Chapter 3

Omics and the Future of Sustainable Biomaterials

Juliet D. Tang*,[1,2] and Susan V. Diehl[2]

[1]USDA ARS Corn Host Plant Resistance Research, P.O. Box 5367, 810 Highway 12E, Mississippi State, Mississippi 39762
[2]Forest Products, Mississippi State University, 201 Locksley Way, Starkville, Mississippi 39759
***E-mail: juliet.tang@ars.usda.gov.**

With global focus on the conversion of biomass into products, fuels, and energy, there is a strong need for information that will lead to new sustainable products, applications, and biotechnological advances. The omics approach to biology is a discovery-driven method that may deliver solutions to these overarching problems. It gives scientists the ability to obtain a systems-level understanding of life that begins with identifying the genome or genes in an organism. The pivotal technology that enabled this revolutionary approach is next generation DNA sequencing. New fields bring new jargon and new analytical methods making it difficult to appreciate the technology or the significance of the science. Our goal in this review is to present an introduction to the most important omics approaches that have been used to gain insight into how wood decay fungi convert lignocellulose into energy and why certain species are metal-tolerant. The expectation is that once relevant genes are identified, biotechnological methods will give rise to novel solutions that will advance wood protection and utilization.

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

The Genomics Era

The ability to use high throughput technology to sequence a genome, which is the complete set of genetic material or DNA (deoxyribonucleic acid) in an organism, has revolutionized the rate of gene discovery. There are four nucleotide bases that make up the linear sequence of DNA. They are G for guanine, A for adenine, T for thymine, and C for cytosine. Despite this apparent simplicity, it is the sequence of these four nucleotides that governs all cellular activity. Thus, by sequencing genomes, the process of decoding life begins.

The race to sequence genomes did not gain momentum until the Human Genome Project was completed in 2003 (*1*). The project involved hundreds of scientists worldwide, took 13 years to complete, and cost $3 billion. At project completion, (i) millions of DNA fragments or reads were sequenced, (ii) computational and visualization tools were developed to map and assemble the reads from unique overlapping regions into a consensus sequence, (iii) a draft of the human genome, which was 3 billion nucleotides long, was generated, (iv) 20,000 to 25,000 genes were identified and annotated both in terms of their structure and function, and (v) a precedence of open access to data and software tools was established through public repositories like the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI), and the DNA Databank of Japan (DDBJ).

Sequencing Technology History

The Sanger method of sequencing and separation by capillary electrophoresis was developed in 1977 and became the major sequencing technology for over 25 years (*2*). The sequenced nucleotides called reads are generated from fragments of genomic DNA that have been cloned into appropriate bacterial vectors. The Sanger method produces highly accurate reads (500 to 900 bases long), but the maximum instrument output is limited to 96 reads per run or 86 Kbp (kilobase pair) of DNA sequence. In 2005 and 2006, the first next generation sequencing (NGS) prototypes were released. These platforms incorporated three major technological improvements: (i) PCR (polymerase chain reaction) replaced bacterial cloning; (ii) DNA was sequenced during the nucleotide addition step rather than by chain termination; and (iii) by spatially separating the microscopic clusters DNA fragments on a solid surface, millions of genomic DNA fragments could be sequenced in parallel during one run (*3*). As a result, DNA sequencing was transformed into a high throughput technology. This drastically accelerated the speed at which genomes could be sequenced and reduced the costs.

Early concerns arose because the first NGS systems produced much shorter reads with higher error rates. These drawbacks were partially compensated for by the large increase in depth of coverage. Depth of coverage refers to the number of times a nucleotide in a particular position of the genome is represented in the reads. Manufacturers have continually pushed the technology to increase read length, maximize output, and reduce base-calling errors. At the same time, new computational algorithms addressed the problem of read error and solved the

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

daunting problem of assembling the millions of short reads into a continuous linear sequence. These algorithms like Velvet (*4*), Euler (*5*), and SOAPdenovo (*6*) are based on the de Bruijn graph, a mathematical concept of graph theory. The momentum of using NGS to sequence microbial genomes really gained speed, however, when convincing evidence showed that NGS alone or in combination with the Sanger method could produce *de novo* (no known reference sequence) assemblies of two fungal genomes, *Grosmannia clavigera* (*7*) and *Sordaria macrospora* (*8*). Because of these technological advances, genomics rapidly rose to the forefront of all aspects of biological science.

Today, there are three major NGS platforms in popular use: Roche, Illumina, and Life Technologies. Each relies on different detection methods to determine the identity of the added nucleotide during the sequencing reaction, and all three have multiplexing capability to increase the number of libraries or samples that can be simultaneously sequenced. The Roche GS-FLX+ offers the longest read length (700 bp), which has the advantage of reducing the number of gaps in the assembly and reducing the depth of coverage needed to get an assembly. The disadvantages are lower output (400 Mbp) and higher error rates for runs of the same nucleotide that are more than 7 bases long. The Illumina HiSeq2000 outstrips the others in terms of output (600 Gbp) but requires higher coverage when assembling a genome without a reference sequence. It is also by far the cheapest for reagents. The major drawback is the short read length (100 bp), which produces shorter contigs and a draft assembly with many gaps. The Life Technologies Ion Torrent is intermediate in terms of output (10 Gbp) and read length (200 bp), but has the fastest run times. The Ion Torrent like the Roche system has problems reading tracts of homopolymers. All three NGS platforms sell affordable personal or compact sequencers. If current cost trends hold, it will not be long before DNA sequencing and *de novo* assembly of microbial genomes will become routinely affordable (*9*).

Genomes of Fungi that Decay Wood

Basidiomycota

Fungi are a premier resource for bioprospecting because they have a tremendous metabolic diversity that includes many novel gene products and enzymes. To facilitate discovery of gene products with practical applications for energy and the environment, the Joint Genome Institute (JGI) of the Department of Energy has sequenced many basidiomycete genomes (*10–14*). Their sequencing pipeline uses the Sanger method or a hybrid approach that mixes Sanger with the Roche and Illumina platforms. Understanding the role genes play in lignocellulose degradation is relevant for the wood protection industry because it identifies potential targets that can be inhibited to prevent decay. Prior to 2010, there were only two sequenced wood decay genomes, *Phanerochaete chrysosporium* (*13*), a white rot fungus, and *Postia placenta* (*14*), a brown rot fungus. Now, three years later, there are sixteen. Seven are brown rot fungi and ten are white rot fungi (Table 1). Their average genome size is 44.59 Mbp (SD = 11.77) and they contain

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

an average of 13,070 genes (SD = 2926). Others, like the genome of *Ophiostoma piceae*, a causative agent of bluestain, have been assigned Bioproject numbers in NCBI and should be released in the near future.

Table 1. Sequenced Genomes of Brown and White Rot Fungi *

Fungus	*Decay Type*	*Genome (Mbp)*	*Predicted Genes*	*Year*	*Cited*
Coniophora puteana	BR	43.0	13,761	2012	(*11*)
Fibroporia radiculosa	BR	33.6	9,262	2012	(*15*)
Fomitopsis pinicola	BR	46.3	14,724	2012	(*11*)
Gloeophyllum trabeum	BR	37.2	11,846	2012	(*11*)
Postia placenta MAD698-R v1	BR	42.5	12,541	2009	(*14*)
Serpula lacrymans S7.9 v2	BR	42.8	12,917	2011	(*10*)
Serpula lacrymans S7.3 v2	BR	47.0	14,495	2011	(*10*)
Wolfiporia cocos	BR	50.5	12,746	2012	(*11*)
Auricularia delicata	WR	75.1	23,577	2012	(*11*)
Ceriporiopsis subvermispora	WR	39.0	12,125	2012	(*18*)
Dichomitus squalens	WR	42.8	12,290	2012	(*11*)
Fomitoporia mediterranea	WR	74.9	11,333	2012	(*11*)
Ganoderma lucidum	WR	39.9	12,080	2012	(*16*)
Heterobasidion irregulare v2	WR, P	33.6	11,464	2012	(*11*)
Phanerochaete chrysosporium v2	WR	35.1	10,048	2006	(*13*)
Punctularia strigosozonata	WR	34.2	11,538	2012	(*11*)
Stereum hirsutum	WR	46.5	14,072	2012	(*11*)
Schizophyllum commune	WR	38.5	13,210	2010	(*12*)
Trametes versicolor	WR	44.8	14,296	2012	(*11*)

* Abbreviations: BR, brown rot; WR, white rot; P, pathogen; Mbp, megabase pair.

The genomes of *S. lacrymans* S7.3 (*10*), *F. radiculosa* (*15*), and *G. lucidum* (*16*) were among the first fungal genomes to be sequenced entirely by NGS. The species, platforms, and strategies were: *S. lacrymans,* Roche/454, single end (400 bp reads from a 3000 bp library); *F. radiculosa,* Illumina, paired end (76 bp reads from a 300 bp library); and *G. lucidum,* Illumina, paired end (100 bp reads from 200 bp and 6000 bp libraries). Single and paired end strategies indicate whether the library fragments were sequenced from one or both ends, respectively. With

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

very short reads, paired end sequencing effectively increases the short read lengths to the fragment size of the library, even if the intervening sequence is unknown. Furthermore, by varying the insert size of the library, larger gaps between known sequence fragments can be bridged. Therefore, when using NGS to sequence a genome *de novo*, longer assemblies are possible by (i) increasing read length, (ii) using a paired end strategy, (iii) generating sequence from libraries of different size, and (iv) combining sequencing data from more than one NGS platform. The draft assembly contains contigs or regions of contiguous sequence that are connected across gaps of known size into scaffolds. Successful assemblies have been obtained with the following technologies and depth of coverage: 8x for Sanger (*S. lacrymans* 7.9, (*10*)), 40x for a hybrid approach that uses Sanger, Roche, and Illumina (*Fomitopsis pinicola* and *F. mediterranea* (*11*)), 150x for Illumina (*F. radiculosa* (*15*)) and 60x for Roche (*O. novo-ulmi* (*17*)).

Genome Annotation

Structural Annotation

The genetic information embedded in the sequence of the four nucleotides (G, A, T, and C) can be decoded because enough biological knowledge exists to predict protein structure and function from the genes in the genomic sequence. All eukaryotic genes are divided into alternating exons and introns, which are the coding and non-coding regions, respectively. When a gene is expressed, it is transcribed into a precursor messenger RNA (mRNA), which is a complementary copy of the gene without the introns. The mature mRNA is transported out of the nucleus into the cytoplasm where it is read by the ribosomes and translated into protein. Proteins, in turn, are responsible for cell structure and function. Except for a few minor variations, the genetic code, or the translation of codon by a 3 letter nucleotide sequence to amino acid, is universal for ALL organisms. The genetic code also includes codons for start and stop translation sites.

When gene prediction only requires the genomic sequence as input, the method is called *ab initio*. Examples of popular *ab initio* gene prediction tools are: GENSCAN (*19*), AUGUSTUS (*20*), FGENESH (*21*), and GeneMark (*22*). GENSCAN, AUGUSTUS, and FGENESH require a training set of about 1000 validated genes to estimate the model parameters. GeneMark, on the other hand, can learn directly from the input sequence. GeneMark-ES, which was used for gene prediction in *F. radiculosa* (*15*), can improve gene prediction accuracy because it adds a parameter based on the intron branch point sequence (*22*), which is conserved in fungi (*23*). An alternative approach to gene prediction uses extrinsic information from products of gene expression like ESTs (expressed sequence tags), cDNA (complementary DNA sequence copied from mRNA), and protein homologues from other species to predict the location of the intron/exon splice sites. Two widely used homology-based gene prediction tools are FGENESH+ (*21*) and Genewise (*24*). While *ab initio* gene predictions have higher sensitivity, homology-based gene predictions have higher specificity. Therefore, many pipelines will produce an initial set of gene predictions *ab initio,*

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

then improve the predictions using homology-based tools. Combiner tools such as Combiner (*25*) and GLEAN (*26*) then choose the best gene structure from the collection of predictions. This kind of structural annotation pipeline was used by JGI (*10–14*) and for *G. lucidum* (*16*).

Table 2 summarizes the gene structure for the genomes listed in Table 1. As a group, the averages for gene and transcript lengths and number of exons per gene are moderately conserved. Exon and intron lengths are the most and least highly conserved, respectively. Since exons encode proteins, constraints related to function are under the most selection pressure to keep exon length relatively unchanged. The accumulation of nucleotide substitutions, insertions, and deletions in introns, on the other hand, generally have less impact on function since introns are non-coding, and eventually lead to greater variation in intron length. This holds true in a comparison with the plants, *Arabidopsis thaliana*, rice, and maize. The average exon lengths are similar: 237 bp for the wood decay fungi (Table 2) and 217, 254, and 259 bp, respectively, for the plants (*27*). Intron lengths, however, are more variable: 82 bp for the wood decay fungi (Table 2) and 167, 413, and 607 bp, respectively, for the plants (*27*).

Table 2. Summary of Gene Structure from the Genomes Listed in Table 1

Gene Feature	*Mean*	*SD*	*% CV*
Gene length bp	1804	159	9
Transcript length bp	1419	96	7
Exon length bp*	237	12	5
Intron length bp	82	10	12
Exons per gene	5.95	0.46	8

* Mean exon length excluded the value for *H. irregulare* (561 bp), which appeared to be an outlier.

Functional Annotation

Assigning names to genes and determining what they do is the process of genome functional annotation. It begins by using the exon sequences in the genes to predict the amino acid sequence of the polypeptide chain(s) that form the proteins. Chemical properties of the amino acids dictate the biochemical activity of a protein. For example, non-covalent forces generated by the amino acid sequence cause regions of the polypeptides to fold into three dimensional structures that look like coils and sheets. Thus, protein structure is essentially modular where different regions within the protein form specific functional and structural domains. Domains may also contain short amino acid sequences that have binding or catalytic activity. Activity of these site-specific features hinges upon its location within a domain. Thus, how a protein interacts with other

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

molecules is determined not only by its amino acid sequence, but also by the spatial geometry of localized sites and domains within the protein.

Decades of protein analysis have shown that domains and site-specific features are conserved within families of proteins that share the same ancestral gene. Therefore, if a newly predicted protein shows high sequence similarity to another experimentally validated protein, they are assumed to be homologs. The workhorse for performing pairwise sequence similarity searches is BLAST or basic local alignment search tool (*28*). One version of BLAST is blastp. It compares a protein query sequence against every sequence in a public or custom protein database, returning the best hits in the alignment with their alignment scores and E-values (the probability of getting the same hit by chance when searching a database of similar size). The algorithm starts by finding seeds or small "word-size" matches then uses dynamic programming to extend the match to find the local alignments. Statistics are then calculated and the high scoring segment pairs are used to rank the best hits. As of April 9, 2013 for fungi alone, there were 734,575 protein records covering 643 taxa in the NCBI protein database that can be used for alignment comparisons (*29*).

Although evolutionary relationships can be inferred from protein sequence homology, it may or may not imply functional similarity. Small changes in a binding or catalytic site may go undetected in a pair-wise sequence alignment comparison, but could significantly alter protein function. Therefore, additional evidence for protein function is drawn from comparisons of protein domains and site-specific features with the signature or predictive models of all described protein families. Generally, combined evidence from sequence homology and family membership is sufficient to assign a putative function to a predicted gene product.

The major tool for determining the similarity of protein signatures is InterProScan (*30*). InterProScan attempts to classify the function of a query sequence by identifying its protein signature and then comparing it against the signatures of known protein families. InterProScan depends upon InterPro (*31*), which is a single searchable resource that integrates protein information from eleven different databases around the world. Each member database has its own process for identifying and using protein domains and/or site-specific features to classify proteins into protein families. For instance, PROSITE models are based on a collection of biologically significant sites (*32*), while those in Pfam are based on domains (*33*). Position-specific scores can be based on a multiple sequence alignment of functionally validated proteins from different species or on hidden Markov Models (HMM) that calculates a transitional probability for one amino acid given the presence of one or more previous amino acids. At present, there are 1668 documented site-specific features in PROSITE (*34*) and 14,831 functionally annotated protein families in Pfam (*35*).

Organizing and adding context to the functional annotations of proteins is the mission of the Gene Ontology (GO) Consortium (*36*). They developed structured vocabularies or ontologies to describe proteins in terms of three domains or categories: cellular component, biological process, and molecular function. Each term within an ontology has a defined relationship with one or more other terms in the same domain. The relationships (is a type of, is a part of, or is regulated by)

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

are described by directed acyclic graphs, meaning that there are child and parent relationships among terms, but the relationships are not necessarily hierarchichal, as one child term can have multiple parents from different levels.

GO annotations complement sequence homology and InterPro annotations plus have the advantage of being easily queried and analyzed by geneset enrichment analysis. The latter determines which GO terms are more highly represented when comparing genesets from different treatments or conditions. For example, during active lignocellulose degradation, one would expect the number of differentially expressed genes with carbohydrate-active enzyme function to be significantly greater compared to glucose-based growth media or non-decay treatments when a preservative was still protecting the wood. Currently, many databases like InterPro (*31*), KEGG pathways (*37*), MetaCyc (*38*), and the Enzyme Commission (EC (*39*)) have already been mapped to GO and the GO terms are retrieved when a protein search of the individual database is performed. In addition, bioinformatics tools like Blast2GO will automate much of the process (*40*). It performs the blastp alignments to the NCBI nr (non-redundant) protein database, retrieves the GO terms, scans InterPro, gets the EC numbers, and will perform geneset enrichment analysis in a single, easy to learn interface.

Predictions for the cellular location of a gene product can also come directly from the protein sequence. A tool that predicts the mitochondrial or extracellular localization of a gene product is TargetP (*41*). The companion tool, SignalP detects the presence of a signal peptide or signal anchor (*41*). The former suggests that a protein may be secreted and the latter, that a protein may be anchored in a membrane. Since fungi secrete many compounds for defense and for digestion of their food, which occurs extracellularly, most fungal genomes include annotations produced by TargetP and SignalP.

Annotations of wood decay fungal genomes show that the majority of the predicted proteins are homologous with proteins in the NCBI nr database, e.g. 83% for *G. lucidum* (*16*), 89% *F. radiculosa* (*15*), and 68% for *S. lacrymans* S7.3 (*10*). The percentage of genes that map to GO terms for these three species ranged from 40 to 58%. The percentage that actually receive a putative functional assignment is lower and only occurs when there is strong similarity to a documented gene product and its corresponding protein family signature. More often than not, a predicted protein will not match an entire signature, but matches only one domain such as kinase. In this case, the annotation is informative but not specific since kinases are one of the largest superfamilies in eukaryotes. A protein with a hypothetical assignment means it had no database match or matched another hypothetical protein. Extrinsic evidence of gene transcription (e.g. from mRNA and/or protein detection) gives further indication that the gene is doing something in the cell and has a biological role. If there is no evidence of gene expression, then it may be a matter of incorrect timing with respect to developmental stage or environmental conditions, or the gene may be a pseudogene.

There is no doubt that automated genome annotation is a powerful predictive technology that can identify the entire functional potential of an organism. Predictions, though, still need expert review or curation and experimental validation. In additions, portions of genes could be missed or misassembled and

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

functional annotations could be incomplete or incorrect. Nevertheless, genomics lays a comprehensive foundation for subsequent systems-level and gene-by-gene investigations.

Other Omes

NGS was selected as Method of the Year in 2007 (*3*) because it has transformed scientific progress in the area of functional genomics, that is, determining how genomes control cell function at the molecular level. Besides genomics, some of the major applications that arose directly or indirectly from NGS are transcriptomics, proteomics, metabolomics, and metagenomics (Table 3). These emerging omics approaches have presented scientists with the challenge of exploring life as a system of interconnected networks of biochemical reactions within one organism or community of organisms. Since the networks are dynamically changing, it is critically important to link the molecular omics data with indicators like physical, morphological, or chemical changes to the wood. These measurements are more accurate indicators of decay stage than time of exposure to the fungus and provide a meaningful biological context for interpreting and comparing different omics datasets.

Table 3. Fields of Study That Have Advanced Because of NGS and Genomics*

Technology	*Discipline*	*Application*
RNA-Seq	Transcriptomics	Identification and quantification of gene expression
2D-LC MS/MS	Proteomics	Identification and quantification of proteins or gene products
LC-MS/MS, GC-EI-MS	Metabolomics	Identification and quantification of metabolites that are products of enzyme activity
NGS	Metagenomics	Identification and quantification of microbial genomes present in a biological sample without culturing

* Abbreviations: RNA-Seq, RNA-sequencing; 2D, 2 dimensional; MS/MS, tandem mass spectrometry; LC, liquid chromatography; GC-EI-MS, gas chromatography-electron ionization-mass spectrometry.

Transcriptomics

Because DNA and mRNA are complementary, the transcriptome, which is the collection of genes transcribed into mRNA at a particular time in a particular cell or tissue type, can be sequenced and quantified by NGS in a technique called

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

RNA-Seq. Quantitative analysis of differential gene expression is based on counts of the number of reads that align to the predicted CDS (coding DNA sequence) of a reference genome. *De novo* transcriptome assembly should also be possible when a reference genome is not available and has been demonstrated in yeast (*42*).

Exploring dynamic changes in gene expression is a critical first step to address the questions of what, when, where, why, and how genes are turned on. Because mRNA has a half-life of minutes to hours (*43*), the transcriptome is a direct product of gene regulation. By varying the experimental conditions, genes that show differential regulation and correlated expression can be identified. By studying these relationships, scientists gain insights into the signaling pathways that turn genes on or off, the transcription factors that regulate gene activities, and the biological roles played by networks of biochemical pathways. The sequence of the transcriptome also serves to update the genome structural annotations, which tend to evolve as more information is gathered.

Commercial solutions for the analysis of RNA-Seq data are becoming more widespread (e.g. DNASTAR, Madison, WI and Golden Helix, Bozeman, MT). Alternatively, open source solutions can be downloaded separately to perform alignment, differential gene expression analysis, GO enrichment, etc. in individual steps. Despite differences in alignment algorithms and statistical models, a comparison of several tools (three for alignment and five for differential gene expression) showed good agreement among the results (*42*).

Proteomics

The predicted protein sequences generated by genomics has fueled the growth of proteomics, another investigative tool of functional genomics. Shotgun proteomics is a popular method of identifying the proteins in complex mixtures and uses multidimensional protein identification technology (MudPIT) (*44*). After the proteins are extracted, reduced, and trypsin-digested, the peptide fragments are separated and analyzed by 2D-LC MS/MS (2 dimensional-liquid chromatography tandem mass spectrometry). During MS1, peptide masses are identified from the charged molecular ions. In MS2, ions are fragmented to create a fragment ion spectrum. Fragmentation tends to occur at preferred sites along the peptide backbone, which generates a characteristic pattern for the peptide. However, since fragmentation is not consistent, the amino acid sequence of the peptide may only partially be identified. Scoring algorithms like Mascot (*45*) and Sequest (*46*) use a combination of the peptide mass, amino acid sequence, and the fragment ion spectrum to rank the best matches against searches of protein databases. Imcreasing both mass accuracy and coverage will improve the confidence of a protein identification from a tryptic digest.

In quantitative proteomics, global differences in protein abundance are measured. Label and label-free approaches have been developed, but labeling with isobaric tags is the most reproducible and accurate (*47*). Isobaric tags have three regions, a reporter, a balance, and a reactive site. Lighter reporters are paired with heavier balances so that their combined masses are equal or isobaric. After digestion, the peptides from the different samples are covalently attached to a

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

different tag at the reactive site. The samples are combined in equal amounts and analyzed by 2D-LC MS/MS. During MS1 analysis, the labeled peptides co-elute and have the same mass, which removes bias due to the separation method. Label detection occurs during fragmentation in the MS2 stage, when the reporter is cleaved off the peptide. The relative peak intensities of the reporter ions approximate the relative amount of protein in the original samples. Two types of tags are available: tandem mass tags (*48*), which can be used in a multiplex of up to 10 samples, and iTRAQ labels (*49*), which have an 8-plex option.

Because proteins are more stable than mRNA, cells have other mechanisms besides synthesis for controlling their activity. These include post-translational modifications, such as phosphorylation, glycosylation, ubiquitination, methylation, and acetylation, that regulate when and where proteins are active. Alternative splicing of the exons from a single gene can also produce proteins with different properties. Thus, when interpreting proteomics data, quantitative differences suggest but do not always equate with differences in activity.

Metabolomics

The only ome that gives a true snapshot of the physiological state of a cell or organism is the metabolome, the collection of small molecule metabolites that are the intermediates and end products of metabolism. Metabolism is the sum of the chemical reactions that sustain life and is a functional manifestation of the genome, transcriptome, and proteome combined. Untargeted metabolic profiling is still an emerging technology that relies mostly on MS-based techniques. As many metabolites as possible are identified and quantified down to the pico- and femtomole levels.

Each MS technique has its strengths and weaknesses (*50*, *51*). GC-EI-MS (gas chromatography-electron ionization-mass spectrometry) is quantitative and provides retention and peak-rich fragmentation spectra, but mass accuracy is generally too low to be useful. High resolution LC-MS/MS instruments like the Orbitrap and quadrupole time of flight mass spectrometers give accurate mass values, but relatively few peaks are found in the MS/MS spectra and the method is only semi-quantitative. Metabolite identifications rely on matching masses, retention indices, and MS/MS spectra to reference compounds. METLIN (*52*), a public database from the Scripps Institute, currently holds 56,612 tandem MS spectra for 11,208 ions (*53*). They also offer Xcms Online, which is an easy to use interface for uploading MS data, searching METLIN, and viewing results (*54*). One feature of the software is the cloud plot (*55*), which gives a global representation of the data in one diagram. For each sample analyzed, the plot simultaneously shows the metabolites, the direction and magnitude of the fold change, its p-value, whether matches were found in the METLIN database, m/z values, elution time, solvent profile, and the total ion chromatogram. The NIST Mass Spectral Library (NIST 11) is more comprehensive than METLIN, but access requires a paid subscription. It has 212,961 EI spectra, 121,586 tandem MS spectra for 15,180 ions, and 346,757 retention indices for 70,835 compounds (*56*).

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

Integrating Omics Data

Mapping omics datasets to biochemicals pathways to determine the link between metabolic networks and biological function can be performed with resources like KEGG (*37*) and MetaCyc (*38*). For example, BioCyc, which is based on the MetaCyc collection of metabolic pathways, uses Pathway Tools (*57*) to create an organism-specific pathway genome database. The metabolic pathways and regulatory networks are predicted from the genomic annotations by automated and/or manual curation. The Omics Viewer in Pathway Tools simplifies the interpretation of large-scale data sets by overlaying transcriptomics, proteomics, and metabolomics data onto the pathway networks so that quantitative differences of the transcripts, proteins, and metabolites can be viewed in their entire metabolic context. Pathway tools also incorporates a genome browser that provides graphical visualization of genomic features like the genes, their orientation, positional coordinates, and transcriptional evidence.

For practical applications like biocide or drug targeting, Pathway Tools is designed to find chokepoint reactions. A chokepoint reaction is the only producer or consumer of a metabolite in a metabolic network. Search modifiers can restrict the output to targets that are specific, selective, safe for humans, and pronounced in their effects. Thus, these kinds of functional genomics studies have the potential to greatly accelerate the discovery rate of targets for small molecule inhibition. This has clear benefits for the wood protection industry. By using a rational approach to the development of wood protection chemicals, molecular targets can be pinpointed and inhibitor effects on both target and non-target species can be predicted.

Metagenomics

A combined approach using Sanger and Roche NGS technology (or one that produces reads at least 400 bases long) has begun to transform studies on microbial ecology. This application, known as metagenomics, eliminates the bias introduced by culturing, since DNA is extracted directly from the environmental sample. It can be used semi-quantitatively to compare the functional potential of different microbial communities and to broadly identify the responsible taxonomic groups. At this point, the technology has not been used to investigate fungal decay communities, but it has been successful for describing key degraders in an anaerobic community on poplar chips (*58*). Deducing structural and functional annotations from genomes of non-eukaryotes is more straightforward because bacteria and *Archaea* lack introns, their protein sequences are more highly conserved, even among distantly related species, and more reference genomes have been sequenced. Programs like Phylogenetic Marker COGs (IMG/M-ER from the JGI website) automate the process of identifying COGs or clusters of orthologous groups from genomic sequence, and, in certain cases, can link the contigs based on their COGs to a phylogenetic group. Phylogenetically identified contigs can then be used to train metagenomic gene classifier algorithms (*59*) to bin the remaining unclassified contigs. Databases that focus on single groups of molecules like CAZy (Carbohydrate-Active enZYmes (*60*)) can also be searched

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

to assign genes more detailed functional annotations regarding polysaccharide degradation. Thus, inferences about the functional role of each taxonomic cluster in a decay community can be made. Given the current read lengths, average exon and intron sizes of fungal genomes, growth of KOGs (Eukaryotic Orthologous Groups of Proteins) and number of fungal genomes sequenced, metagenomics of fungal decay communities is technologically feasible.

Many of the detected genomes will not have a reference genome in a public database, but a more specific inventory of the taxonomic diversity in an environmental sample can be deduced by combining metagenomics with another NGS application called targeted amplicon sequencing. For fungi, the most widely used targeted amplicon is the variable ITS (internal transcribed spacer) region of the ribosomal DNA repeat unit (*61*), which has been extensively sequenced. This approach has been used to profile 1914 OTUs (operational taxonomic units) present on 343 samples taken from decaying Norway spruce logs in two different environments in Sweden (*62*). OTU composition between logs was very different and only a few OTUs were present in most logs; basidiomycetes were more abundant than ascomycetes, but ascomycetes were more widespread; and species diversity increased with stage of decay. Sporocarp production was also determined to be a poor indicator of OTU composition and diversity. Moreover, new primer sequences have been published that are better suited for NGS technology (*63*). They target the ITS2 region of the ribosomal DNA repeat, introduce less bias against species with longer amplicons, and capture more diversity than the traditionally used ITS1F primer (*64*).

Accelerating Discovery

Today's scientific and technological environment is highly competitive and rapidly changing both for university researchers and industries. With a global focus on the conversion of biomass into products, fuels, and energy, there is a strong need for information and understanding that will lead to new sustainable products and applications. Wood decay fungi are unique. They are one of the few groups of organisms that can fully degrade wood. These fungi have evolved unique non-specific yet complex mechanisms to break down the lignocellulosic matrix of wood. The white rot decay fungi preferentially remove lignin or cause simultaneous degradation removing lignin, cellulose and hemicellulose together. Furthermore, white rot fungi only remove wood that is in direct contact with their hyphae, relying on the localized action of their secreted enzymes to digest the wood polymers down to their individual components. Brown rot decay fungi, on the other hand, selectively remove cellulose and hemicelluloses, leaving behind a modified lignin that constitutes humus. They initiate decay more pervasively using hydroxyl free radicals produced by the Fenton reaction. Because of the disruption caused by the free radicals, the secreted carbohydrate-active enzymes like glycoside hydrolases (GH) can reach their substrates even in regions not in direct contact with hyphae and catalyze the breakdown of cellulose and hemicelluloses down to their composite sugars. Comparisons of these fungal genomes through genomics, transcriptomics, proteomics, and metabolomics will

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

open the door to many new discoveries and a better understanding of how these complex decay mechanisms differ among gene family members, species, wood substrates, and environmental factors.

Comparative Genomics

Results from phylogenomics indicate that wood decay fungi have literally changed the landscape of our planet (*11*). During the late Carboniferous Period into the early Permian, large deposits of black coal were formed (*65*). During the Permian Period, coal deposition dramatically decreased, setting limits to the world's coal supply. Using a molecular clock approach to phylogenetic analysis, Floudas et al. (*11*) deduced that the first ligninolytic peroxidase gene arose about 295 million years ago, overlapping with the Permian period. These data led the authors to postulate a causal relationship between the evolution of lignin-degradation in white rot fungi and the decline in black coal formation (*66*).

A comparative study of 27 gene families revealed major differences among the 31 wood decay genomes that could explain the functional differences in decay between white and brown rot fungi (*11*). The genomes of the white rot fungi contained more carbohydrate-active genes (61-148 genes) than the brown rot fungi (32-68 genes). In particular, genes encoding the group of enzymes that act on crystalline cellulose were present in all white rot genomes, but were absent in all brown rot genomes. For lignin degradation the peroxidases, such as lignin peroxidase (LiP), manganese peroxidase (MnP), and versatile peroxidase (VP), were found in the white rot genome but were absent in the brown rot genomes. An evolutionary analysis suggested that the most recent ligninolytic ancestor was a white rot fungus containing a MnP gene, and that LiP arose from a single origin while VP arose twice. While there was expansion of the peroxidase genes in the white rot lines, there appeared to be contraction and eventual loss of the peroxidase genes in the different brown rot lineages.

Comparative genomics has also been used to gain insight into the two modes of white rot decay. Fernandez-Fueyo et al. (*18*) compared the genomes of the white rot fungi, *Ceriporiopsis subvermispora* and *P. chrysosporium*. The species that exhibits selective delignification, *C. subvermispora*, contained over twice the number of MnP genes, ten-fold fewer LiP genes, and seven laccase genes versus no laccase in the simultaneous decay species *P. chrysosporium*. This implies that MnP and laccase may govern selective delignification, while lignin removal during simultaneous decay depends more on LiP.

Suzuki et al. (*67*) used comparative genomics to understand how *P. carnosa* can survive exclusively on softwoods, while other *Phanerochaete* species like *P. chrysosporium* prefer hardwoods. Some differences were found in the carbohydrate-active gene families like the GH5 mannanases. Differences in abundances of lignin-degrading genes were also detected. *P. chrysosporium* had more LiP and fewer MnP genes, which was consistent with its ability to grow better on pulp with a high lignin content. The most striking discovery was the large number of cytochrome P450 monooxygenase genes in *P. carnosa* (266)

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

versus *P. chrysosporium* (149). The authors suggest that *P. carnosa* is able to use the high number of P450 genes to detoxify wood extractives, and in conjunction with its lignolytic genes, degrade the heartwood of softwood.

Functional Genomics of Wood Decay

The discovery that brown rot fungi have multiple functional genes for laccase, but lack the typical cellulases and peroxidases found in white rot fungi arose from research by Martinez et al. (*14*). They combined genomics with gene expression microarrays and proteomics to elucidate the lignocellulose degradation mechanisms of *P. placenta*. With respect to carbohydrate degradation, they found genes for numerous hemicellulases and one endoglucanase, but none had the predicted functional domains and sites to bind and degrade crystalline cellulose directly. Expression of these gene products were confirmed by the microarray study and proteomics analysis of the secretome. They also uncovered numerous expressed genes that had putative functions related to Fenton chemistry like iron reductases, quinone reductase, and oxidases, specifically copper radical oxidases and GMC oxidoreductases.

MacDonald et al. (*68*) used RNA-Seq and qRT-PCR (*69*) to compare the transcriptome of the white rot fungus *P. carnosa* during growth on different wood species and nutrient media. Three MnPs genes were highly expressed on wood, but no unique differences were found when comparing growth on softwood or hardwood. Some of the peroxidase genes, though, were expressed at different levels on the different wood species. The qRT-PCR study revealed a higher abundance of MnPs and LiPs at early stages of cultivation and higher levels of carbohydrate-active enzymes at later stages, suggesting a sequential mode of degradation where lignin is degraded to some extent prior to the carbohydrates.

Using genome screening tools, directed mutagenesis and heterologous expression, Fernandez-Fueyo et al. (*70*) identified 16 peroxidase genes in the selective ligninolytic white rot fungus *C. subvermispora*. The list included 13 putative MnP genes, one generic peroxidase and two unusual LiP/VP genes. The latter were phylogenetically and catalytically intermediate between the standard LiPs and VPs, suggesting the existence of a VP-LiP transitional stage. Other species of white rot fungi also exhibit large MnP gene families. Salame et al. (*71*) used homologous recombination (*72*) to selectively inactivate different *mnp* genes. They found that inactivation (i) did not affect expression of the non-targeted MnPs, (ii) did not reduce decolorization of the substrate orange II, (iii) but did significantly reduce total MnP activity in enzyme assays. They concluded that there was functional redundancy among these genes and that a reduction of one gene was compensated by other gene family members. Less functional redundancy was observed for the laccase gene family in *P. ostreatus*. Pezzella et al. (*73*) used qRT-PCR to quantify transcription of the laccase genes under different conditions and developmental stages. The results depicted a picture of very complex regulation. Some laccase genes were strongly expressed under certain culture conditions, others were expressed only during sporocarp formation, but the majority were poorly expressed under all tested conditions.

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

Another noteworthy study employed metabolomics to determine which metabolite was the iron reductant that drives Fenton chemistry of the brown rot fungus *S. lacrymans* (Boletales) (*74*). It had been proposed that the catechol variegatic acid was the main reductant in this fungus (*10*), although brown rot fungi from other lineages (Gloeophyllales and Polyporales) use 2,5-dimethoxyhydroquinone (DMHQ) (*75*). Korripally et al. (*74*) found that decaying wood contained no variegatic acid but had significant levels of DMHQ, supporting the involvement of this metabolite in Fenton chemistry in all three divergent brown rot lineages.

Functional Genomics of Wood Preservative Tolerance

Certain brown rot fungi like *F. radiculosa* have the ability to overcome copper-based wood preservatives and decay pressure-treated wood (*76*). However, genetic control of these mechanisms is poorly understood. Using RNA-Seq on the transcriptome of *F. radiculosa*, Tang et al. (*77*) identified 917 transcripts that were differentially regulated during decay of wood treated with a copper-based preservative. Over 100 of these genes had functions related to wood decay. Fungal metabolism appeared to change dramatically depending upon whether the RNA was taken when the preservative was protecting the wood or when the preservative lost its efficacy and the wood exhibited high strength loss. Using the results of the transcriptomics analysis as a guide, Tang et al. (*77*) then used qRT-PCR (quantitative reverse transcription-PCR) to perform a time-course study of ten of the differentially regulated genes. They found that the copper-based preservative induced higher than normal expression for several genes in the shortcut pathway of oxalate biosynthesis, oxalate breakdown, and for laccase (*77*). Laccase is a multi-copper oxidase that is induced by copper (*78*). These results reinforce the theory that copper oxalate crystal formation is the mechanism of copper tolerance in brown rot fungi (*79*, *80*) and supports a role for laccase in initiating the Fenton reaction of high-oxalate producing brown rot fungi (*81*). Expression of a GH5 and GH10 gene with putative roles in the degradation of cellulose and hemicelluloses, respectively, were repressed by the preservative until the preservative lost its efficacy and wood showed high strength loss (*77*). The significance of these results is that it provides several targets that could be used to prevent decay initiation in copper-tolerant species of brown rot fungi.

Functional Genomics of Wood Attacking Insects

Proteomic analysis of *Fusarium solani*, isolated from the gut of the longhorned beetle, detected the presence of cellulases, glycosyl hydrolases, xylanases, laccases, and Mn-independent peroxidases (*82*). Tartar et al. (*83*) compared the digestive contributions of the gut of the termite, *Reticulitermes flavipes*, versus its symbionts using a parallel metatranscriptome analysis. They attributed phenoloxidase activity to host-derived laccase genes and found that termite laccases were phylogenetically similar to fungal laccases. Cellulase genes with complementary activity for cellulose degradation were contributed cooperatively by host and symbionts, while genes for hemicellulose degradation

were all symbiont-derived. A later investigation used Roche NGS and proteomics to evaluate the metagene expression of the termite gut and its symbionts when fed diets of different lignin complexity (*84*). The findings revealed two detoxification enzyme families not previously associated with lignin digestion in termites: aldo-keto reductases and catalases. Recombinant versions of these host enzymes showed they significantly enhanced lignocellulose breakdown.

Future Perspectives

It is truly an exciting time to be studying wood biodegradation. Because of NGS, gene discovery is no longer a rate limiting step. Once gene targets and their inhibitors are identified, strategies to prevent decay might involve including the target-specific inhibitors in existing preservative formulations, covalently attaching them to wood by chemical modification, or even genetically engineering the trees to express and concentrate the inhibitors in the wood cell wall layers. All these options stem from a rational approach that uses our understanding of systems biology to develop environmentally sustainable approaches to enhance wood protection and utilization.

References

1. About the Human Genome Project. http://www.ornl.gov/hgmis/home.shtml (accessed June 22, 2013).
2. Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463–7.
3. Kiermer, V. Method of the Year. *Nat. Methods* **2008**, *5* DOI: 10.1038/nmeth1153.
4. Zerbino, D. R.; Birney, E. *Genome Res.* **2008**, *18*, 821–29.
5. Chaisson, M. J.; Brinza, D.; Pevzner, P. A. *Genome Res.* **2009**, *19*, 336–46.
6. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. *Gigascience* **2012**, *1*, 18.
7. Diguistini, S.; Liao, N. Y.; Platt, D.; Robertson, G.; Seidel, M.; Chan, S. K.; Docking, T. R.; Birol, I.; Holt, R. A.; Hirst, M.; et al. *Genome Biol.* **2009**, *10*, R94.
8. Nowrousian, M.; Stajich, J. E.; Chu, M.; Engh, I.; Espagne, E.; Halliday, K.; Kamerewerd, J.; Kempken, F.; Knab, B.; Kuo, H. C.; et al. *PLoS Genet.* **2010**, *6*, e1000891.
9. Chin, C. S.; Alexander, D. H.; Marks, P.; Klammer, A. A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E. E.; et al. *Nat. Methods* **2013**, *10*, 563–9.
10. Eastwood, D. C.; Floudas, D.; Binder, M.; Majcherczyk, A.; Schneider, P.; Aerts, A.; Asiegbu, F. O.; Baker, S. E.; Barry, K.; Bendiksby, M.; et al. *Science* **2011**, *333*, 762–5.
11. Floudas, D.; Binder, M.; Riley, R.; Barry, K.; Blanchette, R. A.; Henrissat, B.; Martinez, A. T.; Otillar, R.; Spatafora, J. W.; Yadav, J. S.; et al. *Science* **2012**, *336*, 1715–9.

12. Ohm, R. A.; de Jong, J. F.; Lugones, L. G.; Aerts, A.; Kothe, E.; Stajich, J. E.; de Vries, R. P.; Record, E.; Levasseur, A.; Baker, S. E.; et al. *Nat. Biotechnol.* **2010**, *28*, 957–63.
13. Vanden Wymelenberg, A.; Minges, P.; Sabat, G.; Martinez, D.; Aerts, A.; Salamov, A.; Grigoriev, I.; Shapiro, H.; Putnam, N.; Belinky, P.; et al. *Fungal Genet. Biol.* **2006**, *43*, 343–56.
14. Martinez, D.; Challacombe, J.; Morgenstern, I.; Hibbett, D.; Schmoll, M.; Kubicek, C. P.; Ferreira, P.; Ruiz-Duenas, F. J.; Martinez, A. T.; Kersten, P.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 1954–59.
15. Tang, J. D.; Perkins, A. D.; Sonstegard, T. S.; Schroeder, S. G.; Burgess, S. C.; Diehl, S. V. *Appl. Environ. Microbiol.* **2012**, *78*, 2272–81.
16. Liu, D.; Gong, J.; Dai, W.; Kang, X.; Huang, Z.; Zhang, H. M.; Liu, W.; Liu, L.; Ma, J.; Xia, Z.; et al. *PLoS One* **2012**, *7*, e36146.
17. Forgetta, V.; Leveque, G.; Dias, J.; Grove, D.; Lyons, R., Jr.; Genik, S.; Wright, C.; Singh, S.; Peterson, N.; Zianni, M.; et al. *J. Biomol. Technol.* **2013**, *24*, 39–49.
18. Fernandez-Fueyo, E.; Ruiz-Duenas, F. J.; Ferreira, P.; Floudas, D.; Hibbett, D. S.; Canessa, P.; Larrondo, L. F.; James, T. Y.; Seelenfreund, D.; Lobos, S.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5458–63.
19. Burge, C.; Karlin, S. *J. Mol. Biol.* **1997**, *268*, 78–94.
20. Stanke, M.; Waack, S. *Bioinformatics* **2003**, *19* (Suppl 2), ii215–25.
21. Salamov, A. A.; Solovyev, V. V. *Genome Res.* **2000**, *10*, 516–22.
22. Ter-Hovhannisyan, V.; Lomsadze, A.; Chernoff, Y. O.; Borodovsky, M. *Genome Res.* **2008**, *18*, 1979–90.
23. Kupfer, D. M.; Drabenstot, S. D.; Buchanan, K. L.; Lai, H.; Zhu, H.; Dyer, D. W.; Roe, B. A.; Murphy, J. W. *Eukaryot. Cell* **2004**, *3*, 1088–100.
24. Birney, E.; Clamp, M.; Durbin, R. *Genome Res.* **2004**, *14*, 988–95.
25. Allen, J. E.; Pertea, M.; Salzberg, S. L. *Genome Res.* **2004**, *14*, 142–8.
26. Elsik, C. G.; Mackey, A. J.; Reese, J. T.; Milshina, N. V.; Roos, D. S.; Weinstock, G. M. *Genome Biol.* **2007**, *8*, R13.
27. Haberer, G.; Young, S.; Bharti, A. K.; Gundlach, H.; Raymond, C.; Fuks, G.; Butler, E.; Wing, R. A.; Rounsley, S.; Birren, B.; et al. *Plant Physiol.* **2005**, *139*, 1612–24.
28. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389–402.
29. NCBI RefSeq Release 59 Statistics. ftp://ftp.ncbi.nih.gov/refseq/release/release-statistics/ (accessed June 22, 2013).
30. Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. *Nucleic Acids Res.* **2005**, *33*, W116–20.
31. Hunter, S.; Jones, P.; Mitchell, A.; Apweiler, R.; Attwood, T. K.; Bateman, A.; Bernard, T.; Binns, D.; Bork, P.; Burge, S.; et al. *Nucleic Acids Res.* **2012**, *40*, D306–12.
32. Sigrist, C. J.; de Castro, E.; Cerutti, L.; Cuche, B. A.; Hulo, N.; Bridge, A.; Bougueleret, L.; Xenarios, I. *Nucleic Acids Res.* **2013**, *41*, D344–7.
33. Punta, M.; Coggill, P. C.; Eberhardt, R. Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; et al. *Nucleic Acids Res.* **2012**, *40*, D290–301.

34. PROSITE. http://prosite.expasy.org (accessed June 22, 2013).
35. Pfam. http://pfam.sanger.ac.uk/ (accessed June 22, 2013).
36. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. *Nat. Genet.* **2000**, *25*, 25–9.
37. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. *Nucleic Acids Res.* **2012**, *40*, D109–14.
38. Caspi, R.; Altman, T.; Dreher, K.; Fulcher, C. A.; Subhraveti, P.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; et al. *Nucleic Acids Res.* **2012**, *40*, D742–53.
39. Enzyme Commission. http://www.chem.qmul.ac.uk/iubmb/enzyme (accessed June 22, 2013).
40. Götz, S.; García-Gómez, J. M.; Terol, J.; Williams, T. D.; Nagaraj, S. H.; Nueda, M. J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. *Nucleic Acids Res.* **2008**, *36*, 3420–35.
41. Emanuelsson, O.; Brunak, S.; von Heijne, G.; Nielsen, H. *Nat. Protoc.* **2007**, *2*, 953–71.
42. Nookaew, I.; Papini, M.; Pornputtapong, N.; Scalcinati, G.; Fagerberg, L.; Uhlen, M.; Nielsen, J. *Nucleic Acids Res.* **2012**, *40*, 10084–97.
43. Kebaara, B. W.; Nielsen, L. E.; Nickerson, K. W.; Atkin, A. L. *Genome* **2006**, *49*, 894–9.
44. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat. Biotechnol.* **2001**, *19*, 242–7.
45. Perkins, D. N.; Pappin, D. J.; Creasy, D. M. *Electrophoresis* **1999**, *20*, 3551–67.
46. Eng, J. K.; McCormick, A. L.; Yates, J. R., 3rd *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–89.
47. Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C. *J. Proteome Res.* **2012**, *11*, 1582–90.
48. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. *Anal. Chem.* **2003**, *75*, 1895–904.
49. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; et al. *Mol. Cell Proteomics* **2004**, *3*, 1154–69.
50. Scheubert, K.; Hufsky, F.; Bocker, S. *J. Cheminf.* **2013**, *5*, 12.
51. Lei, Z.; Huhman, D. V.; Sumner, L. W. *J. Biol. Chem.* **2011**, *286*, 25435–42.
52. Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2012**, *30*, 826–8.
53. METLIN. http://metlin.scripps.edu/ (accessed June 22, 2013).
54. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035–9.
55. Patti, G. J.; Tautenhahn, R.; Rinehart, D.; Cho, K.; Shriver, L. P.; Manchester, M.; Nikolskiy, I.; Johnson, C. H.; Mahieu, N. G.; Siuzdak, G. *Anal. Chem.* **2013**, *85*, 798–804.
56. NIST Standard Reference Database 1A. http://www.nist.gov/srd/nist1a.cfm (accessed June 22, 2013).

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

57. Karp, P. D.; Paley, S. M.; Krummenacker, M.; Latendresse, M.; Dale, J. M.; Lee, T. J.; Kaipa, P.; Gilham, F.; Spaulding, A.; Popescu, L.; et al. *Briefings Bioinf.* **2010**, *11*, 40–79.
58. van der Lelie, D.; Taghavi, S.; McCorkle, S. M.; Li, L. L.; Malfatti, S. A.; Monteleone, D.; Donohoe, B. S.; Ding, S. Y.; Adney, W. S.; Himmel, M. E.; et al. *PLoS One* **2012**, *7*, e36740.
59. Mande, S. S.; Mohammed, M. H.; Ghosh, T. S. *Briefings Bioinf.* **2012**, *13*, 669–81.
60. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233–8.
61. Lindahl, B. D.; Nilsson, R. H.; Tedersoo, L.; Abarenkov, K.; Carlsen, T.; Kjoller, R.; Koljalg, U.; Pennanen, T.; Rosendahl, S.; Stenlid, J.; et al. *New Phytol.* **2013**, *199*, 288–99.
62. Kubartova, A.; Ottosson, E.; Dahlberg, A.; Stenlid, J. *Mol. Ecol.* **2012**, *21*, 4514–32.
63. Ihrmark, K.; Bodeker, I. T.; Cruz-Martinez, K.; Friberg, H.; Kubartova, A.; Schenck, J.; Strid, Y.; Stenlid, J.; Brandstrom-Durling, M.; Clemmensen, K. E.; et al. *FEMS Microbiol. Ecol.* **2012**, *82*, 666–77.
64. Gardes, M.; Bruns, T. D. *Mol. Ecol.* **1993**, *2*, 113–8.
65. Thomas, L. *Coal Geology*; John Wiley & Sons, Ltd: West Sussex, England, 2002; p 384.
66. Robinson, J. M. *Geology* **1990**, *18*, 607–10.
67. Suzuki, H.; MacDonald, J.; Syed, K.; Salamov, A.; Hori, C.; Aerts, A.; Henrissat, B.; Wiebenga, A.; VanKuyk, P. A.; Barry, K.; et al. *BMC Genomics* **2012**, *13*, 444.
68. MacDonald, J.; Doering, M.; Canam, T.; Gong, Y.; Guttman, D. S.; Campbell, M. M.; Master, E. R. *Appl. Environ. Microbiol.* **2011**, *77*, 3211–8.
69. Macdonald, J.; Master, E. R. *Appl. Environ. Microbiol.* **2012**, *78*, 1596–600.
70. Fernandez-Fueyo, E.; Ruiz-Duenas, F. J.; Miki, Y.; Martinez, M. J.; Hammel, K. E.; Martinez, A. T. *J. Biol. Chem.* **2012**, *287*, 16903–16.
71. Salame, T. M.; Knop, D.; Levinson, D.; Yarden, O.; Hadar, Y. *Appl. Environ. Microbiol.* **2013**, *79*, 2405–15.
72. Salame, T. M.; Knop, D.; Tal, D.; Levinson, D.; Yarden, O.; Hadar, Y. *Appl. Environ. Microbiol.* **2012**, *78*, 5341–52.
73. Pezzella, C.; Lettera, V.; Piscitelli, A.; Giardina, P.; Sannia, G. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 705–17.
74. Korripally, P.; Timokhin, V. I.; Houtman, C. J.; Mozuch, M. D.; Hammel, K. E. *Appl. Environ. Microbiol.* **2013**, *79*, 2377–83.
75. Suzuki, M. R.; Hunt, C. G.; Houtman, C. J.; Dalebroux, Z. D.; Hammel, K. E. *Environ. Microbiol.* **2006**, *8*, 2214–23.
76. Clausen, C. A.; Jenkins, K. M. *Chronicles of Fibroporia radiculosa (=Antrodia radiculosa) TFFH 294*; 240; U.S. Department of Agriculture, Forest Service, Forest Products Laboratory: Madison, WI, 2011; pp 1−5.
77. Tang, J. D.; Parker, L. A.; Perkins, A. D.; Sonstegard, T. S.; Schroeder, S. G.; Nicholas, D. D.; Diehl, S. V. *Appl. Environ. Microbiol.* **2013**, *79*, 1523–33.

78. Piscitelli, A.; Giardina, P.; Lettera, V.; Pezzella, C.; Sannia, G.; Faraco, V. *Curr. Genomics* **2011**, *12*, 104–12.
79. Green, F.; Clausen, C. A. *Int. Biodeterior. Biodegrad.* **2003**, *51*, 145, 9.
80. Clausen, C. A.; Green, F. *Int. Biodeterior. Biodegrad.* **2003**, *51*, 139–44.
81. Wei, D.; Houtman, C. J.; Kapich, A. N.; Hunt, C. G.; Cullen, D.; Hammel, K. E. *Appl. Environ. Microbiol.* **2010**, *76*, 2091–7.
82. Scully, E. D.; Hoover, K.; Carlson, J.; Tien, M.; Geib, S. M. *PLoS One* **2012**, *7*, e32990.
83. Tartar, A.; Wheeler, M. M.; Zhou, X.; Coy, M. R.; Boucias, D. G.; Scharf, M. E. *Biotechnol. Biofuels* **2009**, *2*, 25.
84. Sethi, A.; Slack, J. M.; Kovaleva, E. S.; Buchman, G. W.; Scharf, M. E. *Insect Biochem. Mol. Biol.* **2013**, *43*, 91–101.

Deterioration and Protection of Sustainable Biomaterials

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

ACS SYMPOSIUM SERIES **1158**

Deterioration and Protection of Sustainable Biomaterials

Tor P. Schultz, Editor
Silvaware, Inc.
Starkville, Mississippi

Barry Goodell, Editor
Virginia Polytechnic Institute and State University (Virginia Tech)
Blacksburg, Virginia

Darrel D. Nicholas, Editor
Mississippi State University
Mississippi State, Mississippi

Sponsored by the
ACS Division of Cellulose and Renewable Materials

American Chemical Society, Washington, DC

Distributed in print by Oxford University Press

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

Library of Congress Cataloging-in-Publication Data

Deterioration and protection of sustainable biomaterials / Tor P. Schultz, editor, Silvaware, Inc., Starkville, Mississippi, Barry Goodell, editor, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, Virginia, Darrel D. Nicholas, editor, Mississippi State University Mississippi State, Mississippi ; sponsored by the ACS Division of Cellulose and Renewable Materials.
pages cm. -- (ACS symposium series ; 1158)
Includes bibliographical references and index.
ISBN 978-0-8412-3004-0 (alk. paper)
1. Biomedical materials--Congresses. 2. Biomedical materials--Deterioration--Congresses. 3. Sustainable engineering--Congresses. I. Schultz, Tor P., 1953- editor of compilation. II. Goodell, Barry, editor of compilation. III. Nicholas, Darrel D., editor of compilation. IV. American Chemical Society. Cellulose and Renewable Materials Division.
R857.M3D48 2014
610.28′4--dc23

2014013674

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48n1984.

Distributed in print by Oxford University Press

PRINTED IN THE UNITED STATES OF AMERICA

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department

In Deterioration and Protection of Sustainable Biomaterials; Schultz, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2014.