WHOLE-GENOME SEQUENCING OF THE CULTIVATED BUTTON MUSHROOM *AGARICUS BISPORUS*: HISTORY, STATUS, AND APPLICATIONS

RICHARD W. KERRIGAN
Sylvan Research
198 Nolte Drive
Kittanning, Pennsylvania 16201
United States of America
rwk@sylvaninc.com

[on behalf of the Agaricus *Genomics Community]*

ABSTRACT

The culmination of the first half-century of investigations into the genome of *Agaricus bisporus* was marked by the first whole-genome sequencing for a strain of this species. The genome of a second strain of *A. bisporus* was sequenced the following year. While a structural view of the genome is now effectively complete, functional interpretations of the many gene sequences, supported by accumulating experimental data, are still being developed. A first summary of the physiological capabilities and strategies of this 'humicolous' species is now in preparation; when completed it should illuminate the role of the fungus in its natural ecosystems, and provide the most detailed view to date on the interaction of the fungus with its complex commercial compost substrate. Ultimately all functional and behavioral aspects of the mushroom may potentially be decoded by investigating these and additional genome sequences. The availability of whole-genome sequences to the *Agaricus* research community represents a milestone of unprecedented potential.

Keywords: Genomics; Genetics; Breeding

The genome of the 'Button Mushroom' *Agaricus bisporus*, and the genetic diversity incorporated within every cultivated and wild strain of that species, determines the biological potential and fate of each culture and every crop in whatever environment is present. The list of variable, commercially important traits under genetic control is long and growing:

- Yield: vegetative growth rate, optimum temperature for vegetative growth, substrate utilization efficiency,
- Quality: cap pigmentation, cap smoothness, basidiomatal size, shape, firmness, density, bruising responses, processing characteristics, mushroom flavor, agaritine content,
- Development/Morphogenesis: pinning responses and behavior, developmental scheduling of basidiomata,
- Tolerance to environmental stresses: temperature tolerance,
- Resistance to pathogens: responses to pathogens and antagonists including *Pseudomonas, Trichoderma, Lecanicillium fungicola, Mycogone permiciosa,* and dsRNA viruses,
- Breeding efficiency: strain mating behavior, incompatibility phenotype, and life cycle… among others. (For strain identification and intellectual property rights protections, the molecular fingerprint of each strain also derives directly from the primary DNA structure of the genome.)

Knowledge of the genome of *Agaricus* accumulated slowly at first. Twelve key milestones along the way (while acknowledging many other important contributions) are:

1959: Evans, using light microscopy, established the haploid chromosome number at 13 [1].
1972: Raper et al. and Elliott independently confirmed that a heterothallic life cycle existed in *A. bisporus*, masked by the predominant (pseudo-) homothallic behavior of the species, and that traits could segregate and reassort [2, 3].
1982: May and Royse demonstrated the use of codominant (allozyme) genetic markers to confirm successful mated crosses and to provide genetic fingerprints of, and estimated and deduced relationships among, individual strains [4].
1982: Arthur et al. used reassociation kinetics to estimate the size of the haploid genome at 34 Mbp, and also estimated the G-C content and the proportions of single-copy and repeated DNA present [5].
1988: Hintz et al. used restriction mapping to estimate the size of the mitochondrial genome of *A. bisporus* to be 136 kbp [6].
1992: Royer et al. used CHEF analysis to confirm the chromosome number at 13, and the estimated genome size at 30+ Mbp; using Southern hybridizations and linkage-mapped probes, chromosomal length polymorphisms were also demonstrated [7].
1993: Kerrigan et al. provided a comprehensive genetic linkage map to one hybrid heterokaryotic genome of *A. bisporus*, genetically confirmed all facets of normal meiosis and heterothallism (as well as pseudohomothallism), and provided first estimates of varying recombination rates (including non-recombining segments and asymmetrical segregation ratios) over the genome [8].
1993: Xu et al. used knowledge of molecular marker linkage relationships to locate the mating-type gene on chromosome 1, allowing prediction of mating behavior in crosses from genotype data: the beginning of Marker-Assisted Selection (MAS) in *Agaricus* strain breeding [9].
1996: Imbernon et al., using two different hybrid mapping populations, located the recently discovered BSN locus (which exhibits effectively dominant behavior in determining the number of spores formed on the basidium, and thus the effective 'ploidy' of spores and the life cycle they pursue) on chromosome 1, allowing further marker-assisted selection [10].
1999: Callac et al. used linkage markers to locate the primary pileipellis-pigmentation determining locus PPC1 on chromosome 8 [11].
2002: Moquet et al. determined that a primary Quantitative Trait Locus (QTL) determining resistance to *Pseudomonas*-associated 'blotch disease' is closely linked to PPC1 [12].

After the millennium, the pace of molecular genetic, physiological, and other experimental studies on *A. bisporus* continued to accelerate. The important contributions over the past two decades by the teams at Horst/Wageningen (including Sonnenberg and Baars), Warwick/HRI (including Burton, Challen, and Eastwood), INRA Bordeaux (including Foulongne, Callac, Imbernon, Savoie and Largeteau), and, in the private sector, by the Monterey group (including Loftus and Lodder), CTC (Rodier), and my own group at Sylvan, are too numerous to list here. Much of this work is reported in the public record, including patent filings, and can easily be located via online searching. Some work remains confidential and proprietary within private and contracting quasi-public or academic organizations.

As these studies evolved and expanded, it became increasingly obvious that the lack of available genomic sequence for *A. bisporus* was greatly hampering progress in this field. For example, to look for a marker-linked gene target, or the distribution and behavior of a transposable element, one would have to prepare a cosmid library comprising of several thousand individual clones, blot the DNA from the entire genomic library onto several dozen membranes, prepare labelled probes, Southern-hybridize probes to target membranes and develop and capture hybridization signals via enzymatic, radiographic or chemiluminescent methods, prepare flagged cosmid DNA for subcloning (a second iteration of the above steps) or else use direct sequencing and 'walking' along the chromosomal DNA in 500-nt steps, analyze the segment sequence to identify ORFs, transposons, or other features of interest, attempt to classify gene targets based on limited structure-function databases, and then test, via experimentation, hypotheses based on interpretation of the above results. Having carried out precisely such exercises with the team at Sylvan, I can confirm that it is every bit as cumbersome and tedious, not to mention slow, time-consuming and expensive, as it sounds.

The first fungal genome, that of the yeast *Saccharomyces cerevisiae*, was sequenced in 1996. At 12.2 Mbp, it is about one-third the estimated size of the *A. bisporus* genome. Approximately 6000 genes have been identified on 16 chromomes. The first genome sequence of a filamentous fungus, that of *Neurospora crassa*, was published in 2003. This genome, at 43 Mbp, is larger than that of *A. bisporus*, and is believed to incorporate about 10,000 genes.

In 2002, the first of a series of white papers appeared [13], advocating a concerted effort toward obtaining sequences of more fungal genomes. Fifteen candidate fungi were proposed for whole-genome sequencing, and beginning in 2003 many of these sequencing projects were subsequently completed with the resources of the National Human Genome Research Institute (NHGRI), the Whitehead Institute/MIT Center for Genome Research (WICGR), and the Broad Institute. Of particular interest to this readership is the genomics work on agaric fungi (mushrooms), which in the earliest examples includes work on the coprophillic species *Coprinopsis cinereus* (http://genome.jgi-psf.org/Copci1/Copci1.home.html), the mycorrhizal species *Laccaria bicolor* (http://genome.jgi-psf.org/Lacbi2/Lacbi2.home.html) and the wood-inhabiting species *Pleurotus ostreatus* (http://genome.jgi-psf.org/PleosPC15_2/PleosPC15_2.home.html), a cultivated mushroom. While data from these and several other relevant sequencing projects are now available in public databases, it is too early for all of the supporting publications to have appeared in the literature.

In 2003 a core group of *Agaricus* scientists began to build community and industry support for, and seek opportunities for, whole genome sequencing of *A. bisporus*. This group evolved over several years, with key roles played by Kerry Burton, Mike Challen, Anton Sonnenberg, and myself. After submitting a number of proposals, our plan to sequence and compare two distantly-related haploid genomes of *A. bisporus* was accepted by the USA's Department of Energy, Joint Genome Institute, Community Sequencing Program based in Walnut Creek, California. As sequencing neared completion in 2009-2010, the circumstances of some core PIs changed, and Francis Martin generously agreed to take over coordination responsibilities toward collating interpretive data from participating laboratories, and the preparation of a master paper reporting the results of the project.

The focus of our arguments to justify the value of sequencing genomes of *A. bisporus* conformed to the mission of the DOE program: Energy and the Environment. Emphasis in this program is placed upon organisms (or genes) that may participate in the development of new and sustainable fuel and energy sources in formats that conform to current and near-term infrastructure, and on organisms that play key roles, including that of model systems, in ecosystem processes, particularly those that mediate the flow of carbon through the global biological carbon cycle. Additionally, incorporating phylogenetic diversity has always been a consideration of the agencies choosing fungal genomes for sequencing support.

We stressed that *A. bisporus* would be one of the first species in the order *Agaricales*, and would be *the* first species from the large family *Agaricaceae*, to be sequenced. We particularly emphasized that *Agaricus* and *Agaricaceae* belong to an ecological guild adapted to humus-like substrates such as aged forest litter (and commercial mushroom compost). Very little is known about how forest litter is recycled in nature, in spite of the great quantity of leaf, needle and twig material that accumulates and decomposes in forest ecosystems, biomass that contributes greatly to global atmospheric carbon inputs. We predicted that there was a very good chance of finding novel metabolic strategies that might not be seen in the more 'charismatic' white-rot and brown-rot wood-decomposing fungi, nor in coprophillic fungi such as *Coprinopsis*. Our thesis has been that humicolous species (among the 'detritophiles') may have a distinctly different melange, or deployment, of substrate conversion enzymes.

We additionally pointed out that *Agaricus* was now a well-established laboratory organism, and a good model system for various syndromes including the cellular mechanics of pseudohomothallism. In particular, our strategy allowed for the possibility of identifying the BSN locus controlling the basidial spore number trait. Also of interest is the fact that *Agaricus* is the basis of a multibillion dollar food crop industry that employs a large number of people.

Finally, we specifically proposed that one haploid genome, from the commercial European bisporic (var. *bisporus*) homokaryon H97, be sequenced using the Sanger method. A second, wild North American tetrasporic (var. *burnettii*) homokaryon, JB-137-s8, was proposed for sequencing using newer, less expensive short-read technologies, with assembly guided by the finished assembly of the H97 genome. Completing the picture, tens of thousands of ESTs ('Expressed Sequence Tags' obtained from the sequences of messenger RNAs transcribed from actual functioning genes, obtained at several life-cycle stages) would be used to characterize genes of this species, train gene recognition computer algorithms, and directly confirm activity of specific 'gene models'. The mRNAs used were to be obtained from the heterokaryon U1, a widely cultivated hybrid strain and patriarch of the most widely cultivated commercial strain family today. U1 has two haploid nuclear genomes: one from H97, linking half of the EST data directly to the H97 genome sequence, and the other from the H39 homokaryon, providing many opportunities to compare two different allelic gene sequences (H39 vs H97) drawn from within the same European *A. bisporus* var. *bisporus* population.

Thus two scales of comparison were possible: between alleles within the European population, and between genomes of two different varieties of the species from separate populations on two different continents. Other interesting comparisons, for example between wild and cultivated strains, or between compost-adapted wild strains vs. leaf-litter inhabiting strains [14], cannot be made with strong support based only on these first two genomes, but could well be investigated in the near future using additional strains, with the expansion of lower-cost sequencing techniques.

Our final proposal, drafted by Mike Challen, was accepted in 2007 for the 2008 CSP program, and was carried out essentially as we specified. Both genomes are now effectively completed, although (as is typical for most affordable genome projects) a few small gaps remain, and both are now in the public domain on the JGI's genome portal web site http://genome.jgi-psf.org/.

Key findings from the H97 genome sequence assembly are as follows: The 13 nuclear and one mitochondrial chromosome are represented on 29 assembled 'scaffolds', indicating the presence of 15 substantive gaps in the sequence assembly, these gaps comprising not more than 0.7% of the nuclear genome. Some of the 'missing' genome is likely to consist of tandemly arranged ribosomal RNA cassette repeats which are extremely difficult to reassemble into the original genome sequence. Computer search algorithms have identified 10,438 gene 'models', many of which are supported (as representing 'real' genes) by sequenced mRNA transcripts (ESTs). While many of these genes are sufficiently similar to other known genes that their

identities and functions can be predicted with confidence, a considerable number of genes may be unique (in current databases) to *Agaricus*, or may occur more widely but have functions that remain unknown. Genes known to be involved in nutritional pathways are numerous and diverse, allowing comparisons to the profiles of other fungi with different ecological roles. The mitochondrial chromosome of H97 is comprised of 145, 919 nucleotide pairs. Thousands of individual transposable elements (transposons) belonging to numerous and diverse families also inhabit (or infest) the genome; many of these elements have been observed (by Sylvan and by the Wageningen team) to be actively moving from site to site within the genome, and this provides one possible explanation, with many documented precedents of gene disruption, for the development of anomalies such as sectors or stroma from within otherwise healthy, stable cultures.

Since the release of the genome assemblies, teams of researchers have worked to interpret the sequencing and expression data, understand the structure and function of the metabolic pathways deduced from the identities of the confirmed genes, associate those data with the physiological behaviors of the sequenced strains on defined media, and ultimately determine whether *A. bisporus* has a fundamentally distinct 'ecological guild' signature based either on its repertoire of substrate conversion enzymes or on other aspects of its deduced physiology based on gene expression and other data. Most of those analyses are completed, while some are still underway; at this writing the available and developing information is being collated into a cohesive view of the nature of this mushroom and its role in the environment. The finished report is expected to be submitted shortly to a scientific journal, where it will become a part of the public record in the near future. The specific contributions of the many participating laboratories will be detailed and fully credited at that time.

Meanwhile, two opportunities are being pursued concurrently. First, molecular mycologists working on diverse fungi now have two unusual new genomes to compare with those of other organisms; *Agaricus* is becoming one of the standards for genomic and metabolic comparisons, and each study potentially returns additional information of interest to the *Agaricus* community. Second, the laboratories directly involved in the molecular genetics of *Agaricus* now have access to the most powerful tool ever created for enabling progress in this field. It has become a simple matter to obtain the sequence of a gene, compare diversity of a gene sequence among different strains, develop and evaluate genetic markers linked to genes and QTL regions, locate transposons, and carry out many other related tasks. From 2010 forward, the return on monies invested in molecular genetics research on *Agaricus* will be far greater than at any previous time in history.

Problems that could not be solved a few decades ago, at the beginning of my career, which then became comprehensible but intractable, and next became tractable but infeasible, then feasible but prohibitively complex and costly, have ultimately become straightforward, practical, cost-effective and – gradually – even routine to undertake. Such work is well underway. The release of these *Agaricus bisporus* genome sequences, and the forthcoming master paper that will soon document and interpret the sequencing results, represent what may be the pivotal moment in the development of our field.

REFERENCES:

[1] Evans H.J. (1959). Nuclear behaviour in the cultivated mushroom. *Chromosoma* 10: 115-135.

[2] Raper C.A. et al. (1972). Genetic analysis of the life cycle of *Agaricus bisporus*. *Mycologia* 64: 1088-1117.

[3] Elliott T.J. (1972). Sex and the single spore. *Mush. Sci.* 8: 11-18.

[4] May B. and Royse D.J. (1982). Confirmation of crosses between lines of *Agaricus brunnescens* by isozyme analysis. *Exp. Mycol.* 6: 283-292.
[5] Arthur R. et al. (1982). Characterization of the genome of the cultivated mushroom, *Agaricus brunnescens*. *Exp. Mycol.* 7: 127-132.
[6] Hintz W.E.A. et al. (1988). Physical mapping of the mitochondrial genome of the cultivated mushroom *Agaricus brunnescens (= A. bisporus)*. *Curr. Genet.* 14: 43-49.
[7] Royer J.C. et al. (1992). Electrophoretic karyotype analysis of the button mushroom, *Agaricus bisporus*. *Genome* 35: 694-698.
[8] Kerrigan R.W. et al. (1993). Meiotic behavior and linkage relationships in the secondarily homothallic fungus *Agaricus bisporus*. *Genetics* 133: 225-236.
[9] Xu J.-P. et al. (1993). Localization of the mating type gene in *Agaricus bisporus*. *Appl. Environ. Microb.* 59: 3044-3049.
[10] Imbernon M. et al. (1996). *BSN*, the primary determinant of basidial spore number and reproductive mode in *Agaricus bisporus*, maps to chromosome *I*. *Mycologia* 88: 749-761.
[11] Callac P. et al. (1998). Evidence for PPC1, a determinant of the pilei-pellis color of *Agaricus bisporus* fruitbodies. *Fungal Genet. Biol.* 23: 181-188.
[12] Moquet F. et al. (1999). A quantitative trait locus of *Agaricus bisporus* resistance to *Pseudomonas tolaasii* is closely linked to natural cap color. *Fungal Genet. Biol.* 28: 34-42.
[13] http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/FGISEQ2.pdf
[14] Savoie J.-M. et al. (1996). Relative ability of wild isolates and cultivars of *Agaricus bisporus* to degrade conventional mushroom compost: production of extracellular enzyme activities. In: *Mush. Biol. Mush. Prod.* Royse D.J. ed. pp 355-362.