Appendix 9
Glossary

algae: Photosynthetic, aquatic, eukaryotic organisms that contain chlorophyll but lack terrestrial plant structures (e.g., roots, stems, and leaves). Algae can exist in many sizes ranging from single cells to giant kelps several feet long.

algorithm: Formal set of instructions that tells a computer how to solve a problem or execute a task. A computer program typically consists of several algorithms.

annotation: Addition of biologically meaningful descriptions to data (e.g., by labeling regions of sequence data that encode a gene or regulatory region or by identifying the active site of a protein structure).

application programming interface (API): A set of standardized messages or protocols that a program can use to communicate with and request services from another program.

archaea: One of the three domains of life (along with bacteria and eukarya) distinguished through DNA sequence analysis. Archaea are structurally and metabolically similar to bacteria but share some features of their molecular biology with eukaryotes.

architecture: Operational structure of a computer system.

bacteria: One of the three domains of life (along with archaea and eukarya) distinguished through DNA sequence analysis. Also a general term referring to prokaryotic organisms that do not belong to the archaea domain (singular: bacterium).

Bayesian approach: Use of statistical methods that assign probabilities or distributions to future events based on knowledge of prior events.

bioenergy: Energy-related product (e.g., solid, liquid, or gaseous fuels; electricity; and heat) derived from renewable biobased materials (e.g., plant matter and organic waste) or biological processes (e.g., biochemical activities of microbes or plants).

biofilm: Community of microorganisms living together on a surface and embedded in extracellular polymers they create.

biogeochemistry: Study of how interactions among biological and geochemical processes influence the global cycling of such essential elements as carbon, nitrogen, phosphorus, and sulfur.

biogeography: Study of the physical distribution of organisms.

bioinformatics: Science of managing and analyzing biological data using advanced computing techniques.

biomass: Organic material from living organisms, typically plant matter such as trees, grasses, and agricultural crops, that can be burned or converted to liquid or gaseous fuels for energy.

biome: Large geographic region defined by environmental conditions and biological communities found in the area.

bioreactor: Vessel in which biocatalysts or microorganisms involved in the production of a desired biological product are maintained. In industry, bioreactors typically house fermentation reactions and are called fermenters.

biosequestration: Biologically mediated uptake and conversion of carbon dioxide to inert, long-lived, carbon-containing materials.

biota: Living organisms.

C_3 plant: Plants (e.g., soybean, wheat, and cotton) whose carbon-fixation products have three carbon atoms per molecule. Compared with C_4 plants, C_3 plants show a greater increase in photosynthesis with a doubling of CO_2 concentration and less decrease in stomatal conductance, which results in an increase in leaf-level water use efficiency

carbon cycle: The complex carbon flows and transformations among major Earth system components (atmosphere, oceans, and terrestrial systems). The global flow of carbon from one reservoir (carbon sink) to another. Each carbon exchange among reservoirs is mediated by a variety of physical, biogeochemical, and human activities.

carbon dioxide (CO_2): Colorless, odorless gas that absorbs infrared radiation and traps heat in the Earth's atmosphere. CO_2, which is important to the global carbon cycle, is emitted from a variety of processes (e.g., cellular respiration, biomass decomposition, fossil-fuel use) and taken up primarily by photosynthesis and the oceans via air-sea gas exchange.

carbon fixation: Conversion of inorganic carbon dioxide to organic compounds by photosynthesis.

carbon flux: Rate of carbon movement as it flows from one carbon reservoir to another within an organism, ecosystem, or the global carbon cycle.

carbon partitioning: Partitioning to different parts of a plant (e.g., leaf, stem, root, and seed) versus carbon allocation (partitioning between biomass and respiration).

carbon sequestration: Biological or physical process that captures carbon dioxide and converts it into inert, long-lived, carbon-containing materials.

carbon sink: A pool (reservoir) that absorbs or takes up released carbon from another part of the carbon cycle.

carbon source: A pool (reservoir) that releases carbon to another part of the carbon cycle.

cellulose: Large, complex polysaccharide that is a major component of plant cell walls. Each cellulose molecule is a linear chain of thousands of glucose subunits; multiple cellulose chains form cable-like structures that stabilize the matrix of plant cell-wall materials.

chromatin immunoprecipitation (ChIP): In vivo method that uses antibodies targeted to specific DNA-binding proteins to analyze protein-DNA interactions and determine which sequences these proteins bind and where these proteins bind the genome.

climate model: Mathematical model used to understand, simulate, and predict climate trends by quantitatively analyzing interactions among Earth system components (e.g., land, ocean, atmosphere, and biosphere).

cofactor: Inorganic substance (e.g., metal ion) that is a component of an enzyme complex and required for enzyme activity.

co-immunoprecipitation (Co-IP): Technique that uses antibodies to detect interacting proteins. An antibody that specifically binds a target protein is added to a sample of cellular material. The antibody forms a complex with its target and any protein or molecule bound to the target. Then an antibody-binding protein immobilized on a tiny bead is added and used to pull the antibody-protein complex out of solution.

community: All the different species of organisms living together and interacting in a particular environment.

complexes: Aggregates of multiple, interrelated molecular parts.

contig: Group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome.

curation: A process in which experts manually review, validate, update, and add value to biologically meaningful representations of data, information, and knowledge in computer systems.

cyanobacteria: Division of bacteria capable of oxygen-producing photosynthesis and found in many environments, including oceans, fresh water, and soils. Cyanobacteria contain chlorophyll a and other photosynthetic pigments in an intracellular system of membranes called thylakoids. Many cyanobacterial species also are capable of nitrogen fixation.

data mining: Data analysis techniques used to sift through large amounts of data and identify hidden patterns and relationships.

data model: Logical structure for representing data associated with a particular concept and relating it to other data in a database.

data standard: Set of specifications, established by community consensus or authorized by an official standards organization, for representing and organizing data in ways that promote the exchange, comparison, and integration of different datasets.

DNA (deoxyribonucleic acid): Molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases adenine (A), guanine (G), cytosine (C), and thymine (T).

ecosystem: Set of organisms (plants, animals, fungi, and microorganisms) and the physical and chemical factors that make up a particular environment.

electrophoretic mobility shift assay (EMSA): In vitro method for characterizing the interactions between a protein and DNA or RNA. When a protein binds a labeled piece of DNA or RNA, it forms a large molecular complex that moves more slowly down through a gel than free DNA or RNA molecules. Variations of this basic method can identify the specific DNA or RNA sequence that the protein binds, determine the affinity of the protein for a specific sequence, and reveal which proteins in a mixture bind a particular sequence.

eukaryote: A single-celled or multicellular organism (e.g., plant, animal, or fungi) with a cellular structure that includes a membrane-bound, structurally discrete nucleus and other well-developed subcellular compartments. *See also prokaryote.*

expressed sequence tag (EST): A short segment of DNA sequence derived from the mRNA of transcribed (expressed) genes that can be used to uniquely identify and locate full-length, protein-coding genes within a genome.

expression: *See gene expression.*

fluorescence in situ hybridization (FISH): Technique for microscopic imaging that uses fluorescent probes targeted to signature DNA or RNA sequences to identify and locate different populations in a microbial community without having to grow microbes in culture.

fungi: Eukaryotic, heterotrophic organisms—ranging from single-celled yeasts to multicellular molds and mushrooms—that lack chlorophyll, have rigid cell walls, and absorb nourishment from living or dead organic matter.

gene: Fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides, located in a particular position on a particular chromosome, that encodes a specific functional product (i.e., a protein or RNA molecule).

gene calling: Computational process for identifying where genes begin and end within a genome and for assigning meaningful descriptions to DNA segments recognized as genes.

gene expression: Process by which a gene's coded information is converted into structures present and operating in the cell. Expressed genes include those transcribed into mRNA and then translated into proteins, as well as those transcribed into RNA but not translated into proteins [e.g., transfer (tRNA) and ribosomal RNA (rRNA)].

gene family: Group of closely related genes that make similar products.

gene prediction: Computational method for identifying the locations and sequences of possible genes within a genome. Several gene prediction approaches are based on how well an unknown stretch of DNA sequence matches known gene sequences.

gene product: Protein or RNA molecules resulting from the expression of a gene's DNA sequence. The amount of gene product is used to measure a gene's level of expression.

gene regulatory network: Intracellular network of regulatory proteins that control the expression of gene subsets involved in particular cellular functions. A simple network would consist of one or more input signaling pathways, regulatory proteins that integrate the input signals, several target genes (in bacteria, a target operon), and the RNA and proteins produced from those target genes.

genetic code: Nucleotide sequence, coded in triplets along the mRNA, that determines the sequence of amino acids in a protein product. Each set of three nucleotides (codon) in

a gene specifies a particular amino acid or signals the start or stop of protein synthesis.

genome: All the genetic material in the chromosomes of a particular organism. Most prokaryotes package their entire genome into a single chromosome, while eukaryotes have different numbers of chromosomes. Genome size generally is given as total number of base pairs.

genome sequence: Order of nucleotides within DNA molecules that make up an organism's entire genome.

genomics: Study of genes and their function.

genotype: An organism's genetic constitution, as distinguished from its physical characteristics (phenotype).

gross primary production (GPP): Total amount of organic matter created by photosynthesis.

hemicellulose: Any of several polysaccharides (e.g., xylans, mannans, and galactans) that cross-link and surround cellulose fibers in plant cell walls.

heterotroph: Organism that obtains organic carbon by consuming other organisms or the products of other organisms.

high throughput: Analytical or computational analysis done on a massive, automated scale.

horizontal gene transfer (or lateral gene transfer): Exchange of genetic material between two different organisms (typically different species of prokaryotes). This process gives prokaryotes the ability to obtain novel functionalities or cause dramatic changes in community structure over relatively short periods of time. *See also vertical gene transfer.*

horizontal queries: Queries that associate equivalent data entities across species, samples, or habitats (e.g., homologous genes between species, community composition across samples, and abundance or enrichment of metabolic pathways across habitats).

informatics: Science of managing and analyzing data using advanced computing techniques.

interoperability: Ability of two or more computer systems to work together by exchanging services or communicating, sharing, and interpreting data using common protocols.

isobaric tag for relative and absolute quantitation (iTRAQ): Chemical probe that labels the N-terminus of all peptides in up to eight different biological samples, thus enabling the identification and quantitation of corresponding proteins using mass spectrometry–based proteomic approaches. Samples subjected to different experimental

conditions can be tagged with different iTRAQ labels and then mixed together (multiplexed) to enable simultaneous quantitative analysis.

isotherm: Line on a map indicating points of equal temperature.

isotope-coded affinity tag (ICAT): Chemical probe that labels cysteine residues in proteins, thus enabling the selective isolation and quantitation of particular subsets of proteins using mass spectrometry–based proteomic approaches. Samples subjected to different experimental conditions can be tagged with ICAT labels of different molecular mass and then mixed together to enable comparisons of protein abundance levels in a single analysis step.

knowledgebase: Comprehensive collection of knowledge stored in databases and used to solve problems in a particular subject area such as biology.

latency: Delays that can affect system response time.

lateral gene transfer: *See horizontal gene transfer.*

life-cycle management: End-to-end management of a project.

lignin: Complex, insoluble polymer whose structure surrounds and gives strength and rigidity to cellulose fibers in the cell walls of woody plants. Lignin makes up a significant portion of the mass of dry wood and, after cellulose, is the second most abundant form of organic carbon in the biosphere.

lignocellulose: Refers to plant cell-wall materials primarily made up of lignin, cellulose, and hemicellulose.

LIMS: Acronym for laboratory information management system, which is a computerized system used by laboratories to track samples; automate data capture from laboratory instruments; and facilitate storage, presentation, and sharing of data among collaborating researchers.

machine reasoning: Ability of a computer to make selections or solve problems using approaches that model human reasoning and learning.

mass spectrometry: Method involving specialized instruments for measuring the mass and abundance of molecules in a mixture and identifying mixture components by mass and charge.

messenger RNA (mRNA): RNA that serves as a template for protein synthesis. *See also transcription and translation.*

metabolic flux analysis: System-level understanding and quantitation of the flow of molecules through metabolic networks.

metabolism: Collection of all biochemical reactions that an organism uses to obtain the energy and materials it needs to sustain life. An organism uses energy and common biochemical intermediates released from the breakdown of nutrients to drive the synthesis of biological molecules.

metabolites: Small molecules (<500 Da) that are the substrates, intermediates, and products of enzyme-catalyzed metabolic reactions.

metabolomics: Type of global molecular analysis that involves identifying and quantifying the metabolome—all metabolites present in a cell at a given time.

metadata: Data that describe specific characteristics and usage aspects of other data (e.g., what data are about, when and how data were created, who can access the data, and the formats available).

metagenomics: Study of the collective DNA isolated directly from a community of organisms living in a particular environment.

metaomics: High-throughput, global analysis of DNA, RNA, proteins, or metabolites isolated directly from a community of organisms living in a particular environment.

metaproteomics: High-throughput, global analysis of proteins isolated directly from a community of organisms living in a particular environment. Metaproteomics can reveal which genes are actively translated into functional proteins by a community.

metatranscriptomics: High-throughput, global analysis of RNA isolated directly from a community of organisms living in a particular environment. Metatranscriptomics can reveal which genes are actively expressed by a community.

microalgae: Microscopic, unicellular aquatic plants.

microarray: Analytical technique used to measure the mRNA abundance (gene expression) of thousands of genes in one experiment. The most common type of microarray is a glass slide onto which DNA fragments are chemically attached in an ordered pattern. As fluorescently labeled nucleic acids from a sample are applied to the microarray, they bind the immobilized DNA fragments and generate a fluorescent signal indicating the relative abundance of each nucleic acid in the sample.

microbiome: A community of microorganisms that inhabits a particular environment. For example, a plant microbiome includes all the microorganisms that colonize a plant's surfaces and internal passages.

microorganism: Sometimes called a microbe, this is any microscopic prokaryotic or eukaryotic organism, including bacteria, archaea, and protists.

model: Mathematical representation used in computer simulations to calculate the evolving state of dynamic systems.

model ecosystem: A specific type of ecosystem that is widely studied in great detail by a community of researchers to provide insights into the processes controlling the behavior of other ecosystems.

modeling: Use of statistical and computational techniques to create working computer-based models of biological phenomena that can help to formulate hypotheses for experimentation and predict outcomes of research.

molecular machine: Highly organized assembly of proteins and other molecules that work together as a functional unit to carry out operational, structural, and regulatory activities in cells.

motif: A sequence motif is a short, recurring pattern of nucleotides (in DNA or RNA) or amino acids (in proteins) that can signal a particular function or molecular event (e.g., a sequence where a protein binds). A structural motif is a recurring, three-dimensional arrangement of structural elements observed in different proteins.

net primary production (NPP): Fraction of photosynthetically fixed organic matter that remains after accounting for carbon lost to cellular respiration and other biological processes.

nitrogen fixation: Process carried out by certain species of bacteria and archaea in which atmospheric nitrogen (N_2) is converted to organic nitrogen-containing compounds that can be used by other organisms.

noncoding RNA (ncRNA): Any RNA molecule that does not serve as a template for protein synthesis.

nonprocedural relational operators: Programming language constructs that are used to compare and test the relationship between two values or entities. With nonprocedural relational operators, the user specifies what output is needed but does not specify the procedure to obtain the output.

object-relational system: System that combines object-related database concepts with relational databases.

omics: Collective term for a range of new high-throughput biological research methods (e.g., transcriptomics, proteomics, and metabolomics) that systematically investigate entire networks of genes, proteins, and metabolites within cells.

ontology: Organized, hierarchical structure of concepts relevant to a particular knowledge domain. An ontology identifies which of several equivalent terms should be used to represent a concept and defines how different terms and concepts are related. Ontologies are developed to ensure the consistent use of language across multiple databases and information systems.

operon: In prokaryotic genomes, a linear group of genes transcribed together on the same mRNA molecule and controlled by the same regulatory element.

organelle: Specialized structure within a cell with a specific function.

parallelization: Simultaneous use of multiple computers to carry out a particular task or solve different parts of the same problem.

parameterization: In climate modeling, approach used to represent phenomena that are too small-scale or complex to be included in a model.

pathway: Series of molecular interactions that occur in a specific sequence to carry out a particular cellular process (e.g., sense a signal from the environment, convert sunlight to chemical energy, break down or harvest energy from a carbohydrate, synthesize ATP, or construct a molecular machine).

phenology: Study of recurring biological phenomena (e.g., seasonal leaf loss in trees) and how changes in climate or the surrounding environment can impact the timing of these periodic events.

phenomics: Collective study of multiple phenotypes (e.g., all phenotypes associated with a particular biological function).

phenotype: Physical characteristics of an organism.

photosynthate: Organic carbon produced by photosynthesis.

photosynthesis: Process by which plants, algae, and certain types of prokaryotic organisms capture light energy and use it to drive the transfer of electrons from inorganic donors (e.g., water) to carbon dioxide to produce energy-rich carbohydrates.

photosystem: Large, membrane-bound molecular complex consisting of multiple proteins containing pigment molecules (e.g., chlorophylls) that absorb light at a particular wavelength and transfer the energy from the absorbed photon to a reaction center that initiates a series of electron-transport reactions.

phototroph: Organism capable of photosynthesis.

phylogeny: Study of the relatedness and evolutionary relationships among different groups of organisms.

phytoplankton: Free-floating, microscopic photosynthetic organisms (e.g., algae, cyanobacteria, and dinoflagellates) found in the surface waters of marine and freshwater environments.

post-translational modification: Any of several chemical modifications (e.g., phosphorylation, disulfide bond formation, cleavage of inactive sequence) involved in converting a newly translated amino acid sequence into a functional protein.

post-translational regulation: Process that controls the expression of gene products in cells by influencing the conversion of a newly translated amino acid sequence into a functional protein.

***Prochlorococcus*:** Type of unicellular cyanobacterium that is an extremely abundant primary producer in the world's oceans. *Prochlorococcus* is the smallest known oxygenic phototroph. Its abundance and phototrophic metabolism make it important in global carbon cycling through CO_2 fixation.

prokaryote: Single-celled organism lacking a membrane-bound, structurally discrete nucleus and other subcellular compartments. Bacteria and archaea are prokaryotes. *See also eukaryote.*

protein: Large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nucleotides in the gene that codes for the protein. Proteins maintain distinct cell structure, function, and regulation.

protein complex: Aggregate structure consisting of multiple protein molecules.

proteome: Collection of proteins expressed by a cell at a particular time and under specific conditions.

proteomics: Large-scale analysis of the proteome to identify which proteins are expressed by an organism under certain conditions. Proteomics provides insights into protein function, modification, regulation, and interaction.

protists: Microscopic, eukaryotic organisms that have simple cellular organization. Protists include plant-, animal-, and fungus-like organisms that range in function from photosynthetic primary producers (e.g., green algae and diatoms) to predators and parasites.

protozoa: Single-celled, eukaryotic microorganisms that use cellular appendages called flagella to propel them through their environments.

provenance data: Data describing all the details of the experiment environment (e.g., manipulation of samples; software, tools, and methods used to conduct the experiment) so that researchers can visualize the experimental process and potentially reproduce the results of a specific experiment.

quality assurance (QA): Approach used to ensure that data systems will perform to a required standard for quality.

quality control (QC): Methods used to determine if the products of a process meet or exceed a defined standard for quality.

quantitative trait loci: All the DNA regions within a genome associated with the different genes that influence a particular complex trait.

recalcitrance: Natural resistance of plant cell-wall materials to physical and biological deconstruction.

regulator: Protein (e.g., a repressor) that controls the expression or activity of other molecules in a cell.

regulatory elements: Segments of the genome (e.g., regulatory regions, genes that encode regulatory proteins or small RNAs) involved in controlling gene expression.

regulatory circuit: *See gene regulatory network.*

regulatory region or sequence: Segment of DNA sequence to which a regulatory protein binds to control the expression of a gene or group of genes that are expressed together.

regulon: Set of operons controlled by the same regulator. Operons belonging to the same regulon can be located in different regions of a genome.

respiration: Series of biochemical redox reactions in which the energy released from the oxidation of organic or inorganic compounds is used to generate cellular energy in the form of ATP.

ribosomal RNA (rRNA): Specialized RNA found in the catalytic core of the ribosome, a molecular machine that synthesizes proteins in all organisms.

1010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101

RNA (ribonucleic acid): Molecule that plays an important role in protein synthesis and other chemical activities of the cell. RNA's structure is similar to that of DNA. Classes of RNA molecules include messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and other small RNAs, each serving a different purpose.

Robetta server: Source of automated tools for analyzing and predicting protein structures.

RuBisCo (Ribulose-1,5-bisphosphate carboxylase/oxygenase): Enzyme that catalyzes the first major step of photosynthetic carbon fixation by adding a molecule of carbon dioxide to a short 5-carbon sugar called ribulose bisphosphate. The resulting 6-carbon sugar is split into two 3-carbon molecules that can be used to build larger sugar molecules. RuBisCo also catalyzes photorespiration, which releases CO_2.

scalability: Ability of a computer system to respond to increased demands.

schema: Description of the structure and organization of all the elements of a database.

semantic Web technologies: Technologies based on a common set of design principles that improve the efficiency of searching and sharing information on the Web by making Web content, which is designed to be read by humans, computer readable.

shotgun sequencing: Common approach to sequencing microbial genomes that involves breaking the genome into random fragments, which are cloned into vectors and sequenced. Computational analysis is used to compare all DNA sequence reads from random fragments and assemble the entire genome by aligning overlapping sequences.

signal-transduction pathway: Series of biochemical reactions that receive extracellular chemical signals. These signals are transmitted and amplified within the cell and ultimately used to stimulate or repress a certain type of molecular activity (e.g., gene expression).

simulation: Combination of multiple models into a meaningful representation of a whole system that can be used to predict how the system will behave under various conditions. Simulations can be used to run in silico experiments to gain first insights, form hypotheses, and predict outcomes before conducting more expensive physical experiments.

single nucleotide polymorphisms (SNPs): DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered.

stimulon: Set of genes controlled by the same stimulus.

synthetic biology: Field of study that aims to build novel biological systems designed to carry out particular functions by combining different biological "parts" or molecular assemblies.

system architecture: Conceptual design depicting how data and services are partitioned and linked among the different components of interconnected database systems.

systems biology: Use of global molecular analyses (e.g., measurements of all genes and proteins expressed in a cell at a particular time) and advanced computational methods to study how networks of interacting biological components determine the properties and activities of living systems.

systems microbiology: Systems biology approach that focuses on understanding and modeling microorganisms at molecular, cellular, and community levels.

taxa: Categories (e.g., phylum, order, family, genus, or species) used to classify animals and plants (singular: taxon).

taxonomy: Hierarchical classification system for naming and grouping organisms based on evolutionary relationships.

terabyte: Unit of computer storage representing one trillion (or 10^{12}) bytes.

transcript: Messenger RNA molecule (mRNA) generated from a gene's DNA sequence during transcription.

transcription: Synthesis of an RNA copy of a gene's DNA sequence; the first step in gene expression. *See also translation.*

transcription factor: Protein that binds to regulatory regions in the genome and helps control gene expression.

transcription start site (TSS): Position within the DNA sequence of a gene where the enzyme RNA polymerase initiates synthesis of mRNA.

transcriptomics: Global analysis of expression levels of all RNA transcripts present in a cell at a given time.

transfer RNA (tRNA): A class of small RNA molecules that have triplet nucleotide sequences that are complementary to the triplet nucleotide coding sequences of mRNA. During protein synthesis, each tRNA bonds with a particular amino acid that is added to the growing amino acid chain as specified by the order of nucleotides in the mRNA.

translation: Process in which the genetic code carried by mRNA directs the synthesis of proteins from amino acids.

transporter: Protein that transports a molecule from one location to another; in most cases, transporters are membrane proteins that control the movement of molecules in and out of cells.

vertical gene transfer: Inheritance or passing of genetic material from one generation to another. *See also horizontal gene transfer.*

vertical queries: Queries that span multiple data levels (e.g., from correlating climate data and habitats to genes found in different samples).

visualization: Representation of data using images that add meaning and facilitate user access, navigation, and retrieval of data.

wiki ("what I know is"): A method for a community to collectively accumulate knowledge.

xylem: Water-carrying tissue in vascular plants that gives stalks and stems rigidity. Xylem is a major component of wood where the cells of this tissue have thick, lignin-rich walls.

yeast two-hybrid (Y2H): Method for studying and identifying novel interactions between a protein of interest and other proteins. The protein of interest is fused to one of two domains of a transcription-activating molecule. The second domain is fused to potential binding partners of the protein of interest. When the protein of interest interacts with its binding partner, the two domains of the transcription-activating molecule come together and initiate the expression of a reporter enzyme that carries out some characteristic functionality (e.g., confers antibiotic resistance, produces a blue color).