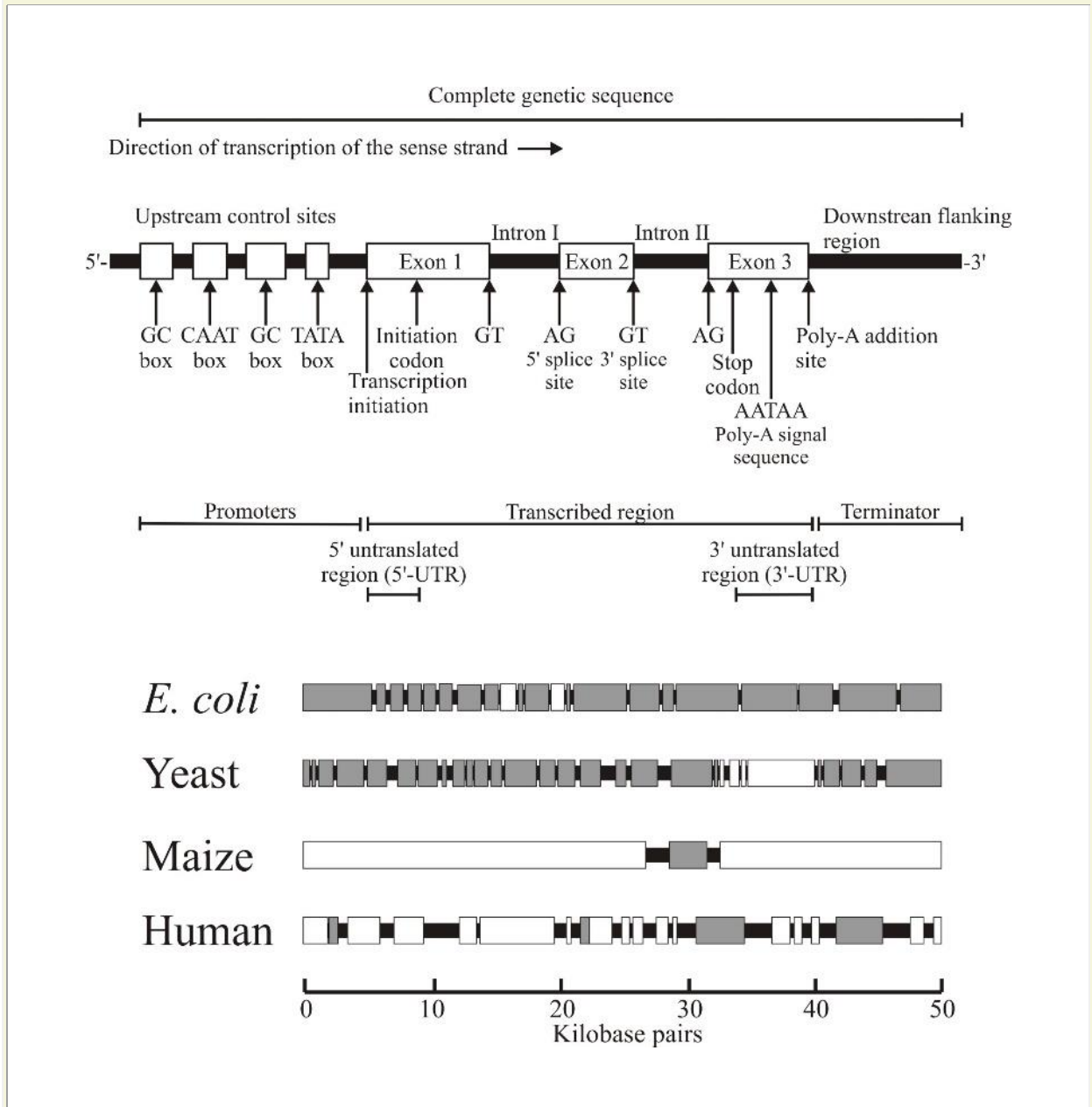


Fig. 1. Top: the basic structure of a typical eukaryotic gene. The schematic diagram indicates the structure of a type II gene; that is, a protein-encoding gene transcribed by polymerase II. The diagram is not drawn to scale and the relative sizes of the different sections differ between genes and between the eukaryotic Kingdoms. Bottom: comparison of 50 kbp segments of the genomes of the prokaryote *Escherichia coli* and three eukaryotes to show how the 'density' of genetic information varies. In each case the grey boxes correspond to gene sequences, and the white boxes correspond to stretches of repeated sequences. Adapted from Moore & Novak Frazer, 2002.

So, despite the differences that undoubtedly exist, fungi are typical eukaryotes, featuring all the basic cell biology expected of this grade of organisation. Even though the yeast genome is only in the same size range as some of the more advanced prokaryotes, the genetic structure and functioning of genes of filamentous fungi are representative of all eukaryotes and we can use their sequences to learn about genomics (Moore & Novak Frazer, 2002).

Analysis of the genetic sequences that make up the genome of an organism, and comparisons of the genomes of different organisms (exercises that have come to be known as the science of genomics) only became possible from the mid-1990s. Establishing the exact DNA sequence of a genome is a major undertaking, but is only the prelude to intensive analysis. The priority of genomics is to establish the number and function of genes in an organism.

We will describe methods developed for sequencing and studying whole fungal genomes, but first we want to discuss a few more fungus-specific genetic details; namely introns, alternative splicing, transposons, genomic variation, including gene clusters and horizontal transfer, and ploidy variation.

2.1. Introns

An intron is any nucleotide sequence within a gene that is removed by RNA splicing during maturation of the final RNA product. The term intron refers to both the DNA sequence within a gene and the corresponding sequence in the primary RNA transcript. Sequences that are joined together in the final mature RNA, after RNA splicing, are called exons.

Introns are found in the genes of most organisms and many viruses and can be located in a wide range of genes, including those that generate proteins, ribosomal RNA (rRNA), and transfer RNA (tRNA). When proteins are generated from intron-containing genes, *RNA splicing* takes place as part of the RNA processing pathway that follows transcription and precedes translation.

The word intron is derived from the term *intra-genic region*, meaning a region inside a gene; they are sometimes called intervening sequences [<https://en.wikipedia.org/wiki/Intron>]. The term 'intervening sequence', though, can refer to any of several families of internal nucleic acid sequences, in addition to introns, that are not present in the final gene product, including *inteins* ('protein introns' which are segments of a protein able to excise themselves and join the remaining portions [the *exteins*] with a peptide bond in a process termed protein splicing), *untranslated sequences* (UTR), and nucleotides removed by *RNA editing*.

At least four distinct classes of introns have been identified:

- introns in nuclear protein-coding genes that are removed by *spliceosomes* (called spliceosomal introns);
- introns in nuclear and transfer RNA genes that are removed by proteins (*tRNA introns*);
- self-splicing group I introns that are removed by RNA catalysis;
- self-splicing group II introns that are removed by RNA catalysis;
- there is a fifth type, called Group III introns, which are possibly related to spliceosomal introns but too little is known about how their splicing takes place.

Eukaryotic protein-coding genes are interrupted by spliceosomal introns, which are removed from transcripts before protein translation. The first fungal genomes characterised had low intron densities: the yeasts *Schizosaccharomyces pombe* (average 0.9 introns per gene) and *Saccharomyces cerevisiae* (average 0.05 introns per gene). However, among filamentous ascomycete fungi, *Neurospora crassa* and *Aspergillus nidulans* have much higher intron densities (2 to 3 per gene), and average intron densities in basidiomycete and zygomycete fungi have proved to be among the highest known among eukaryotes (4 to 6 per gene on average). Several fungal species share many intron positions with distantly related species; many intron positions are shared between plants and animals but there has been a general loss of introns in fungi.

Both the fungal ancestor and fungus-animal ancestor (of the Opisthokont lineage) were very intron rich, with intron densities matching or exceeding the highest known average densities in modern species of fungi and approaching the highest known across eukaryotes. *Fungal evolution has been dominated by intron loss* with nearly complete intron loss along some fungal lineages. Avoiding extremes, the average picture is of moderate intron densities in the common ancestors followed by a tripling of intron number in vertebrates and plants, massive intron loss in yeasts like *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, and variable intron loss in other fungi (Irimia & Roy, 2014; Phasha *et al.*, 2017).

2.2. Alternative splicing

Introns are transcribed into complementary RNA (the *primary transcript RNA*). They have to be removed from the primary transcript by the *spliceosome* to generate the *messenger RNA* which is translated into protein. If there are several introns in the gene sequence, there may be several ways of removing them, and several alternative mRNAs that could be spliced together as a result.

For example, one gene in the ascomycete *Verticillium dahliae* (cause of wilt diseases in many plants) has been shown to use six different splice sites to produce up to five mature mRNAs. This is called *alternative splicing* of the primary transcripts of protein-coding genes. It is a major post-transcriptional regulatory mechanism which, in addition to regulation of transcription itself, provides the complex diversity of the transcriptome and proteome that characterises eukaryotes.

Alternative splicing is common throughout eukaryotes. Transcriptome sequencing has shown that almost 94% of human genes are alternatively spliced. In plants, estimates of alternative splicing vary from about 60% of intron-containing genes in *Arabidopsis*, 52% in soybean, 40% in cotton, 40% in maize, to 33% in rice. In fungi, it has been estimated that on average, about 6 to 7% of the genes are affected by alternative splicing, but the extent of alternative splicing varies across the kingdom (and varies between species, too).

In general, the number of splice variants found is lowest in the yeasts (3 in *Schizosaccharomyces pombe*; 9 in *Saccharomyces cerevisiae*), somewhat higher in filamentous ascomycetes (20 in *Neurospora crassa*; 100 in *Aspergillus nidulans*; 231 in *Fusarium graminearum*; 861 in *Coccidioides immitis*), and higher still in basidiomycetes (4,819 in *Schizophyllum commune*).

The higher rates of alternative splicing are associated with developmental complexity and with a pathogenic lifestyle, particularly in genes involved in functions of stress response and dimorphic switching. It has been shown that alternatively spliced transcripts are regulated differentially in development.

Alternative splicing is an important regulatory mechanism, which in many eukaryotes increases the coding capacity from a limited set of genes to provide the additional complexity to the proteome that may be required for more elaborate cell functions. However, even in mammalian cells, including humans, not normally thought of as having 'a limited set of genes', *most genes are alternatively spliced*, and mutations in alternative transcripts can give rise to diseases such as cancer.

In fungi, genes involved in virulence in fungal pathogens, genes specifying transcription factors, genes involved in cell growth and morphogenesis have all been reported to be regulated by alternative splicing (Grützmann *et al.*, 2014; Irimia & Roy, 2014; Gehrman *et al.*, 2016; Jin *et al.*, 2017).

2.3. Transposons

Transposons (transposable elements, or TEs) are ubiquitous and vital components of almost all prokaryotic and eukaryotic genomes. Eukaryotic transposons are classified into two main classes:

Class I elements, also known as *retrotransposons*, use an RNA intermediate during transposition, which is transcribed from its DNA template; the reverse transcriptase which does this is often encoded by the TE itself.

Class II TEs form a large and diverse group of mobile elements, but the most important of these in fungi are those with a '*cut-and-paste*' transposition mechanism that does not involve an RNA intermediate. These transpositions are catalysed by an endonuclease (transposase), which is encoded within the TE; the basic architecture of which comprises a transposase and terminal inverted repeats (*TIRs*), which are the excision sites for the transposase. The transposase makes a staggered cut at the excision sites producing sticky ends and cuts out the DNA transposon to ligate it into target sites elsewhere in the genome.

Some transposases bind non-specifically to any target site in the DNA; others bind to specific target sequences. DNA polymerase fills in the single-strand gaps resulting from the sticky ends and DNA ligase closes lesions in the sugar-phosphate backbone. This results in excision site duplication and the insertion sites of DNA transposons can be identified by short direct repeats (resulting from the DNA polymerase repair of the staggered cut in the target DNA) followed by inverted repeats of the excision sites (required for any future TE excision by transposase).

Cut-and-paste TEs may be duplicated if their transposition takes place during S phase of the cell cycle while the DNA is being replicated and this can result in gene duplication, which plays an important role in genomic evolution.

Fungal genomes are exceptionally variable in their TE content, varying over the range 0.02 to 29.8% of their genome consisting of transposable elements and, like other eukaryotes, each

fungal transposable element is either of class I or of class II. Here again, though, there is tremendous variability, with some genomes (for example, two strains of *Pleurotus ostreatus*) being populated mainly by Class I elements.

A survey of 1,730 fungal genomes for transposable elements found DNA TEs across the whole data set but with an uneven distribution in terms of both TE classification and fungal classification. TE content generally correlated with genome size, and TE count is associated with the lifestyle, being elevated in mycorrhizas and diminished in animal parasites.

Interestingly, TEs are opposed by several fungal genome defence mechanisms including Repeat-Induced Point mutation (RIP) and RNA interference (where RNA molecules inhibit gene expression or translation, by neutralizing targeted mRNA molecules; now called RNAi, but also known as co-suppression, post-transcriptional gene silencing and quelling).

Fungi that possessed RIP and RNAi systems had more total TE sequences but fewer elements retaining a functional transposase. This indicates stringent control over transposition and an expression of epigenetic defence intended to suppress TE expression and limit their proliferation (Gladyshev, 2017).

There are very few DNA transposons in genomes belonging to the oldest fungal lineages; the Cryptomycota, Microsporidia, Chytridiomycota, and Blastocladiomycota. Lower terrestrial fungi vary in their TE composition: Glomeromycotina have large genomes with more than 80,000 copies of DNA TEs, but only 59 have been found in *Mortierella alpina* (Mortierellomycotina) (though this genome had about 4,000 remnant copies, that lacked transposase), and 165 in *Mortierella elongata*.

Genome architectures of Ascomycota also varied significantly. Most members of Saccharomycetes had fewer than 20 TE copies with a transposase domain whereas species of *Erysiphe*, *Tuber*, and *Pseudogymnoascus* could have thousands of DNA TEs. Among Basidiomycota, two contrasting genome architectures have been distinguished: those with compact genomes with only a handful of transposons (Ustilaginomycotina, Microbotryomycetes) and those with large genomes with a very large number of transposons, e.g. Agaricomycetes (with up to a thousand TEs) and Pucciniomycetes (with several thousand TEs) (Kempken & Kück, 1998; Castanera *et al.*, 2016; Muszewska *et al.*, 2017).

The occurrence of 'cut-and-paste' transposons in many eukaryotic lineages and their similarity to the prokaryotic insertion sequences suggest that eukaryotic TEs may be older than the last common eukaryotic ancestor.

TEs shape genomes by recombination and transposition; they lead to chromosomal rearrangements; they create new gene neighbourhoods; they alter gene expression by introducing new regulatory sequences for established host genes and they play key roles in adaptation to new life styles like mutualism/symbiosis and pathogenicity by duplicating host genes, so they can take on new roles without endangering their original functions. Pritham (2009) described eukaryotic genomes as containing: '...a menagerie of populations of transposable elements...' and stated that '...it is evident that these elements have played an important role in genome evolution...'

TEs are thought to have been responsible for assembling the *Metabolic Gene Clusters* (MGC), which are common features of most *fungi* genomes but rarely found in other eukaryotes, though they are common in prokaryotes. MGCs are defined as: ‘...tightly linked sets of mostly non-homologous genes involved in a common, discrete metabolic pathway...’

They encode various functions in fungi; nutrient acquisition, synthesis and/or degradation of metabolites, etc., and as well as encoding the enzymes that perform these anabolic or catabolic processes, MGCs often contain appropriate *regulatory sequences*, and those that code for production of toxins also include the mechanisms needed to protect their fungal resident from the toxins. This modular nature of MGCs contributes to the metabolic and ecological adaptability of fungi. MGCs enable easy pathway amplification by gene duplication, and the duplication event can also be engineered by the TEs that assembled the cluster.

Indeed, as well as assembly, TEs are capable of *transposing MGCs*, either to new sites in the same genome, or perhaps to other nuclei in the same heterokaryons, or to other organisms entirely. This last possibility is called *horizontal gene transfer*. There is evidence for many horizontal gene transfer events in fungi; events that have greatly enhanced the basic lifestyle of the fungi concerned (Richards *et al.*, 2011; Slot *et al.*, 2017; Steenkamp *et al.*, 2018).

3. PLOIDY AND GENOMIC VARIATION

A characteristic feature of fungi is the presence of large number of nuclei in a common cytoplasm. Even in fungi with septate hyphae the septa are perforated to some degree, so the mycelium is essentially coenocytic (although the fact that neighbouring hyphal cells can show very different differentiation states on the two sides of what appear to be open septal pores suggests that hyphal compartments can be physiologically distinct). As far as the primary genetic function of nuclei is concerned, the fungal mycelium is commonly heterokaryotic. Heterokaryosis refers to the presence of two or more genetically distinct nuclei within the same hypha. It is uncommon in all other organisms, but heterokaryosis is a hallmark of kingdom Fungi (Roberts & Gladfelter, 2016; Strom & Bushley, 2016).

Hyphal fusion between different fungal individuals is limited by vegetative compatibility barriers. However, these compatibility barriers are not absolute, and exchange of nuclei between hyphae of different species is now believed to enhance fungal diversification. Such an event produces a fungal chimera, which is an organism that contains cells or tissues from two or more different species, and this can enhance diversification at the species level by allowing horizontal gene transfer between mycelia that are too distantly related to hybridise sexually (Roper *et al.*, 2013).

Polyploidy, featuring past and recent whole-genome duplications, is a major evolutionary process in eukaryotes, particularly in plants and, but to a less extent, in animals; and it also occurs in fungi. Many fungi undergo ploidy changes during adaptation to adverse or new environments. Some fungi exist as stable haploid, diploid, or polyploid (triploid, tetraploid) hyphae, while others change ploidy under some conditions and revert to the original ploidy level under other conditions. Aneuploidy (an abnormal chromosome number) is sometimes observed in fungi exposed to new or stressful environments or resulting from an earlier ploidy change. The parasexual cycle is a sequence resulting in random chromosome loss over

several divisions, so a diploid is reduced to a haploid state through a series of aneuploid intermediates (Stukenbrock & Croll, 2014).

Ploidy can increase through mating, endoreduplication, which is replication of the nuclear genome in the absence of mitosis, or failure of cytokinesis after replication. Evidently, some fungi have evolved the ability to tolerate large genome size changes and generate vast genomic heterogeneity without using the meiotic reduction division; indeed, the evolutionary history of *Saccharomyces* species has been shaped by past and recent whole genome duplication events (Albertin & Marullo, 2012; Todd *et al.*, 2017). Species of *Armillaria* are unusual in having diploid tissues in the (mushroom) fruit body though this is produced by a dikaryotic mycelium.

4. SEQUENCING FUNGAL GENOMES

Very little of what we have described in the Sections above could have been written before the complete sequences of fungal genomes became available, so the topic of genome sequencing deserves some discussion.

The priority of genome sequencing is to establish the number, disposition and function of genes in an organism. Genomics is the systematic study of an organism's genome. Consideration of the many uses of a genome sequence started by focussing on the human genome (Sharman, 2001) and came up with these activities:

- studying the proteins and RNA of the *proteome* and *transcriptome* (and perhaps deciding how to change them to serve our own purposes);
- establishing the genetic basis of interactions between organisms, especially pathogenesis and the mechanisms of disease, but including more benign relationships such as mutualisms and mycorrhizas;
- comparing genome sequences from related organisms to examine genome evolution and relationships between organisms at the genomic level: for example, how/if genes are conserved in different species; how relationships between genomes compare with conventional taxonomic classifications, which are of course based upon the outcome of information encoded in the genome; and studying mechanisms of speciation.

The Human Genome Project began with Sanger sequencing technology; the *chain termination method* which is now referred to as the 'first generation technology' of genome sequencing. This procedure is technically undemanding, but time-consuming and labour-intensive. Over the years various aspects were automated and the development of marker nucleotides labelled with four different fluorescent labels (fluorochrome labelling) enabled development of sequencing machines that have a fluorescence detector that can discriminate between the different labels.

Automatic sequencing machines rely on capillary electrophoresis rather than slab gels. The capillary is filled with buffer solution at a certain pH value. Fluorescently labelled PCR products of various lengths are separated in the capillary according to their size, but the separating force is the difference in charge to size ratio (not their ability to filter-flow through a gel).

In other words, size is measured by the overall negative charge, and the longer the fragment, the more negative the charge it bears. As the fragments are driven toward the positive electrode of a capillary by the electric field, they pass a laser beam that triggers a flash of light from the fluorochrome attached to the marker nucleotide that is characteristic of the base type (for example, green for A, yellow for T, blue for G, red for C).

In this way, the genome is carefully read by the machine in one pass; and, of course, the machine can examine many capillary gels in each run. A single machine can sequence half a million bases per day; and then continue into the night without complaint. This is the start of the improvement in speed and accuracy, and reduction in manpower and cost of genome sequencing technology that has happened since completion of the human genome project in 2003.

Subsequent replacement of the electrophoretic capillary with a flow cell, miniaturisation, and use of high-throughput and massively parallel processing brought us to present day '*Next-Generation Sequencing*' (NGS), also called *second-generation sequencing*; a phrase used to describe several different modern sequencing technologies, which differ in engineering configurations and sequencing chemistry.

Some of these platforms can sequence one million to 43 billion 'short reads', of sequence fragments of 50-400 bases each, in each instrument run [view: https://en.wikipedia.org/wiki/Massive_parallel_sequencing]. For more details, we suggest you check out the European Bioinformatics Institute (EMBL-EBI) **online video lectures** at this URL: <https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course>.

Genomics is the systematic study of the genome of an organism and 'systematic study' may well involve comparison with the genomic sequences of other organisms; and phylogenetic study may involve comparisons with many other genomes. Genomics characteristically involves large data sets because it deals with DNA sequences by the megabase. Overall the word genomics has come to embrace a considerable range of activities that can be 'structural' (these have defined endpoints that are reached when the structural determination is complete) or 'functional' (which are more open-ended because additional aspects of function can be added continually).

Genomics requires the use of a combination of different methods, including:

- DNA mapping and sequencing;
- Collecting genome variation;
- Transcriptional control of genes;
- Transcriptional networks that integrate functions of, potentially, many genes;
- Protein interaction networks, which are similarly potentially very extensive;
- Signalling networks.

Genomics has enabled the expansionist approach to be taken to biology. Rather than being restricted by the techniques to concentrate on how individual parts of the organism work in isolation, the biologist can now expect to investigate how many (ultimately, perhaps all) parts of the organism work together. The newly coined expression '*omics*', although originally

informal, is being increasingly used to refer to fields of study of genome biology by adding the ending '-omics'. The related suffix '-ome' is used to describe the objects of study of such fields.

Some examples are:

- Genomics/genome, the complete gene complement of an organism;
- Transcriptomics/transcriptome, all mRNA expressed transcripts;
- Proteomics/proteome, all translated proteins;
- Metabolomics/metabolome, the set of metabolites, the small molecule intermediates and products, of primary and secondary metabolism.

All of these fields of study contribute to **Systems Biology**, which is an holistic (rather than reductionist) scientific approach focussing, often with mathematical and computational modelling, on a wide range of complex interactions in biological systems (see the Wikipedia definition at https://en.wikipedia.org/wiki/Systems_biology). Horgan & Kenny (2011) explain the rationale this way: '...The basic aspect of these approaches is that a complex system can be understood more thoroughly if considered as a whole...Systems biology and omics experiments differ from traditional studies, which are largely hypothesis-driven or reductionist. By contrast, systems biology experiments are hypothesis-generating, using holistic approaches where no hypothesis is known or prescribed but all data are acquired and analysed to define a hypothesis that can be further tested...'

Apart from the four 'omics' fields of study outlined above, there are several other similar terms in use today: taxonomics/taxome is the sum of all the described species and higher groups (genera, families, phyla) of all life, or the sum of all valid taxa of a particular lifeform (often specified, for example, beetle taxome, rust taxome, etc.); phylogenomics (at the time of writing 'phylogenome' is not defined) involves the reconstruction of evolutionary relationships by comparing sequences of whole genomes or portions of genomes; an interactome is the whole set of molecular interactions in a specific biological cell; and a functome, is the complete set of functional molecular units in biological cells.

The omics wiki site [<http://omics.org/>] describes many more. Check out the *History of Omics: as a generic name for various omics and a standalone biology discipline* by Jong Bhak at this URL: http://omics.org/index.php/History_of_Omics. He describes using a computer program to generate tens of thousands of omics terms, one of which is **textome**, which is 'the complete set of biological literature that contain useful information when combined to generate new information through bioinformatics'.

The ability to sequence and compare complete genomes is improving our understanding of many areas of biology. Such data more directly reveal evolutionary relationships and indicate how pathogens spread and cause disease. They enable us to approach a comprehensive understanding of the activities of living cells and how they are controlled at the molecular level. The information has practical value, too. This is why so many pharmaceutical companies are involved in genome projects: the hope is that it will be possible to identify genes responsible for, or which have influence on, diseases, and then design therapies to combat disease directly (Sharma, 2015; Taylor *et al.*, 2017).

Functional genomics now dominates; sequencing an entire genome is only the beginning of functional studies of the transcriptome, proteome, and the metabolome so that our understanding of the organism is holistic. The genome alone is not context sensitive because it is the full set of genetic information. Instead, the transcriptome, proteome and metabolome are all context sensitive because what they comprise depends upon the instantaneous regulatory status of the cell. Or, as the old-time segregational (Mendelian) geneticists at the beginning of the twentieth century put it: phenotype = genotype + environment.

5. ANNOTATING THE GENOME

The process of '*annotating the genome*' starts once the genome sequence has been established and its assembly completed. Annotation is the association of its component sequences with specific functions, and, if the *Saccharomyces cerevisiae* example is a guide, this process can continue for a long time. Annotation requires sophisticated computation, that is: it is an *in silico* analysis. Gene identification is probably the most difficult problem and relies on computer programs that align sequences and use '*gene finder*' programs.

Gene finding is easier with bacterial genomes, in which computer programs can find 97-99% of all genes automatically. In eukaryotes both gene finding, and gene function assignment remain challenging tasks.

Sense is made of genome sequences by annotation *in silico* to:

- identify ORFs by their start and finish codons, and allowing for the minimum length of functional proteins;
- detect the presence of recognisable functional motifs in segments of the deduced gene or protein;
- compare against known protein or DNA sequences using homologous genes from the same or other genomes.

Further annotation is done experimentally by:

- classical gene cloning and functional analysis;
- analysis of cDNA clones or EST sequences (an *expressed sequence tag* or EST is a short component sequence of a transcribed cDNA, so it is a portion of an expressed gene), and gene expression data.

No single method of genome annotation is comprehensive; all have their limitations, so they must be used in concert. Many of the genes identified in sequencing projects will be 'new' in the sense that when the sequence is identified the gene function is unknown. Establishing the cellular role of such new ORFs requires a different set of bioinformatics tools that integrate sequence information with the accumulated knowledge of metabolism so that conjectures can be made about likely functions. Those predictions are then tested experimentally by using heterologous expression, gene knockouts, and characterisation of purified proteins. Parallel analysis of phylogenetically diverse genomes can also help in understanding the physiology of the organism whose genome is being sequenced. When the sequence of the whole genome has been established and annotated, the genome can be compared with others on the databases.

The yeast *Saccharomyces cerevisiae* is a well-established model organism which, since the days of Louis Pasteur, has a long history in physiology, biochemistry and molecular biology. The genome of yeast continues to be a useful model for eukaryotes, comprising a grand total of 12.1 Mb distributed over 16 chromosomes, which range in size between 250 kb and more than 2.5 Mb. The yeast genome-sequencing project was started in 1989. The sequence of chromosome III was the first to be published in 1992, chromosomes II and XI followed in 1994, and the sequence of the entire genome was released in April 1996. Quality control measures ensured a 99.97% level of accuracy of the sequence.

Today, the place to learn about this genome is the Saccharomyces Genome Database (SGB) website at <https://www.yeastgenome.org/> and the Yeast Genome Snapshot at <https://www.yeastgenome.org/genomesnapshot>. As of September 2019, there were 6,604 open reading frames (ORFs) which possibly encode metabolically active proteins, of which 5,180 were verified, 735 were uncharacterised, and 689 were considered dubious.

On average, a protein-encoding gene is found every two kb in the yeast genome. The ORFs vary from 100 to more than 4,000 codons, although two-thirds are less than 500 codons, and they are evenly distributed on the two strands of the DNA. In addition to these, the yeast genome contains 27 rRNA genes in a large tandem array on chromosome XII, 77 genes for small nuclear RNAs, 277 tRNA genes (belonging to 42 codon families) scattered across the chromosomes, and 51 copies of the yeast retrotransposons (Ty elements).

There are also non-chromosomal elements, most notably the yeast mitochondrial genome (80 kb) and the 6 kb 2 μ plasmid DNA, but there may be other plasmids, too. So, 25 years after the genome was sequenced, only 78% of the ORFs have been verified; a rate of progress that makes it even more amazing that on April 17, 2018, SGB announced a single publication in the journal *Nature* by a team of researchers jointly led by Joseph Schacherer and Gianni Liti, that reported the whole-genome sequences and phenotypes of no fewer than 1,011 different *Saccharomyces cerevisiae* yeast strains (Peter *et al.*, 2018). Gathering isolates from many different geographical locations and ecological niches (wineries, breweries and bakeries, but also from rotting bananas, sea water, human blood, sewage, termite mounds, and more), the authors surveyed evolutionary relationships among the strains to describe the worldwide population distribution of this species and deduce its historical spread.

This unusually large-scale population genomic survey demonstrates that the likely geographic origin of *S. cerevisiae* lies somewhere in East Asia. Budding yeast began spreading around the globe about 15,000 years ago and was subjected to several independent domestication events during its worldwide journey. For example, whereas genomic markers of domestication appeared about 4,000 years ago in sake yeast, such markers appeared in wine yeast only 1,500 years ago. Domesticated isolates exhibited high variation in ploidy, aneuploidy and genome content, while genome evolution in wild isolates was mainly driven by the accumulation of single nucleotide polymorphisms, most of which are present at very low frequencies.

The alleged purpose of study of a model organism like yeast is the expectation that its analysis will enable the identification of genes relevant to disease in humans; and this expectation seems to be fulfilled. Comparing the sequences of human genes available in the sequence

databases with yeast ORFs shows that over 30% of yeast genes have homologues among the human sequences, most of these representing basic cell functions. Finding this sort of homology can contribute to the understanding of human disease.

The first example of this seems to be *Friedreich ataxia*, which is the most common type of inherited ataxia (loss of control of bodily movements) in humans, the biochemistry of which was uncovered by demonstrating homology to a yeast ORF of known function. Friedreich's ataxia is caused by enlargement of a GAA repeat in an intron that results in decreased expression of the frataxin gene; frataxin is a highly conserved iron-binding protein present in most organisms, and Friedreich's ataxia pathology is associated with disruption of iron-sulfur cluster biosynthesis, mitochondrial iron overload, and oxidative stress.

Frataxin is the human mitochondrial protein that has homologues in yeast. In yeast, mutants defective in the frataxin homologue accumulate iron in mitochondria and show increased sensitivity to oxidative stress. Biosynthesis of Fe-S clusters in yeast is a vital process involving the delivery of elemental iron and sulfur to scaffold proteins and the architecture of the protein complex to which frataxin contributes is essential to ensure concerted and protected transfer of potentially toxic iron and sulfur atoms to the mitochondrion. This comparison suggests that Friedreich's ataxia is caused by mitochondrial dysfunction and may point towards novel methods of treatment (Pastore & Puccio, 2013; Ranatunga *et al.*, 2016).

In many ways, this kind of comparison alone can justify all the effort devoted to sequencing the yeast genome. Functional genomics studies the roles of genes and proteins to define gene/protein function. The outcome is known as the Gene Ontology. Originally, ontology was a branch of metaphysics; a philosophical inquiry into the nature of being. For the computer scientist, ontology is the rigorous collection and organisation of knowledge about a specific feature. The aims of Gene Ontology (GO) are to:

- develop and standardise the vocabulary about the attributes of genes and gene products that is species-neutral, and equally applicable to prokaryotes and eukaryotes, and uni- and multicellular organisms;
- annotate genes and gene products within sequences, and coordinate understanding and distribution of annotation data;
- and provide bioinformatics tools to aid access to all these data.

To achieve all this, there are three organising principles of GO to describe the function of any gene/protein sequence as follows:

- Biological process; effectively the answer to the question why does the sequence exist? This can be cast in very broad terms describing the biological goals accomplished by function of the sequence, for example mitosis, meiosis, mating, purine metabolism, etc.
- Molecular function; effectively what does the sequence do? The tasks performed by individual gene products, for example transcription factor, DNA helicase, kinase, phosphatase, phosphodiesterase, dehydrogenase, etc.
- Cellular component; where is that function exercised? The location in subcellular structures and macromolecular complexes. For example, nucleus, telomere, cell wall, plasma membrane, endoplasmic reticulum lumen, etc.

The ontology data are freely available from the Gene Ontology Consortium's website at this URL: <http://www.geneontology.org/>. General information about genomics is accessible through the Broad Institute's listings at <https://www.broadinstitute.org/>.

Annotation has been automated by *annotation programs* (available online) that quickly identify ORFs for hypothetical genes in a genome. Many sequences are conserved across large evolutionary distances, so many functional assignments can be inferred using information already available from other organisms; this sequence search and comparison process can also be automated.

Annotating the genes of filamentous fungi, even other Ascomycota and close relatives of *Saccharomyces cerevisiae*, is more demanding because their genomes are much larger and their gene structure more complex than those of yeast. Genes of filamentous fungi often contain multiple introns (section 3, above), with some within the open reading frame of the gene (very few yeast genes contain introns, those that do have a single intron at the start of the coding sequence, often interrupting the initiation codon). Also, the intron-boundary sequences may not become evident until the transcriptome is analysed, and alternative splicing events catalogued (section 4, above).

The greater complexity of gene structure in filamentous fungi demands independent data on gene expression to make confident functional assignments. Methods have been described that use cDNA or EST sequence alignments, and gene expression data to predict reliably the function of *Aspergillus nidulans* genes. We recommend you read the discussion and explanation of the approach by Sims *et al.* (2004).

Yandell & Ence (2012) have published '*A beginner's guide to eukaryotic genome annotation*' and further information and advice is freely available online at:

- Bioinformatics Platform for Agroecosystem Arthropods at <https://bipaa.genouest.org/is/how-to-annotate-a-genome/>;
- *Annotation for Amateurs* tutorial on the PlantGDB website at this URL: <http://www.plantgdb.org/tutorial/annotatemodule/>;
- DNA annotation pages on the *Wikipedia* page at https://en.wikipedia.org/wiki/DNA_annotation.

The most up-to-date information on the genes of any organism in which you are interested can be obtained from the website devoted to that organism (use your preferred web search engine to find it). For example, entering '*coprinopsis cinerea genome*' into the search engine finds the *Coprinopsis cinerea* home page, which gives you general information about the organism and its genome, on the JGI Genome Portal, which is at this URL: [\[https://genome.jgi.doe.gov/Copci1/Copci1.home.html\]](https://genome.jgi.doe.gov/Copci1/Copci1.home.html). This page has a menu of hyperlinks across the top that give access to the deepest detail about the genome of this species. The main Internet sites for fungal genomic data are discussed in section 9, below.

Bioinformatics is essentially the use of computers to process biological information when computation is necessary to manage, process, and understand very large amounts of data. Although there are many bioinformatics tools and databases, using them effectively often requires specialised knowledge; where this is lacking, the BioStar platform can help. Biostar is

an online forum where experts and those seeking solutions to problems of computational biology exchange ideas. BioStar can be accessed at <https://www.biostars.org/> (Parnell *et al.*, 2011).

Bioinformatics is particularly important as an adjunct to genomics research, because of the large amount of complex data this type of research generates, so to a great extent the word, and the approaches it encompasses, have become synonymous with the use of computers to store, search and characterise genomics, transcriptomics, proteomics and metabolomics data. But there are other large data sets in need of analysis that rightly fall within range of the fundamental definition of the word 'bioinformatics'. These are large data sets arising from:

- Survey data and censuses, particularly, but not only, those involving automatic data capture, and 'surveys of surveys' (known as *metadata*).
- Data generated by mathematical models that seek to simulate a biological system and its behaviour in time.

The aim of functional genomics is to determine the biological function of all the genes and their products, how they are regulated and how they interact with other genes and gene products. Add interactions with the environment and this is fully integrated biology; what has come to be known as systems biology (Klipp *et al.*, 2009; Nagasaki *et al.*, 2009; Horgan & Kenny, 2011). Comprehensive studies of such large collections of molecules as occur in the transcriptome, proteome, and metabolome require what are described as high throughput methods of analysis at each stage from the generation of mutants through to the determination of which proteins are associated with which functions. Each stage generates massive amounts of data that are qualitatively and quantitatively different, which must be integrated to allow construction of realistic models of the living system (Delneri *et al.*, 2001).

Methods that generate large numbers of gene mutants and simultaneously identify the mutants and/or their products in ways amenable to automation was the start of the **high throughput approach** (Ross-Macdonald *et al.*, 1999; Cho *et al.*, 2006; Caracuel-Rios & Talbot, 2008; Foster *et al.*, 2006; Honda & Selker, 2009). There is scope for large scale international collaboration in this sort of exercise and 1999 saw the establishment of a collection of mutant yeast strains, each bearing a defined deletion in one of 6,000+ potential protein encoding genes in yeast (Winzeler *et al.*, 1999). This is the EUROSCARF collection (EUROpean Saccharomyces Cerevisiae ARchive for Functional analysis; see <http://www.euroscarf.de/>).

Transcriptome analyses and can be studied in a fully comprehensive manner using hybridisation-array analysis, which is described as a massively parallel technique because it allows so many sequences to be examined at one time. Remember, though, that mRNA molecules transmit instructions for synthesising proteins; they do not function otherwise in the workings of the cell, so transcriptome analyses are considered to be an indirect approach to functional genomics. The transcriptome comprises the complete set of mRNAs synthesised in the cell under any given well-defined set of physiological conditions. Unlike the genome, which has a fixed collection of sequences, the transcriptome is context dependent, which means that its content of sequences depends on the cell response to the current set of physiological circumstances, and the make-up of that set will change when the physiological circumstances change.

Those physiological circumstances will be adapted in response to changes in both the intracellular and extracellular environment of the cell; its nutritional status, state of differentiation, age, etc. The mRNA of genes that are newly expressed (up-regulated) will appear in the sequence collection, and the mRNA of genes that are not expressed (down-regulated) in the new circumstance will disappear from, or be greatly reduced in, the sequence collection. The pattern of mRNA content in the transcriptome reveals the pattern of gene regulation.

Hybridisation arrays are now used widely to study the transcriptome because of their ability to measure the expression of many genes with great efficiency. Microarrays permit assessment of the relative expression levels of hundreds, even thousands, of genes in a single experiment. Hybridisation arrays are also called *DNA micro- or macroarrays*, *DNA chips*, *gene chips*, and *bio chips* (Nowrousian, 2007, 2014).

The web definition of 'DNA microarray' is: a collection of microscopic DNA spots attached to a solid surface forming an array; used to measure the expression levels of many genes simultaneously (https://en.wiktionary.org/wiki/DNA_microarray).

The array of single-stranded DNA molecules is typically distributed on glass, a nylon membrane, or silicon wafer (any of which might be called 'a chip'), each being immobilised at a specific location on the chip in a predetermined (and computer-recorded) grid formation. Microarrays and macroarrays differ in the size of the sample spots of DNA; in *macroarrays* the size of the spot is over 300 µm, in *microarrays* it is less than 200 µm. Macroarrays are normally spotted by high-speed robotics onto nylon membranes, microarrays are made on glass or quartz surfaces (usually called custom arrays) (GeneChip®, from Affymetrix Inc.; see <https://www.affymetrix.com/>) (Lipshutz *et al.*, 1999).

The immobilisation onto the solid matrix is the most crucial aspect of the technique as it must preserve the biological activity of the molecules. The spotted material can be genomic DNA, cDNA, PCR products (any of these sized between 500 to 5,000 base pairs) or oligonucleotides (20 to 80-mer oligos). The identities and locations of the single-stranded DNAs are known, so when the chip is treated with a suspension of experimental cDNA molecules prepared from a set of mRNAs, the cDNAs complementary to those on the chip will bind to those specific spots. The complementary binding pattern can be detected and since the DNAs at each position on each grid are known, the complementary binding pattern indicates the pattern of gene expression in the sample.

Macroarrays are hybridised using a radioactive probe; normally ^{33}P , an isotope of phosphorus which decays by β -emission so that the decay, and therefore the position of the complementary binding can be imaged with a *phosphorimager*, a device in which β -particle emissions excite the phosphor molecules on the plate in a way that can be detected by scanning the plate with a laser and the attached computer converts the energy it detects to an image in which different colours represent different levels of radioactivity.

Microarrays are exposed to a set of targets either separately (single dye experiment) or in a mixture (two dye experiment) to determine the identity/abundance of complementary sequences. Laser excitation of the spots yields an emission with a spectrum characteristic of

the dye(s), which is measured using a scanning confocal laser microscope. Monochrome images from the scanner are imported into software in which the images are pseudo-coloured and merged and combined with information about the DNAs immobilised on the chip. The software outputs an image which shows whether expression of each gene represented on the chip is unchanged, increased (up-regulated) or decreased (down-regulated) relative to a reference sample. In addition, data is accumulated from multiple experiments and can be examined using any number of data mining software tools.

There are many uses for DNA microarrays. Apart from expression profiling to examine the effect of physiological circumstance on gene expression on which we have so far concentrated, hybridisation arrays can be used to:

- dissect metabolic pathways and signalling networks;
- establish transcription factor regulatory patterns, target genes and binding sites;
- compare gene expression in normal tissue with that of diseased tissue, initially to establish which genes are involved in response to disease, and when that is done to diagnose disease;
- identify gene expression of different tissues and different states of cell differentiation to establish tissue-specific and/or differentiation-specific genes;
- study reaction to specific drugs, agrochemicals, antibiotics or toxins to identify drug targets, side effects, and resistance mechanisms.

The proteome is the complete set of proteins synthesised in the cell under a given set of conditions. The traditional method for quantitative proteome analysis combines protein separation by high-resolution 2-dimensional isoelectric focusing (IEF)/SDS-PAGE (2DE) with mass spectrometric (MS) or tandem mass spectrometric (MS/MS) identification of selected protein spots detected in the 2DE gels by use of specific protein stains. Continued improvement in technology is steadily increasing the throughput of protein identifications from complex mixtures and permitting quantification of protein expression levels and how they change in different circumstances (Aebersold, 2003; Bhadauria *et al.*, 2007; Rokas, 2009).

An important feature that arises from analysis of the proteome is the enormous extent and complexity of the network of interactions among proteins and between proteins and other components of the cells. These networks can be visualised as maps of cellular function, depicting potential interactive complexes and signalling pathways. To quote from Jewett *et al.*, 2006: 'Metabolomics consists of strategies to *quantitatively* identify cellular metabolites and to understand how trafficking of these biochemical messengers through the metabolic network influences phenotype'.

Metabolomics is especially important in fungi because these organisms are widely used to produce chemicals. The main difficulty in metabolome analysis is not technical as there are sufficient analytical tools (mass spectrometry is most often used) and mathematical strategies available for extensive metabolite analyses. However, the indirect relationship between the metabolome and the genome raises conceptual difficulties. The biosynthesis or degradation of a single metabolite may involve many genes, and the metabolite itself may impact on many more. Consequently, the bioinformatics tools and software required must be exceptionally powerful (Shen & Zhu, 2019).

6. FUNGAL GENOMES AND THEIR COMPARISON

Saccharomyces cerevisiae is the best-studied fungus, and the fission yeast *Schizosaccharomyces pombe* is also an important model organism for which a complete genome is available (of 13.8 Mb with 4,824 protein-coding genes). However, neither of these yeasts is an adequate model for filamentous fungi, which have more genes (approximately 8,400) and bigger genomes (30 to 40 Mb); both features are presumably related to the wider morphogenetic, metabolic, and ecological capabilities of filamentous fungi.

Certainly, it is already clear that several genes present in filamentous fungi are not present in yeasts, so *comparative genomics* is a growing business. There are several genome projects underway and planned; they include many filamentous fungi. The principal one being the Earth BioGenome Project (EBP), which has been described as ‘a moonshot for biology’. This proposed 10-year project aims to sequence, catalogue, and characterise the genomes of all of Earth’s eukaryotic biodiversity (Lewin *et al.*, 2018).

Comparative genomics is a science in its own right (Gibson & Muse, 2009), and with more than 2,000 already fully sequenced or in progress, the range of fungal genomes available is the widest sampling of genomes from any eukaryotic kingdom. Many of the fungal genomes fall into groups of related species that are ideal for comparative studies (see, for example, Jones, 2007). Nevertheless, the fungi chosen for sequencing initially were mostly pathogens or model organisms and dealing with this bias was one aim of the 1,000 Fungal Genomes Project [<http://1000.fungalgenomes.org/home/>] the motto for which is ‘*Sequencing unsampled fungal diversity*’. Another approach is to design genome sequencing programmes with some specific objective in mind such as development of alternative bioenergy sources, bioremediation, and fungus-environment interactions (Baker *et al.*, 2008).

The first global initiative to sequence and annotate fungal genomes was managed and coordinated by the Broad Institute of MIT and Harvard under what was called the *Fungal Genome Initiative* (FGI), which is still described at this URL: <https://www.broadinstitute.org/fungal-genome-initiative>. FGI prioritised sequence data from fungi that are important to medicine, agriculture and industry and established a sequence database for that purpose. Over 100 fungi have been sequenced in this programme, including human and plant pathogens as well as fungi that serve as basic models for molecular and cellular biology. Fungal genome websites at the Broad Institute have been changed as the sequencing projects have been completed. Formerly interactive websites have been replaced with static pages providing information on fungal projects, along with links to sites where datasets can still be downloaded, and the primary repositories for all fungal genomic data now are *MycoCosm*, *FungiDB* and *Ensembl Fungi*.

MycoCosm [<https://genome.jgi.doe.gov/programs/fungi/index.isf>] is hosted by the Joint Genome Institute (JGI), a *Department of Energy Office of Science User Facility* managed by *Lawrence Berkeley National Laboratory at the University of California*. MycoCosm is: ‘...JGI’s web-based fungal genomics resource, which integrates fungal genomics data and analytical tools for fungal biologists. It provides navigation through sequenced genomes, genome analysis in context of comparative genomics and genome-centric view...and offers the largest available collection of fungal genomes, for comparative genomics across phylo- and eco-

groups, along with interactive web-based tools for genome downloading, searching and browsing, and a form for nominating new species for sequencing to fill gaps in the *Fungal Tree of Life*.’

This portal also hosts the *1,000 Fungal Genomes Project*, an international collaboration set up to sequence 1,000 fungal genomes (though this number has now been greatly exceeded) [<https://jgi.doe.gov/our-science/science-programs/fungal-genomics/1000-fungal-genomes/>]; and the *Genomic Encyclopedia of Fungi*, which focuses on genomes of fungi that contribute to plant health (including symbiosis, pathogenicity and biocontrol), biorefinery mechanisms (conversion of biopolymers to sugars for fuel production), and fungal diversity [<https://jgi.doe.gov/our-science/science-programs/fungal-genomics/genomic-encyclopedia-of-fungi/#feedstock>] (Grigoriev *et al.*, 2014).

- A list of fungal genomes in the Mycocosm system can be found at this URL: <https://genome.jgi.doe.gov/fungi/fungi.info.html>;
- and progress with the 1,000 Fungal Genomes Project is regularly reported at this URL: <http://1000.fungalgenomes.org/home/>

FungiDB (<http://fungidb.org/fungidb/>; Stajich *et al.*, 2012) is now one of the *EuPathDB* family of databases (this being the *eukaryotic pathogen genomics database resource*) that supports a wide range of microbial eukaryotes; FungiDB (Aurrecoechea *et al.*, 2017) includes many fungal (and oomycete) species, including non-pathogens. This resource provides automated analysis of multiple genomes, curated information, with comments and supporting evidence from the user community. In addition, FungiDB offers sophisticated tools for integrating and mining diverse Omics datasets that fungal biologists will find useful. The FungiDB web site also gives access to a YouTube tutorials channel, web tutorials (videos and PDF-downloads), and teaching exercises.

Ensembl Fungi is a browser for fungal genomes (<http://fungi.ensembl.org/index.html>). The genomes are taken from the databases of the *International Nucleotide Sequence Database Collaboration* (the *European Nucleotide Archive* at the European Bioinformatics Institute [<https://www.ebi.ac.uk/>], *GenBank* at the US National Center for Biotechnology Information [<https://www.ncbi.nlm.nih.gov/>], and the DNA Data Base of Japan [<https://www.ddbj.nig.ac.jp/index-e.html>]). The portal offers an extensive range of tools, downloads and documentation.

- A list of all fungi on the website of Ensembl Fungi is located at this URL: <http://fungi.ensembl.org/species.html>.

As of 2019, well over 1,000 fungal genomes have already been sequenced and annotated or are in the process of being sequenced and annotated (and that total does *not* include the 1,011 *Saccharomyces cerevisiae* genomes published by Peter *et al.*, 2018). We strongly recommend that you visit the websites listed above because the genomic data are updated regularly as improvements and amendments are made to the sequences; but also because the index pages provide hyperlinks that allow you to access, and download, the genome sequences and information about many aspects that we cannot deal with here, including: basic statistics about genome size, gene density, etc., search facilities allowing you to find similarities to other sequences, feature searches to explore and view annotated features on the sequence, gene indexes to find specific genes by a variety of methods, ability to browse

the DNA sequence, find clones, and graphically view sequence regions, opportunity to download sequence, genes, markers, and other genome data.

7. MANIPULATING GENOMES: GENE EDITING

Ultimately, you may think in terms of applying all this knowledge to the creation of something entirely new. That is, to developing a biological system of some form that does not already exist in the biosphere. Of course, since the dawn of agriculture, practical people have been involved in modifying the genomes of their cultivated plants and animals by a combination of *artificial selection* and *selective breeding*, producing crop species (like maize) or domesticated animals (like high milk-yield cattle) that simply could not exist in the wild.

Indeed, although we were unaware of it at the time, by selecting brews or ferments that produced the most satisfactory end products in brewing, baking and other food fermentations (cheese, salami, soy, miso) we have also been unconsciously applying selection pressure to the fungi and bacteria involved in those processes for a very long time; the process is called '*domestication*').

The 'modern' version of this is called *synthetic biology*. Wikipedia defines synthetic biology as: '...the design and construction of novel artificial biological pathways, organisms or devices, or the redesign of existing natural biological systems...' [https://en.wikipedia.org/wiki/Synthetic_biology]. Kaznessis (2007) adds the crucial rider that synthetic biological engineering is emerging from molecular biology as a distinct discipline based on quantification. And that's the real defining feature, this is a branch of biology that verges on engineering (Silver *et al.*, 2014) because it depends on large scale computer processing of large amounts of *numerical* data.

The process starts with genetic engineering and in practice the process proceeds from *functional analysis* by experimentation and can eventually lead to *functional design*. The essential flow of activity is as follows: gene sequence → change or disrupt the DNA (deletion, inactivation by insertion, point mutation) → mutant phenotype → function → alter function → change sequence → new (improved?) phenotype.

This is the essence of functional genomics; being the study of gene function on the genomic scale. In filamentous fungi it is a field of research that has made great advances in very recent years and which continues to advance at rapid pace. Transformation and gene manipulation systems have been developed and applied to many economically important filamentous fungi and oomycetes; overall, the integration of information from the various processes that occur within a cell provides a more complete picture of how genes give rise to biological functions and a better understanding of how those functions can be redesigned.

An important development has been *Agrobacterium*-mediated transformation (AMT). *Agrobacterium tumefaciens* is a gram-negative bacterium which is a common plant pathogen causing crown gall tumours. This tumorous growth of the plant tissue is induced when the bacterium transfers some bacterial DNA (called T-DNA) to the host plant. T-DNA is located on a 200 kbp plasmid (the tumour-inducing or Ti plasmid). The T-DNA integrates into the plant genome, then T-DNA genes that encode enzymes to produce plant growth regulators are

expressed, and their expression results in uncontrolled growth of the plant cells. However, for use as a cloning vector, the T-region of the Ti plasmid can be deleted and replaced by other DNA sequences because plasmid virulence, transfer and integration are controlled by genes elsewhere on the plasmid.

What is significant for our present discussion is that *Agrobacterium tumefaciens* is able to transfer its T-DNA to a very wide range of fungi and produces a significantly higher frequency of more stable transformants than alternative transformation methods (Michielse *et al.*, 2005). AMT is a relatively simple system to work with, primarily because it does not require the production of protoplasts or sphaeroplasts. Indeed, a major attraction of AMT is the variety of starting materials that can be used: protoplasts, spores, mycelium, and pieces of fruit body tissues have all produced successful transformation. The approach seems to be applicable to the full range of fungi (zygomycetes, Ascomycota and Basidiomycota) and shows great potential for fungal biotechnology and medicine (Michielse *et al.*, 2005; Sugui *et al.*, 2005; Idnurm *et al.*, 2017).

But by far the most crucial development in recent years has been *gene editing*. The process depends on engineered nucleases, which can be designed to cut at any location in the genome of any species and introduce modified DNA sequences into the endogenous (host organism) sequence. There are three major classes of engineered nuclease enzyme (we will describe the fourth gene editing system, the CRISPR-Cas system, separately below):

- zinc-finger nucleases (ZFNs),
- transcription activator-like effector nucleases (TALENs) and,
- engineered *meganucleases*.

Engineered nucleases create site-specific double-strand breaks at desired locations in the genome. These fusion proteins serve as readily targetable 'DNA scissors' for gene editing applications that enable targeted genome modifications to be accomplished such as sequence insertion, deletion, repair and replacement in living cells. The induced double-strand breaks are repaired through nonhomologous end-joining or homologous recombination, and the whole process results in precisely targeted mutations ('edits') being incorporated into the experimental genome. This type of gene editing was selected by the journal *Nature Methods* as the 2011 Method of the Year (Anonymous Editorial, 2011). Fundamental to the use of engineered nucleases in genome editing is that the engineered enzymes produce double stranded breaks (DSBs) in the DNA of the target organism. Double strand breaks are cytotoxic lesions that threaten genome integrity and most organisms have mechanisms to repair DSBs (Ceccaldi *et al.*, 2016).

The concept underlying *ZFNs* and *TALENs* technologies is that of a non-specific DNA cutting catalytic domain (obtained from an endonuclease with discrete and separate DNA recognition and cleaving sites) being linked to peptides that recognise specific DNA sequences such as *zinc fingers* (*ZFNs*) and *transcription activator-like effectors* (*TALEs*). Zinc finger motifs occur in several transcription factors. The C-terminal part of each finger is responsible for the specific recognition of a short region (about 3 base pairs) of the DNA sequence. Combining 6 to 8 zinc fingers whose recognition sites have been characterised produces a protein that can target around 20 base pairs of a specific gene. Although the nuclease portions of both *ZFNs* and *TALENs* constructs have similar properties, the difference between these engineered

nucleases is in their DNA recognition peptide. ZFN 'zinc fingers' rely on a combination of cysteine and histidine residues to react with their metal ions so codons for those amino acids identify the nuclease target sequence.

Transcription Activator-Like Effectors (TALEs) are proteins secreted by *Xanthomonas* plant pathogenic bacteria that bind promoter sequences in the host and activate expression of plant genes that aid bacterial infection. They recognise plant DNA sequences through a central repeat domain consisting of a variable number of about 34 amino acid repeats. TALEs can be engineered to bind to practically any desired DNA sequence, so when combined with a Nuclease, the TALENs (which are *artificial, engineered, restriction enzymes*) can cut DNA at the specific location(s) desired by the experimenter. TALEN constructs are used in a similar way to ZFNs but have three advantages in targeted mutagenesis: (i) DNA binding specificity is higher, (ii) off-target effects are lower, and (iii) construction of DNA-binding domains is easier. Meganucleases, discovered in the late 1980s, are *endonucleases characterised by a large recognition site* (DNA sequences of 12 to 40 base pairs). Sites of this length generally occur only once in any given genome, so meganucleases are *the most specific of the naturally occurring restriction enzymes*. Such meganucleases are quite common, but the most valuable tools for gene engineering have been derived from the LAGLIDADG family of endonucleases, so-called for the conservation of a specific amino acid sequence motif which is defined by each letter as a code that identifies a specific residue (the motif is: *Leucine-Alanine-Glycine-Leucine-Isoleucine-Aspartic acid-Alanine-Aspartic acid-Glycine*). This motif binds to a specific DNA sequence; change the amino acid sequence and it will bind to a different DNA sequence.

The 'engineering' aspect of this is that mutagenesis and high throughput screening methods have been used to create meganuclease variants that recognise a defined catalogue of unique DNA sequences. Others have been fused to various meganucleases to create hybrid enzymes that recognise a new sequence and yet others have had the DNA interacting amino acids of the meganuclease altered to design sequence specific meganucleases; all contributing to what is called *rationally designed meganuclease*. Meganucleases have the benefit of causing less toxicity in cells than ZFNs because of more stringent DNA sequence recognition; however, the construction of sequence-specific enzymes for all possible sequences is costly and time consuming. Nevertheless, it can be done. View https://en.wikipedia.org/wiki/Genome_editing to learn more about engineered nucleases.

The **CRISPR-Cas9**-based system has become a common platform for genome editing in a variety of organisms. CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) are genetic elements, which provide bacteria with adaptive immunity to viruses and plasmids. They consist of short sequences that originate as remnants of genes from past infections, sandwiched between unusual, repeated bacterial DNA sequences; the 'clustered regularly interspaced short palindromic repeats' that give CRISPR its name. The CRISPR-associated protein Cas9 is an endonuclease that uses a guide sequence within an RNA duplex, tracrRNA:crRNA, to form base pairs with DNA target sequences, enabling Cas9 to introduce a site-specific double-strand break in the DNA. The dual tracrRNA:crRNA was engineered as a single guide RNA (sgRNA) that retains two critical features:

- a sequence at the 5' side that determines the DNA target site by Watson-Crick base-pairing with the target DNA, and
- a duplex RNA structure at the 3' side that binds to Cas9.

From this, Doudna & Charpentier (2014) created a simple two-component system in which experimenter-determined changes in the guide sequence of the sgRNA direct Cas9 to target the specific DNA sequence of interest to the experimenter. Cas9-sgRNA-mediated DNA cleavage produces a blunt double-stranded break in the target DNA that triggers repair enzymes to disrupt or replace DNA sequences at or near the cleavage site. Catalytically inactive forms of Cas9 can also be used for programmable regulation of transcription and visualisation of genomic loci.

The simplicity of the CRISPR-Cas9 system has made this a cost-effective and easy-to-use technology to target, edit, modify, regulate, and mark genomic loci of a wide array of cells and organisms precisely and efficiently. By introducing plasmids containing Cas genes and specifically constructed CRISPRs into living eukaryotic cells, the eukaryotic genome can be cut at any desired position. This is the quickest and cheapest method for gene editing and requires the least amount of expertise in molecular biology because it is RNA rather than protein that is engineered to guide the nuclease to the target. This is a major advantage that CRISPR has over the ZFN and TALEN methods; it can target different DNA sequences using its ± 80 -nucleotide sgRNAs, while both ZFN and TALEN methods require construction and testing of the proteins created for targeting each DNA sequence. The CRISPR-Cas system was selected by the journal *Science* as its 2015 Breakthrough of the Year (McNutt, 2015); you can read about the latest developments in the '*CRISPR revolution*' topic page written by Jon Cohen (a staff writer for *Science*) at this URL: <http://www.sciencemag.org/topic/crispr>. Gene editing technologies have been developed for application to animals (Dunn & Pinkert, 2014), plants (Mohanta *et al.*, 2017) and fungi (Nødvig *et al.*, 2015; Chen *et al.*, 2017; Pudake *et al.*, 2017; Zheng *et al.*, 2017), and these publications all make fascinating reading.

Anzalone *et al.* (2019) describe a technique they call **prime editing** (or search-and-replace genome editing) as being:

'...a versatile and precise genome editing method that directly writes new genetic information into a specified DNA site using a catalytically impaired Cas9 fused to an engineered reverse transcriptase, programmed with a prime editing guide RNA (pegRNA) that both specifies the target site and encodes the desired edit.'

The authors claim that prime editing greatly expands the scope and capabilities of genome editing, and in principle could correct about 89% of known pathogenic human genetic variants. So, what could it do for fungi?

8. BIOTECHNOLOGICAL OPPORTUNITIES: FUNGI AS CELL FACTORIES

Techniques described above enable the most detailed manipulation of the fungal genome. It's only a slight exaggeration to say that if you can dream of a genetic manipulation, then there's a genome editing technology that can now make it happen (McCluskey & Baker, 2017). Hibbett *et al.* (2013) call the science 'genome-enabled mycology' describing it as being characterised '...by the pervasive use of genome-scale data and associated computational tools in all aspects of fungal biology. Genome-enabled mycology is integrative and often requires teams of researchers with diverse skills in organismal mycology, bioinformatics and molecular biology...'

Their paper discusses the technical and social changes that need to be made to enable all fungal biologists to make use of the new data; and it starts a special issue of the journal *Mycologia* which is devoted to genome-enabled mycology (*Mycologia* 2013, issue 6 of volume 105, contents at this URL: <https://www.tandfonline.com/toc/umyc20/105/6>).

'*Fungal cell factories*' is a phrase that is often used in 'blue skies' discussions about what could be done in the future. This is because the metabolic activities of fungi have already been harnessed for so long in applications ranging from food fermentation to pharmaceutical production that they are naturally thought of as indispensable biotechnological tools. The more so because fungal bioprocesses of earlier generations, like those that produce citric acid and penicillin, and those of today's generation producing lovastatin, have had such positive impacts on human society.

The metabolic and enzymatic diversity encoded in the genomes of fungi will continue to be developed for production of new generations of enzymes, pharmaceuticals, chemicals and biofuels. Though there must be many applications which will only emerge with time and further knowledge; there are some which are immediately obvious. Currently, fungal derived enzymes that degrade plant derived biomass are being utilised for the development of bioprocesses for biofuel and renewable chemical production, particularly the growing demand for sustainable production of biochemicals that substitute for chemicals otherwise obtained from fossil fuels.

Filamentous fungi are of great interests as biocatalysts in biorefineries as they naturally produce and secrete a variety of different organic acids that can be used as building blocks in the chemical industry; ideally, in a lignocellulosic biorefineries, the fungus could be considered in a combined approach where it hydrolyses plant biomass wastes and ferments the resulting sugars into different organic acids.

Genomics and metabolomics analyses enable rapid identification of novel secondary metabolites open to industrial exploitation through the design of high yielding fungal cell factories (Karagiosis & Baker, 2012; Khan *et al.*, 2014; Nielsen & Nielsen, 2017; Badalyan & Zambonelli, 2019; Badalyan *et al.*, 2019). There is no shortage of novel methods to obtain new metabolites by engineering fungal secondary metabolism, but increased *yield* is the key essential and regulation of secondary metabolite biosynthesis is incompletely understood.

However, the identification of the *mcrA* gene as a principal regulator of *Aspergillus* secondary metabolism indicates that further advance in this direction is imminent. The *mcrA* gene is conserved, and it encodes a transcription factor that regulates transcription of hundreds of genes including at least ten secondary metabolite gene clusters in *Aspergillus terreus* and *Penicillium canescens* (Scharf & Brakhage, 2013; Oakley *et al.*, 2016).

Production of recombinant proteins by filamentous fungi was initially focussed on exploiting the extraordinary enzyme synthesis and secretion ability of fungi to produce single recombinant protein products, especially by industrial strains of *Aspergillus*, *Trichoderma*, *Penicillium* and *Rhizopus* species. Two disadvantages of filamentous fungi as hosts for recombinant protein production became apparent immediately: one is their common ability to produce homologous proteases which could degrade the heterologous protein product and the other is that the protein glycosylation patterns in filamentous fungi and in mammals are quite different.

Specifically, fungi lack the functionally important terminal sialylation of the glycans that occurs in mammalian cells.

So, without engineering, filamentous fungi, despite their other advantages, are not the most suitable microbial hosts for production of recombinant human glycoproteins for therapeutic use. Nevertheless, strategies to prevent proteolysis have already met with some success and new scientific information being generated through genomics and proteomics research will extend the biomanufacturing capabilities of recombinant filamentous fungi, enabling them to express genes encoding multiple proteins, making filamentous fungi even better candidates to produce proteins and protein complexes for therapeutic use (Ward, 2012; Fernández & Vega, 2013; Nevalainen & Peterson, 2014).

Most of what we have discussed so far in this Section has implied industrial production by submerged (liquid) fermentation of fungi, but it is essential to remember that solid state fermentation is a crucial process for producing enzymes, organic acids, flavour compounds, pharmaceutical agents and food processing; see review by Ghosh (2016). Of course, it is also the foundation of the mushroom cultivation industry (Petre, 2015). This last is especially important in relation to potential improvements in the biotechnological procedures for producing mushrooms as healthy and highly nutritive food in their own right, while at the same time using mushroom farming as a bioremediation tool, by using recalcitrant wastes as substrates for the mushroom farming compost before crop production and/or by using spent mushroom compost for soil remediation after cropping (Purnomo *et al.*, 2011; Camacho-Morales & Sánchez, 2015).

Ganoderma is a particularly interesting edible commercial mushroom because it is mainly farmed for use as a traditional Chinese medicine. Fruit bodies of the *Ganoderma lucidum* species complex contain many bioactive compounds; indeed, well over 400 secondary metabolites have been isolated from various *Ganoderma* species (Baby *et al.*, 2015; Ahmad, 2019). A mixture of *Ganoderma lucidum* polysaccharides (known as GLP) is the main bioactive component in the water soluble extracts of this mushroom, and there is some evidence that GLP possesses potential anticancer activity (Sohretoglu & Huang, 2018).

A wide range of pharmaceutically-interesting metabolites have been found in extracts of *Ganoderma*, and some have been found to be stimulators of neural stem cell proliferation *in vitro*, which could be of value in treatment of neurodegenerative diseases (Yan *et al.*, 2015), while others have promise for therapeutic modification of the immune system (Berovic & Podgornik, 2015; Sudheer *et al.*, 2018).

Other extracts have been assessed for genotoxicity and anti-genotoxicity using comet assays of mouse lymphocytes; no evidence was found for genotoxic chromosomal breakage nor cytotoxic effects by *Ganoderma* extract in the mouse, nor did it protect against the effects of the mutagen ethyl methanesulfonate. This study found no evidence for the extract having any value in protecting against the test mutagen (Chiu *et al.* 2000). A more recent study found that although the aqueous extract of *Ganoderma lucidum* exhibited no genotoxic effect, it did have an antigenotoxic effect. This study used the hen's egg test for micronucleus induction as a genotoxicity assay (formation of micronuclei during cell division indicates induced chromosome

instability and fragmentation), which is different from the assay techniques using mouse lymphocytes mentioned earlier (Çelik & Özparlak, 2019).

Shah (2012) stresses the importance of genotoxicity testing for pharmaceuticals to ensure compliance with the guideline of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH; a unique project that brings together regulatory authorities of Europe, Japan and the United States with pharmaceutical industry representatives).

Most of the above deals with aspects of fungal biology that are fairly widely known; but there are some ***fungus-specific topics*** about which we remain astonishingly ignorant, but which could provide therapeutic targets (Moore *et al.*, 2020):

- How do some fungi and fungal-like organisms accomplish the three-dimensional positioning of wall- and membrane-forming vesicles to subdivide large volumes of cytoplasm to create motile or non-motile spores ('free cell formation')? They are the ***only*** organisms that do this.
- What mechanisms are used to ensure that the nuclear membrane remains intact as the nuclear division progresses? This is another ***unique*** characteristic of present day fungi.
- How does the Spitzenkörper operates? How it is assembled, regulated, and directed? It is ***only*** found in filamentous fungi.
- What controls the multinucleated nature of most hyphae and some specific tissues? How do fungi (and ***only*** fungi) control the synchronicity of mitotic divisions and then the (rapid) migration and distribution of daughter nuclei within their hyphae?
- How is hyphal fusion is managed, and how do incompatible fusions trigger the death of hyphae?
- What regulates the placement of septa in hyphae and how are different states of differentiation controlled on the two sides of perforated septa?
- How is the yeast-hyphal dimorphism controlled? It is important in the life-style of so many pathogenic fungi and control may be analogous to multicellular growth regulation in plants (Cogo *et al.*, 2018).
- How is hyphal branching controlled in time and space?
- How is the behaviour of multitudes of independent hyphal apices orchestrated to create tissue layers (like hymenia) and how are determinate-growth controls selectively applied to such large hyphal communities?

You will notice that we have not mentioned the fungal plasma membrane (which is ***unique*** in using ergosterol) or the fungal wall (which is also ***exclusive*** to fungi); this is because, as well as being established targets for selective toxicity, these structures have been investigated exhaustively. It's too easy to try a slightly different agent that targets the wall or membrane; another batch of chemicals, another lot of routine tests. We need now to be more imaginative when trying to identify new antifungal drugs and fungicides; and we have the tools to achieve this goal.

We would have more success if we directed our attention to some of the other structures and/or processes that are ***unique to fungi***. Bearing in mind the bulleted list immediately above we would suggest: (a) vesicle transportation and positioning; (b) nuclear membrane dynamics;

(c) Spitzenkörper dynamics; (d) nucleus migration and control of nuclear division; (e) hyphal fusion and how programmed cell death is triggered by the non-self-recognition system; (f) septation dynamics and physiological function; (g) control of dimorphism; (h) dynamics of hyphal branching; (i) selective control of determinate growth of hyphal apices.

9. FUNGI FOR BIOREMEDIATION

Petroleum-degrading bacteria can be used for bioremediation of soil contaminated with petroleum sludge, which is the oily sludge generated in refineries when inlet and outlet tanks are cleaned. Disposal of the sludge causes environmental issues as well as human health concerns, but bacterial remediation of contaminated soil can restore its ability to support healthy growth of crop plants (Varma *et al.*, 2017). Successful, and rapid, degradation rates of Total Petroleum Hydrocarbon (TPH) has also been demonstrated by mixed populations of petroleum-degrading bacteria with white-rot fungi grown in solid-state fermentation. Enzymes secreted by the white-rot fungus (laccases), acting together with the bacteria, degraded the petroleum hydrocarbons in the contaminated soil (Liu *et al.*, 2017), and this includes non-petroleum oil 'spills/pollutants' such as palm oil mill effluents (Subowo, 2019).

Petroleum oil spill is no longer the most dramatically-damaging form of pollution of the natural environment due to human activity because plastics have now reached the number one spot in global concern. Plastics have become an essential part of modern life; recent estimates suggest that in 2016, world plastics production totalled around 335 million metric tons. The phrase '*plastic materials*' covers a range of polymers, including polyvinylchloride (PVC), polyurethanes, polystyrene, polyamides and polyesters with a range of properties and susceptibility to degradation (Sabev *et al.*, 2006; Shah *et al.*, 2008).

The key to the potential for fungi being able to bioremediate plastic wastes is the capacity of basidiomycetes to degrade the complex structure of lignocellulose, which is due to their ability to secrete all the extracellular enzyme systems necessary to degrade wood. Ability to mineralise lignin is limited to white-rot fungi. Lignin is a polymer of three **phenyl**propanoid alcohols and benzene rings, ether linkages and carbon-carbon bonding predominate in its structure. Lignin is cleaved by an oxidative process that depends on two major groups of enzymes called heme peroxidases and laccases. Together, these enzymes digest away the lignin, which provides the main pigmentation of wood, leaving the white cellulose component (which is why they are called 'white-rots') for the cellulase enzymes to digest. The process of catabolic lignin degradation involves:

- cleavage of ether bonds between monomers;
- oxidative cleavage of the propane side chain;
- demethylation;
- benzene ring cleavage to ketoadipic acid which is fed into the tricarboxylic acid cycle as a fatty acid (Moore *et al.*, 2020).

The enzymes that achieve this have a high potential for biotechnological applications, especially mycoremediation (Barh *et al.*, 2019). Recent genomic studies of basidiomycetes have provided valuable information about the variety of enzymes they make available (Peralta *et al.*, 2017); we should all hope that '*synthetic biology*' (Osbourn *et al.*, 2012) can fulfil its promise and be used to re-design existing biological systems for this essential

mycoremediation function. Thankfully, the importance of fungi and of mycology for continued world development and improvement is beginning to be realised (Lange *et al.*, 2012).

Generally low production costs have led to plastics being used in a vast range of applications; as you can see from a casual glance around your surroundings, from flooring materials, heat insulation, shoe soles, cable sheaths, pipework, packaging, food containers, furnishings, electronic devices and a host of other products that have become essential to modern life. It has been estimated that a total of 8,300 million metric tons of plastics have been produced in the world during the past 65 years; and that, as of 2015, approximately 6,300 million metric tons of plastic waste had been generated, around 9% of which had been recycled, 12% was incinerated, and 79% was accumulated in landfills or just discarded into the natural environment. Between 5 million and 13 million metric tons of plastic end up in the ocean every year (Geyer *et al.*, 2017). The use of plastics has increased twenty fold since 1964, and it is expected to double by 2035 (Velis, 2014). It has been reported that the world's synthetic plastic production will be approximately 1800 million tons in 2050 (Gallo *et al.*, 2018) and approximately 12,000 million metric tons of plastic waste will be in the environment by that year (Geyer *et al.*, 2017; <https://www.nature.com/articles/s41467-018-04565-2.pdf>).

Poly-(ethylene terephthalate) (PET) is one of the most abundantly produced synthetic polymers and is accumulating in the environment at a staggering rate as discarded packaging and textiles. Unfortunately, the properties that make PET so useful to us in our daily lives also endow it with an alarming resistance to biodegradation, with the potential of it lasting for centuries in most natural environments. Most applications that employ PET, such as single-use beverage bottles, clothing, packaging, and carpeting employ crystalline PET, which is recalcitrant to catalytic or biological depolymerisation due to the limited accessibility of the ester linkages.

PET can be depolymerised to its constituents if the ester bonds of the polymer can be cleaved. Doing this with available chemical techniques is too costly to be a viable recycling solution. Recently, a newly discovered bacterium isolated from outside a bottle-recycling facility in Japan, *Ideonella sakaiensis*, was shown to exhibit the rare ability to grow on PET as a major carbon and energy source. When grown on PET, this strain produces two enzymes capable of hydrolysing PET and the reaction intermediate, mono(2-hydroxyethyl) terephthalic acid. Both enzymes are required to enzymatically convert PET efficiently into its two environmentally benign monomers, terephthalic acid and ethylene glycol; so, yielding the monomers for further plastics manufacture (Yoshida *et al.*, 2016; Austin *et al.*, 2018).

As long ago as 2003, Barratt *et al.* (2003) demonstrated that fungi are the predominant microorganisms responsible for degradation of soil-buried polyester polyurethane. More recently, Álvarez-Barragán *et al.* (2016) found several fungi able to use polyester polyurethane or a polyether polyurethane varnish as the only carbon source and found that the fungi were also able to degrade solid polyester polyurethane foams. FTIR-spectroscopy and GC-MS was used to show the hydrolysis of ester and urethane bonds in the polyurethane substrates. Khan *et al.* (2017) isolated a strain of *Aspergillus tubingensis* that was capable of degrading polyurethane from the soil of a general city waste disposal site in Islamabad, Pakistan. Between them, these studies have isolated strains from 12 different species of fungi that can significantly degrade polyurethanes, demonstrating that there is no difficulty in isolating efficient polyurethane-

degrading fungi from nature, and that the mechanisms they use to degrade the polymer are accessible to study and could provide the basis for the development of biotechnological processes for polyurethane biodegradation and recycling.

A different approach is to study the degradation rates between different polymers to find biodegradable polymers that have the potential to be compostable. Al Hosni *et al.* (2019) studied biodegradation of the four polymers polycaprolactone (PCL), polyhydroxybutyrate (PHB), polylactic acid (PLA) and poly(1,4 butylene) succinate (PBS) in soil and compost over ten months at 25°C, 37°C and 50°C. PCL showed the fastest degradation rate under all conditions and was completely degraded after 91 days when buried in compost and incubated at 50 °C. Fungi growing on the polymer surfaces were identified by sequence analysis. *Aspergillus fumigatus* was the most commonly found at 25°C and 37°C, while *Thermomyces lanuginosus* was abundant at 50 °C, and could degrade PCL over a range of soil conditions.

For some recalcitrant plastics, such as PVC, although the polymer itself is highly resistant to degradation, it is the plasticisers (organic acids blended into the material to increase flexibility of the product) that are often themselves highly susceptible to enzymatic microbial attack; as in, for example plasticised PVC. Consequently, broad spectrum biocides are often incorporated into polymer blends to inhibit fungal and bacterial growth and so extend the lifetime of the final product; which, of course, only increases the adverse environmental impact when the plastic is discarded.

Ahuactzin-Pérez *et al.* (2018) discovered that *Pleurotus ostreatus* degrades and uses (as carbon and energy source) high concentrations of di-(2-ethyl hexyl) phthalate (DEHP), and *Fusarium culmorum* has been shown to produce a range of esterase enzymes when challenged with DEHP (Ferrer-Parra *et al.* 2018; Portillo-Ojeda *et al.*, 2018; Ocaña-Romo *et al.*, 2018; González-Márquez *et al.*, 2019). Phthalates are plasticisers, primarily used as additives in plastics like polyvinyl chloride (PVC) polymers to make them more flexible. Because they are not chemically bound, phthalates are easily released from plastic articles, through direct release, leaching, and abrasion, and phthalate esters are one of the most frequently detected **persistent** organic pollutants in the environment (Gao & Wen, 2015). In laboratory animal studies, some phthalates have been associated with developmental and reproductive toxicity and they are generally considered to be toxins that interfere with endocrine systems in mammals (Hauser & Calafat, 2005).

If common ascomycete and basidiomycete fungi can produce enzyme systems enabling them to use such pollutants, and the solid polymers themselves, for growth, they provide an opportunity for bioremediation of plastic waste. *Pleurotus ostreatus* degrades lignin efficiently, grows well in both liquid and solid fermentation systems, and is an ideal candidate for genome engineering into a **plastic-eating Oyster mushroom**.

DEDICATION

We must end on an unwelcome sad note. As the text of this review was reaching completion our good friend, greatly valued colleague and co-author Geoffrey David Robson died suddenly on 15th May 2018. We dedicate this paper to his memory as a mycologist of distinction.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

Aebersold R. 2003. Quantitative proteome analysis: methods and applications. *Journal of Infectious Diseases*. 187: S315-S320.

Ahmad M.F. 2019. *Ganoderma lucidum*: a macro fungus with phytochemicals and their pharmacological properties. In: *Plant and Human Health*, Vol 2 (eds Ozturk M. & Hakeem K.), pp. 491-515. Springer Nature. Cham, Switzerland. ISBN: 9783030033439.

Ahuactzin-Pérez M., Tlecuil-Beristain S., García-Dávila J., Santacruz-Juárez E., González-Pérez M., Gutiérrez-Ruiz M.C. & Sánchez C. 2018. A novel biodegradation pathway of the endocrine-disruptor di(2-ethyl hexyl) phthalate by *Pleurotus ostreatus* based on quantum chemical investigation. *Ecotoxicology and Environmental Safety*. 147: 494-499.

Albertin W. & Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. *Proceedings of the Royal Society B: Biological Sciences*. 279: 2497-2509.

Al Hosni A.S., Pittman J.K. & Robson G.D. 2019. Microbial degradation of four biodegradable polymers in soil and compost demonstrating polycaprolactone as an ideal compostable plastic. *Waste Management*. 97: 105-114.

Álvarez-Barragán J., Domínguez-Malfavón L., Vargas-Suárez M., González-Hernández R., Aguilar-Osorio G. & Loza-Tavera H. 2016. Biodegradative activities of selected environmental fungi on a polyester polyurethane varnish and polyether polyurethane foams. *Applied and Environmental Microbiology*. 82(17): 5225-5235.

Anonymous Editorial. 2011. Method of the Year 2011. *Nature Methods*. 9: 1.

Anonymous Editorial. 2017. Stop neglecting fungi. *Nature Microbiology*. 2: 17120.

Anzalone A.V., Randolph P.B., Davis J.R., Sousa A.A., Koblan L.W., Levy J.M., Chen P.J., Wilson C., Newby G.A., Raguram A. & Liu, D.R. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. URL: <https://www.nature.com/articles/s41586-019-1711-4>.

Aurrecochea C., Barreto A., Basenko E.Y., Brestelli J., Brunk B.P. and 30 others. 2017. EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Research*. 45: D581–D591.

Austin H.P., Allen M.D., Donohoe B.S., Rorrer N.A., Kearns F.L. and 16 others. 2018. Characterisation and engineering of a plastic-degrading aromatic polyestherase. *Proceedings of the National Academy of Sciences of the United States of America*. 115: E4350-E4357.

Baby S., Johnson A.J. & Govindan B. (2015). Secondary metabolites from *Ganoderma*. *Phytochemistry*. 114: 66-101.

Badalyan S.M., Barkhudaryan A. & Rapior S. 2019. Recent progress in research on the pharmacological potential of mushrooms and prospects for their clinical application. In: *Medicinal Mushrooms: Recent Progress in Research and Development*, (eds Agrawal, D. & Dhanasekaran, M.). Chapter 1, pp. 1-70. Springer Nature Singapore Pte Ltd. Singapore.

Badalyan S.M. & Zambonelli A. 2019. Biotechnological exploitation of macrofungi for the production of food, pharmaceuticals and cosmeceuticals. In: *Advances in Macrofungi: Diversity, Ecology and Biotechnology*, (eds Sridhar K.R. & Deshmukh S.K.), pp. 199-230. CRC Press, an imprint of Taylor & Francis Group. Boca Raton, FL, USA: 366 pp.

Baker S.E., Thykaer J., Adney W.S., Brettin T.S., Brockman F.J., d'Haeseleer P., Martinez A.D., Miller R.M., Rokhsar D.S., Schadt C.W., Torok T., Tuskan G., Bennett J., Berka R.M., Briggs S.P., Heitman J., Taylor J., Turgeon B.G., Werner-Washburne M. & Himmel M.E. 2008. Fungal genome sequencing and bioenergy. *Fungal Biology Reviews*. 22: 1-5.

Barh A., Kumari B., Sharma S., Annepu S.K., Kumar A., Kamal S. & Sharma V.P. 2019. Mushroom mycoremediation: kinetics and mechanism. Chapter 1 (pp. 1-22) in: *Smart Bioremediation Technologies: Microbial Enzymes*, (ed Bhatt P.). Pp. 408. Academic Press an imprint of Elsevier Inc. London, UK. ISBN: 9780128183076.

Barratt S.R., Ennos A.R., Greenhalgh M., Robson G.D. & Handley P.S. 2003. Fungi are the predominant micro-organisms responsible for degradation of soil-buried polyester polyurethane over a range of soil water holding capacities. *Journal of Applied Microbiology*. 95(1): 78-85.

Berovic M. & Podgornik B.B. 2015. Cultivation of medicinal fungi in bioreactors. In: *Mushroom Biotechnology: Developments and Applications*, (ed Petre M.), pp. 155-171. Academic Press, an imprint of Elsevier Inc. London, UK. 242 pp. ISBN: 9780128027943.

Bhadauria V., Zhao W.-S., Wang L.-X., Zhang Y., Liu J.-H., Yang J., Kong L.-A. & Peng Y.-L. 2007. Advances in fungal proteomics. *Microbiological Research*. 162: 193-200.

Bunnik E.M. & Le Roch K.G. 2013. An introduction to functional genomics and systems biology. *Advances in Wound Care*. 2: 490-498.

Camacho-Morales R.L. & Sánchez J.E. 2015. Biotechnological use of fungi for the degradation of recalcitrant agro-pesticides. In: *Mushroom Biotechnology: Developments and Applications*, (ed Petre M.), pp. 203-214. Academic Press, an imprint of Elsevier Inc. London. 242 pp.

Caracuel-Rios Z. & Talbot N.J. 2008. Silencing the crowd: high-throughput functional genomics in *Magnaporthe oryzae*. *Molecular Microbiology*. 68: 1341-1344.

Castanera R., López-Varas L., Borgognone A. LaButti K., Lapidus A., Schmutz J., Grimwood J., Pérez G., Pisabarro A.G., Grigoriev I.V., Stajich J.E. & Ramírez L. 2016. Transposable

elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics* 12: article e1006108.

Ceccaldi R., Rondinelli B. & D'Andrea A.D. 2016. Repair pathway choices and consequences at the double-strand break. *Trends in Cell Biology*. 26: 52-64.

Çelik B. & Özparlak H. 2019. Determination of genotoxic and antigenotoxic effects of wild-grown Reishi mushroom (*Ganoderma lucidum*) using the hen's egg test for analysis of micronucleus induction. *Biotechnic & Histochemistry*, article number: 1052-0295.

Chen J., Lai Y., Wang L., Zhai S., Zou G., Zhou Z., Cui C. & Wang S. 2017. CRISPR/Cas9-mediated efficient genome editing via blastospore-based transformation in entomopathogenic fungus *Beauveria bassiana*. *Scientific Reports*. 7: article 45763.

Chiu S.W., Wang Z.M., Leung T.M. & Moore D. 2000. Nutritional value of *Ganoderma* extract and assessment of its genotoxicity and anti-genotoxicity using comet assays of mouse lymphocytes. *Food and Chemical Toxicology*. 38: 173-178.

Cho Y., Davis J.W., Kim K.-H., Wang J., Sun Q.-H., Cramer R.A. Jr, Lawrence C.B. 2006. A high throughput targeted gene disruption method for *Alternaria brassicicola* functional genomics using linear minimal element (LME) constructs. *Molecular Plant-Microbe Interactions*. 19: 7-15.

Cogo A.J.D., Dutra Ferreira K. dos R., Okorokov L.A., Ramos A.C., Façanha A.R. & Okorokova-Façanha A.L. 2018. Spermine modulates fungal morphogenesis and activates plasma membrane H⁺-ATPase during yeast to hyphae transition. *Biology Open*. 7: article bio029660.

Delneri D., Brancia F.L. & Oliver S.G. 2001. Towards a truly integrative biology through the functional genomics of yeast. *Current Opinion in Biotechnology*. 12: 87-91.

Doudna J.A. & Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 346: article 1258096.

Dunn D.A. & Pinkert C.A. 2014. Gene editing. In: *Transgenic Animal Technology. A Laboratory Handbook* (3rd edition), (ed Pinkert C.A.), pp. 229-248. Elsevier Inc. London, UK. ISBN 9780124104907.

Dyer P.S., Munro C.A. & Bradshaw R.E. 2017. Fungal genetics. In: *Oxford Textbook of Medical Mycology*. Eds: Kibbler C.C., Barton R., Gow N.A.R., Howell S., MacCallum D.M. & Manuel R.J. Chapter 5, pp. 35-42. Oxford University Press Oxford, UK: 400 pp.

Fernández F.J. & Vega M.C. 2013. Technologies to keep an eye on: alternative hosts for protein production in structural biology. *Current Opinion in Structural Biology*. 23: 365-373.

Ferrer-Parra L., López-Nicolás D.I., Martínez-Castillo R., Montiel-Cina J.P., Morales-Hernández A.R., Ocaña-Romo E., González Márquez A., Portillo-Ojeda M., Sánchez-Sánchez D.F. & Sánchez C. 2018. Partial characterization of esterases from *Fusarium culmorum* grown

in media supplemented with di (2-ethyl hexyl phthalate) in solid-state and submerged fermentation. *Mexican Journal of Biotechnology*. 3: 82-94.

Foster S.J., Monahan B.J., Bradshaw R.E. 2006. Genomics of the filamentous fungi - moving from the shadow of the bakers yeast. *Mycologist*. 20: 10-14.

Gao D. & Wen Z.-D. 2015. Phthalate esters in the environment: a critical review of their occurrence, biodegradation, and removal during wastewater treatment processes. *The Science Of The Total Environment*. 541: 986-1001.

Gallo F., Fossi C., Weber R., Santillo D., Sousa J., Ingram I., Nadal A. & Romano D. 2018. Marine litter plastics and microplastics and their toxic chemicals components: the need for urgent preventive measures. *Environmental Sciences Europe*. 30: 13-27.

Gehrmann T., Pelkmans J.F., Lugones L.G., Wösten H.A.B., Abeel T. & Reinders M.J.T. 2016. *Schizophyllum commune* has an extensive and functional alternative splicing repertoire. *Scientific Reports*. 6: article 33640.

Geyer R., Jambeck J.R. & Law K.L. 2017. Production, use, and fate of all plastics ever made. *Science Advances*. 3: article number e1700782.

Ghosh J.S. 2016. Solid state fermentation and food processing: a short review. *Journal of Nutrition & Food Sciences*. 6: 453 (7 pages).

Gibson G. & Muse S. 2009. *A Primer of Genome Science*, 3rd Edn. Sinauer Associates, Inc./Macmillan Publishers Limited. Basingstoke, UK: Pp.350.

Gladyshev E. 2017. Repeat-Induced Point mutation (RIP) and other genome defense mechanisms in fungi. *Microbiology Spectrum*. 5: FUNK-0042-2017.

González-Márquez A., Loera-Corral O., Santacruz-Juárez J, Tlécuítl-Beristain S., García-Dávila J., Viniestra-González G. & Sánchez C. 2019. Biodegradation patterns of the endocrine disrupting pollutant di (2-ethyl hexyl) phthalate by *Fusarium culmorum*. *Ecotoxicology and Environmental Safety*. 170: 293-299.

Grigoriev I.V., Nikitin R., Haridas S., Kuo A., Ohm R., Otilar R., Riley R., Salamov A., Zhao X., Korzeniewski F., Smirnova T., Nordberg H., Dubchak I. & Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. 42: D699-D704.

Grützmann K., Szafranski K., Pohl M., Voigt K., Petzold A. & Schuster S. 2014. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA Research*. 21: 27-39.

Hauser R. & Calafat A.M. 2005. Phthalates and human health. *Occupational and Environmental Medicine*. 62: 806-818.

Hibbett D.S., Stajich J.E. & Spatafora J.W. (2013). Toward genome-enabled mycology. *Mycologia*. 105: 1339-1349.

Honda S. & Selker E.U. 2009. Tools for fungal proteomics: multifunctional *Neurospora* vectors for gene replacement, protein expression and protein purification. *Genetics*. 182: 11-23.

Horgan R.P. & Kenny L.C. 2011. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*. 13: 189-195.

Idnurm A., Bailey A.M., Cairns T.C., Elliott C.E., Foster G.D., Ianiri G. & Jeon J. 2017. A silver bullet in a golden age of functional genomics: the impact of *Agrobacterium*-mediated transformation of fungi. *Fungal Biology and Biotechnology*. 4: 6.

Irimia M. & Roy S.W. 2014. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harbor Perspectives in Biology*. 6: article a016071.

Jewett M.C., Hofmann G. & Nielsen J. 2006. Fungal metabolite analysis in genomics and phenomics. *Current Opinion in Biotechnology*. 17: 191-197.

Jin L., Li G., Yu D., Huang W., Cheng C., Liao S., Wu Q. & Zhang, Y. (2017). Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus *Verticillium dahlia*. *BMC Genomics*. 18: 130.

Jones M.G. 2007. The first filamentous fungal genome sequences: *Aspergillus* leads the way for essential everyday resources or dusty museum specimens? *Microbiology*. 153: 1-6.

Karagiosis S.A. & Baker S.E. 2012. Fungal Cell Factories. In: *Food and Industrial Bioproducts and Bioprocessing*, (ed Dunford N.T.). Wiley-Blackwell. Oxford, UK:

Kaznessis Y.N. 2007. Models for synthetic biology. *BMC Systems Biology*. 1: 47.

Kempken F. & Kück U. 1998. Transposons in filamentous fungi - facts and perspectives. *BioEssays*. 20: 652-659.

Khan A.A., Bacha N., Ahmad B., Lutfullah G., Farooq U. & Cox R.J. 2014. Fungi as chemical industries and genetic engineering for the production of biologically active secondary metabolites. *Asian Pacific Journal of Tropical Biomedicine*. 4: 859-870.

Khan S., Nadir S., Shah Z.U., Shah A.A., Karunarathna S.C., Xu J., Khan A., Munir S. & Hasan F. 2017. Biodegradation of polyester polyurethane by *Aspergillus tubingensis*. *Environmental Pollution*. 225: 469-480.

Klipp E., Liebermeister W., Wierling C., Kowald A., Lehrach H. & Herwig R. 2009. *Systems Biology: A Textbook*. Wiley-VCH Verlag GmbH & Co. KgaA. Weinheim, Germany. ISBN-10: 3527318747, ISBN-13: 978-3527318742.

Lange L., Bech L., Busk P.K., Grell M.N., Huang Y., Lange M., Linde T., Pilgaard B., Roth D. & Tong, X. 2012. The importance of fungi and of mycology for a global development of the bioeconomy. *IMA Fungus*. 3: 87-92.

Lewin H.A., Robinson G.E., Kress W.J., Baker W.J., Coddington J. and 19 others. 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*. 115: 4325-4333.

Lipshutz R.J., Fodor S.P., Gingeras T.R. & Lockhart D.J. 1999. High density synthetic oligonucleotide arrays. *Nature Genetics*. 21 (January supplement): 20-24.

Liu B., Liu J., Ju M., Lic X. & Wang P. 2017. Bacteria-white-rot fungi joint remediation of petroleum-contaminated soil based on sustained-release of laccase. *RSC Advances*. 7: 39075-39081.

McCluskey K. & Baker S.E. 2017. Diverse data supports the transition of filamentous fungal model organisms into the post-genomics era. *Mycology*. 8: 67-83.

McNutt M. 2015. Editorial: breakthrough to genome editing. *Science*. 350: 1445.

Michielse C.B., Hooykaas P.J.J., van den Hondel C.A.M.J.J. & Ram A.F.J. 2005. *Agrobacterium*-mediated transformation as a tool for functional genomics in fungi. *Current Genetics*. 48: 1-17.

Mishra N.C. & Tatum E.L. 1973. Non-Mendelian inheritance of DNA-induced inositol independence in *Neurospora*. *Proceedings of the National Academy of Sciences of the United States of America*. 70: 3875-3879.

Mohanta T.K., Bashir T., Hashem A., Abd-Allah E.F. & Bae H. 2017. Genome editing tools in plants. *Genes*. 8: 399.

Moore D. & Novak Frazer L. 2002. *Essential Fungal Genetics*. Springer-Verlag Inc. New York.

Moore D., Robson G.D. & Trinci A.P.J. 2020. *21st Century Guidebook to Fungi*, second edition. Cambridge University Press. London, UK. pp 600. ISBN: 9781108745680; www.cambridge.org/9781108745680.

Muszewska A., Steczkiewicz K., Stepniewska-Dziubinska M. & Ginalski K. 2017. Cut-and-paste transposons in fungi with diverse lifestyles. *Genome Biology and Evolution*. 9: 3463-3477.

Nagasaki M., Saito A., Doi A., Matsuno H. & Miyano S. 2009. *Foundations of Systems Biology*. Springer-Verlag. London, UK. ISBN-10: 1848820224, ISBN-13: 978-1848820227.

Nevalainen H. & Peterson R. 2014. Making recombinant proteins in filamentous fungi - are we expecting too much? *Frontiers in Microbiology*. 5: 75.

Nielsen J.C. & Nielsen J. 2017. Development of fungal cell factories for the production of secondary metabolites: linking genomics and metabolism. *Synthetic and Systems Biotechnology*. 2: 5-12.

Nørdvig C.S., Nielsen J.B., Kogle M.E. & Mortensen U.H. 2015. A CRISPR-Cas9 system for genetic engineering of filamentous fungi. PLoS ONE. 10: article e0133085.

Nowrousian M. 2007. Of patterns and pathways: microarray technologies for the analysis of filamentous fungi. Fungal Biology Reviews. 21: 171-178.

Nowrousian M. 2014. Genomics and transcriptomics to analyze fruiting body development. In: The Mycota, Fungal Genomics, XIII: (2nd ed.) (ed Nowrousian M.), pp. 149-172. Springer-Verlag: Berlin, Germany. ISBN: 978-3-642-45217-8.

Oakley C.E., Ahuja M., Sun W.-W., Entwistle R., Akashi T., Yaegashi J., Guo C.-J., Cerqueira G.C., Russo W.J., Wang C.C.C., Chiang Y.-M. & Oakley B.R. 2016. Discovery of *McrA*, a master regulator of *Aspergillus* secondary metabolism. Molecular Microbiology. 103: 347-365.

Ocaña-Romo E., Ferrer-Parra L., López-Nicolás D.I., Martínez-Castillo R., Montiel-Cina J.P., Morales-Hernández A.R., González-Márquez A., Portillo-Ojeda M.L., Sánchez-Sánchez D.F. & Sánchez C. 2018. Partial characterization of esterases from *Fusarium culmorum* grown in media containing di (2-ethyl hexyl phthalate) in solid-state and submerged fermentation. New Biotechnology. 44: Abstracts Supplement, p. S137, abstract P25-11.

Osborn A.E., O'Maille P.E., Rosser S.J. & Lindsey K. 2012. Synthetic biology. New Phytologist. 196: 671-677.

Parnell L.D., Lindenbaum P., Shameer K., Dall'Olio G.M., Swan D.C., Jensen L.J., Cockell S.J., Pedersen B.S., Mangan M.E., Miller C.A. & Albert I. 2011. BioStar: an online question & answer resource for the bioinformatics community. PLoS Computational Biology. 7: article e1002216.

Pastore A. & Puccio H. 2013. Frataxin: a protein in search for a function. Journal of Neurochemistry. 126: 43-52.

Peralta R.M., da Silva B.P., Gomes Côrrea R.C., Kato C.G., Vicente Seixas F.A. & Bracht A. 2017. Enzymes from basidiomycetes - peculiar and efficient tools for biotechnology. In: Biotechnology of Microbial Enzymes: Production, Biocatalysis and Industrial Applications, (eds Brahmachari G., Demain A.L. & Adrio J.L.), pp. 119-149. Academic Press. An imprint of Elsevier Inc. Amsterdam, The Netherlands. ISBN: 9780128037256.

Peter J., De Chiara M., Friedrich A., Yue J.-X., Pflieger D. and 16 others. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature. 556: 339-344.

Petre M. (ed). 2015. Mushroom Biotechnology: Developments and Applications. Academic Press, an imprint of Elsevier Inc. London, UK. 242 pp. ISBN: 9780128027943.

Phasha M.M., Wingfield B.D., Coetzee M.P.A., Santana Q.C., Fourie G. & Steenkamp E.T. 2017. Architecture and distribution of introns in core genes of four *Fusarium* species. G3: Genes, Genomes, Genetics. 7: 3809-3820.

Portillo-Ojeda M.L., Arteaga-Mejía M., González-Márquez A. & Sánchez C. (2018). Effect of the pH on growth and esterase activity of *Fusarium culmorum* grown on media supplemented with di (2-ethylhexyl) phthalate in submerged fermentation. *New Biotechnology*. 44: Abstracts Supplement p. S138, abstract P25-12.

Pritham E.J. 2009. Transposable elements and factors influencing their success in eukaryotes. *Journal of Heredity*, 100: 648-655.

Pudake R.N., Kumari M., Sahu B.B. & Sultan E. 2017. Targeted gene disruption tools for fungal genomics. In: *Modern Tools and Techniques to Understand Microbes*, (eds Varma A. & Sharma A.), pp 81-102. Springer Inc. Cham, Switzerland. ISBN 9783319491950.

Purnomo A.S., Mori T., Takagi K. & Kondo R. 2011. Bioremediation of DDT contaminated soil using brown-rot fungi. *International Biodeterioration & Biodegradation*. 65: 691-695.

Ranatunga W., Gakh O., Galeano B.K., Smith D.Y., Söderberg C.A.G., Al-Karadaghi S., Thompson J.R. & Isaya G. 2016. Architecture of the yeast mitochondrial iron-sulfur cluster assembly machinery: the sub-complex formed by the iron donor, Yfh1 protein, and the scaffold, Isu1 protein. *Journal of Biological Chemistry*. 291: 10378-10398.

Richards T.A., Leonard G., Soanes D.M. & Talbot N.J. 2011. Gene transfer into the fungi. *Fungal Biology Reviews*, 25: 98-110.

Roberts S.E. & Gladfelter A.S. 2016. Nuclear dynamics and cell growth in fungi. In: *The Mycota, Vol. I. Growth, Differentiation and Sexuality*, 3rd edn, (ed Wendland J.), pp. 27-46. Springer International Publishing. Cham, Switzerland.

Rokas A. 2009. The effect of domestication on the fungal proteome. *Trends in Genetics*. 25: 60-63.

Roper M., Simonin A., Hickey P.C., Leeder A. & Glass N.L. 2013. Nuclear dynamics in a fungal chimera. *Proceedings of the National Academy of Sciences of the United States of America*. 110: 12875-12880.

Ross-Macdonald P., Coelho P.S., Roemer T., Agarwal S., Kumar A., Jansen R., Cheung K.H., Sheehan A., Symoniatis D., Umansky L., Heidtman M., Nelson F.K., Iwasaki H., Hager K., Gerstein M., Miller P., Roeder G.S. & Snyder M. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*. 402: 413-418.

Sabev H.A., Barratt S.R., Greenhalgh M., Handley P.S. & Robson G.D. 2006. Biodegradation and biodeterioration of man-made polymeric materials. In: *Fungi in Biogeochemical Cycles*, (ed Gadd G.M.), pp. 212-235. Cambridge University Press. Cambridge, UK. ISBN: 9780521845793.

Scharf D.H. & Brakhage A.A. 2013. Engineering fungal secondary metabolism: a roadmap to novel compounds. *Journal of Biotechnology*. 163: 179-183.

Shah A.A., Hasan F., Hameed A. & Ahmed S. 2008. Biological degradation of plastics: A comprehensive review. *Biotechnology Advances*. 26: 246-265.

Shah S.U. 2012. Importance of genotoxicity & S2A guidelines for. *IOSR Journal of Pharmacy and Biological Sciences*. 1: 43-54.

Sharma K.K. 2015. Fungal genome sequencing: basic biology to biotechnology. *Critical Reviews in Biotechnology*. 36: 743-759.

Sharman A. 2001. The many uses of a genome sequence. *Genome Biology*. 2: reports 4013.1-4013.4.

Shen X. & Zhu Z.-J. 2019. MetFlow: an interactive and integrated workflow for metabolomics data cleaning and differential metabolite discovery. *Bioinformatics*. 35: 2870-2872.

Silver P.A., Way J.C., Arnold F.H. & Meyerowitz J.T. 2014. Engineering explored. *Nature*. 509: 166-167.

Sims A.H., Gent M.E., Robson G.D., Dunn-Coleman N.S. & Oliver S.G. 2004. Combining transcriptome data with genomic and cDNA sequence alignments to make confident functional assignments for *Aspergillus nidulans* genes. *Mycological Research*. 108: 853-857.

Slot J.C., Townsend J.P. & Wang Z. 2017. Fungal gene cluster diversity and evolution. *Advances in Genetics*, **100**: 141-178.

Sohretoglu D. & Huang S. 2018. *Ganoderma lucidum* polysaccharides as an anti-cancer agent. *Anti-Cancer Agents in Medicinal Chemistry*. 18: 667-674.

Stajich J.E, Harris T., Brunk B.P., Brestelli J., Fischer S., Harb O.S., Kissinger J.C. Li, W. Nayak V., Pinney D.F., Stoeckert C.J. & Roos D.S. 2012. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Research*. 40: D675-D681.

Steenkamp E.T., Wingfield M.J., McTaggart A.R. & Wingfield B.D. 2018. Fungal species and their boundaries matter - definitions, mechanisms and practical implications. *Fungal Biology Reviews*, 32: 104-116.

Strom N.B. & Bushley K.E. 2016. Two genomes are better than one: history, genetics, and biotechnological applications of fungal heterokaryons. *Fungal Biology and Biotechnology*. 3: 4.

Stukenbrock E.H. & Croll D. 2014. The evolving fungal genome. *Fungal Biology Reviews*. 28: 1-12.

Subowo Y.B. 2019. Screening of basidiomycetes with laccase activity for lignin degradation on POME. *Institute of Physics (IOP) Conference Series: Earth and Environmental Science*. 308: conference 1, article number 012013.

Sudheer S., Alzorqi I., Manickam,S. & Ali A. (2018). Bioactive compounds of the wonder medicinal mushroom '*Ganoderma lucidum*'. In: *Bioactive Molecules in Food*. Reference Series

in Phytochemistry, (eds Mérillon J.M. & Ramawat K.), pp 1-31. Springer International Publishing AG. Cham, Switzerland. ISBN: 9783319545288.

Sugui J.A., Chang Y.C. & Kwon-Chung K.J. 2005. *Agrobacterium tumefaciens*-mediated transformation of *Aspergillus fumigatus*: an efficient tool for insertional mutagenesis and targeted gene disruption. Applied and Environmental Microbiology. 71: 1798-1802.

Taylor J.W., Branco S., Gao C., Hann-Soden C., Montoya L., Sylvain I. & Gladieux P. 2017. Sources of fungal genetic variation and associating it with phenotypic diversity. Microbiology Spectrum. 5: FUNK-0057-2016.

Todd R., Forche A. & Selmecki A. 2017. Ploidy variation in fungi: polyploidy, aneuploidy, and genome evolution. Microbiology Spectrum. 5: FUNK-0051-2016.

Varma S.S., Lakshmi M.B., Rajagopal P. & Velan M. (2017). Degradation of Total Petroleum Hydrocarbon (TPH) in contaminated soil using *Bacillus pumilus* MVSV3. Biocontrol Science. 22: 17-23.

Velis C. A. 2014. Global recycling markets - plastic waste: A story for one player – China. Report prepared by FUELogy and formatted by D-waste on behalf of International Solid Waste Association - Globalisation and Waste Management Task Force. ISWA, Vienna.

Ward O.P. 2012. Production of recombinant proteins by filamentous fungi. Biotechnology Advances. 30: 1119-1139.

Willis K.J. (ed) 2018. State of the World's Fungi 2018. Royal Botanic Gardens, Kew. London, UK. Report, 92 pp. URL: <https://stateoftheworldsfungi.org/2018/> (el PDF del informe es una descarga gratuita).

Winzeler E.A., Shoemaker D.D., Astromoff A., Liang H., Anderson K. and 46 others. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science. 285: 901-906.

Yandell M. & Ence D. (2012). A beginner's guide to eukaryotic genome annotation. Nature Reviews Genetics. 13: 329-342.

Yan Y.-M., Wang X.-L., Luo, Q., Jiang L.-P., Yang C.-P., Hou B., Zuo Z.-L., Chen Y.-B. & Cheng Y.-X. (2015). Metabolites from the mushroom *Ganoderma lingzhi* as stimulators of neural stem cell proliferation. Phytochemistry. 114: 155-162.

Yoshida S., Hiraga K., Takehana T., Taniguchi I., Yamaji H., Maeda Y., Toyohara K., Miyamoto K., Kimura Y. & Oda K. 2016. A bacterium that degrades and assimilates poly(ethylene terephthalate). Science. 351: 1196-1199.

Zheng Y.-M., Lin F.-L., Gao H., Zou G., Zhang J.-W., Wang G.-Q., Chen G.-D., Zhou Z.-H., Yao X.-S. & Hu D. 2017. Development of a versatile and conventional technique for gene disruption in filamentous fungi based on CRISPR-Cas9 technology. Scientific Reports. 7: 9250.