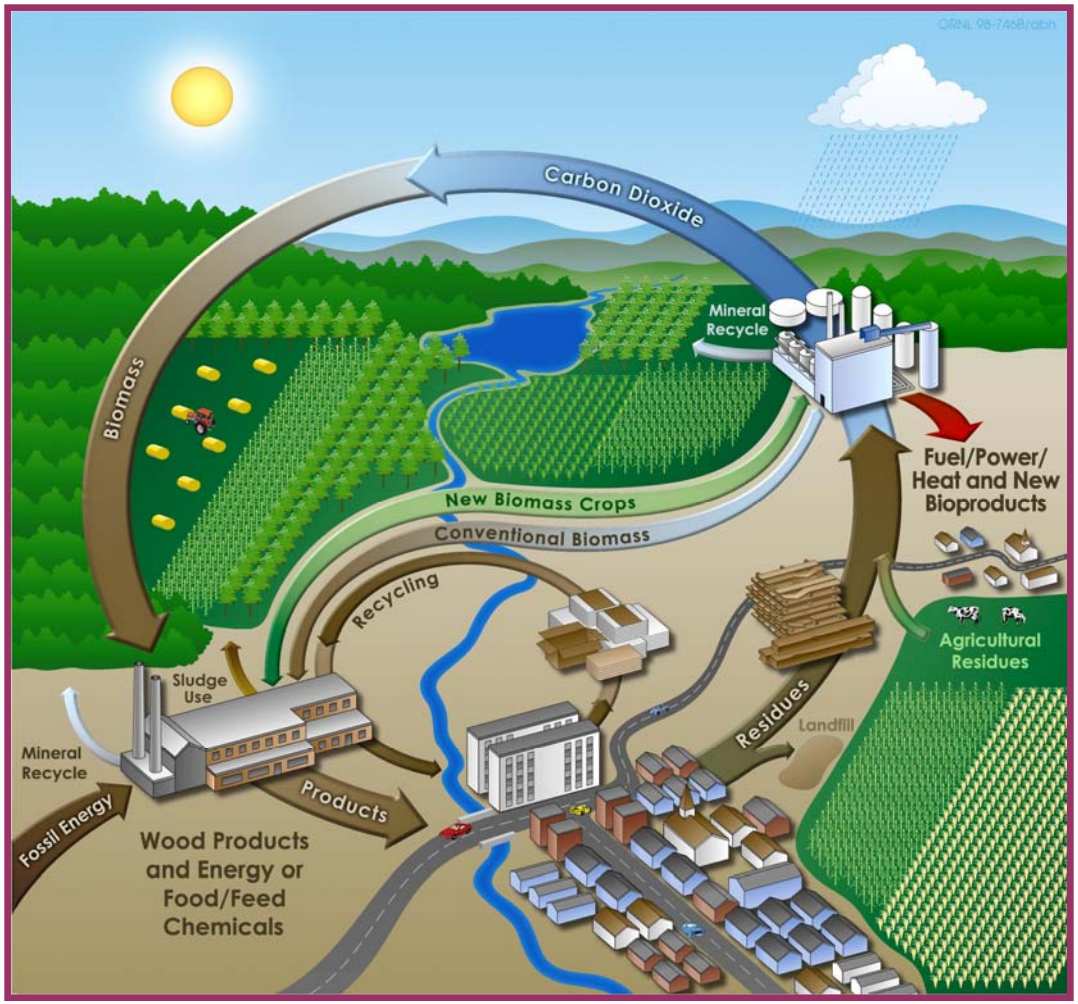
The Role of Systems Biology in Bioenergy at ORNL

Brian Davison

Chief Scientist, Systems Biology and Biotechnology

Oak Ridge National Laboratory

March 2007



President's State of the Union Address January 2006

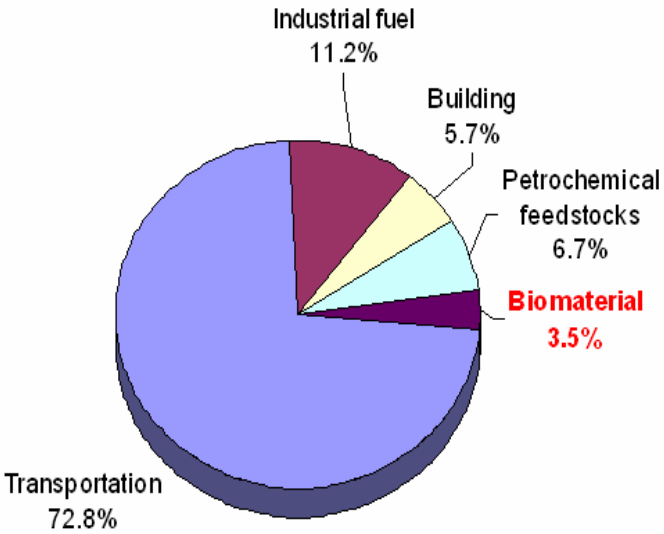
Keeping America competitive requires affordable energy. And here we have a serious problem: America is addicted to oil, which is often imported from unstable parts of the world.

The best way to break this addiction is through technology…. and we are on the threshold of incredible advances…

So tonight I announce…push for breakthroughs in two vital areas…change how we power our homes and offices,…change how we power our automobiles.

The “30 by 30” goal

- US petroleum consumption
  - 329 billion gallons/year
  - >1000 gallons/year per person
- Goal set by Congress: Replace 30% gas and diesel consumption with biofuels by the 2030
- Requires approximately one billion dry tons of biomass for 60 billion gallons using current technology



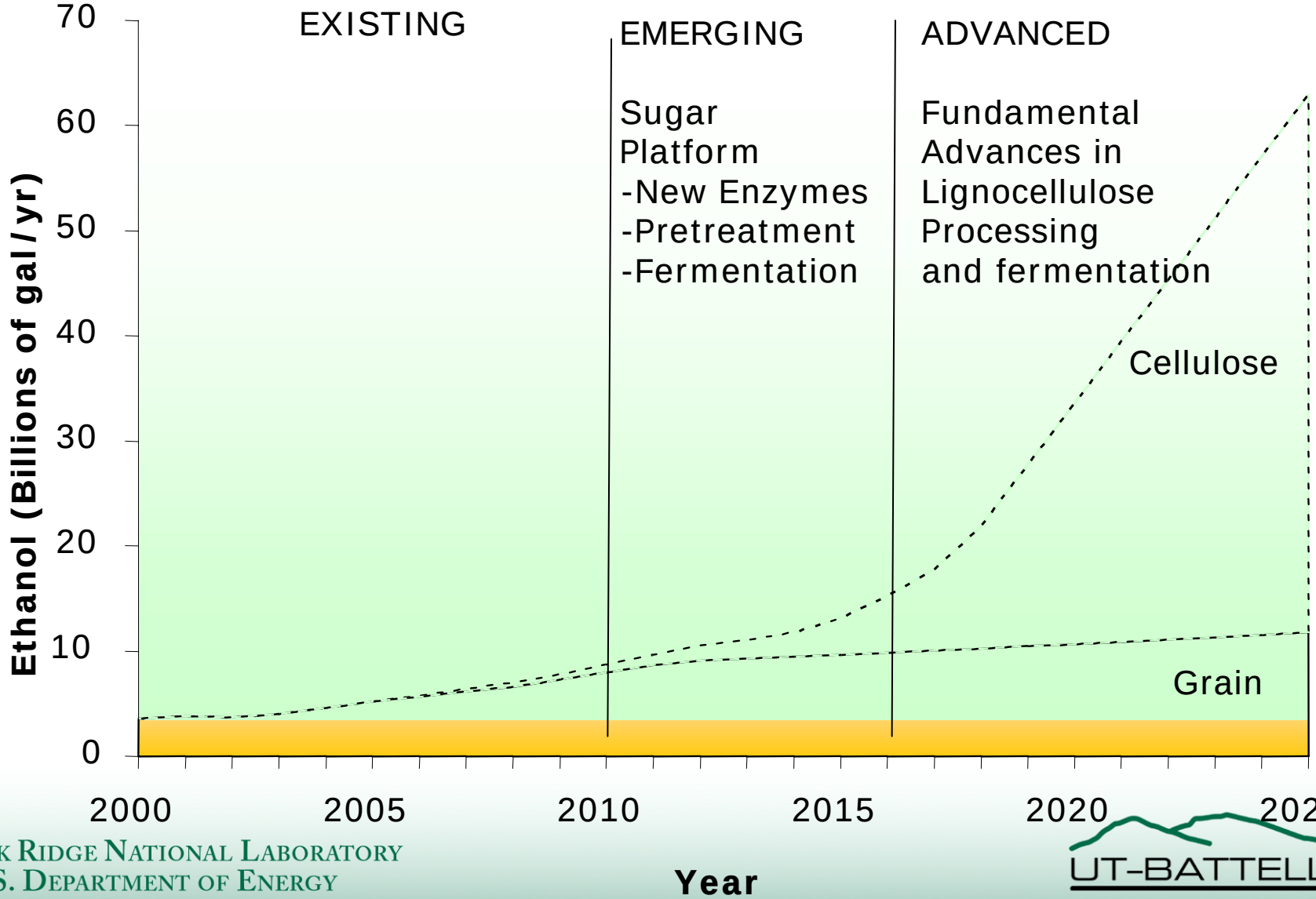
Predicted Petroleum Use in 2025-2030

Source: Winning the Oil Endgame, RMI 2005, and *NRC 2000

State of the Union 2007

- Decrease U.S. petroleum consumption by 20% in 10 years —“20 in 10”
- Grow production of renewable fuels to 35 billion gallons per year

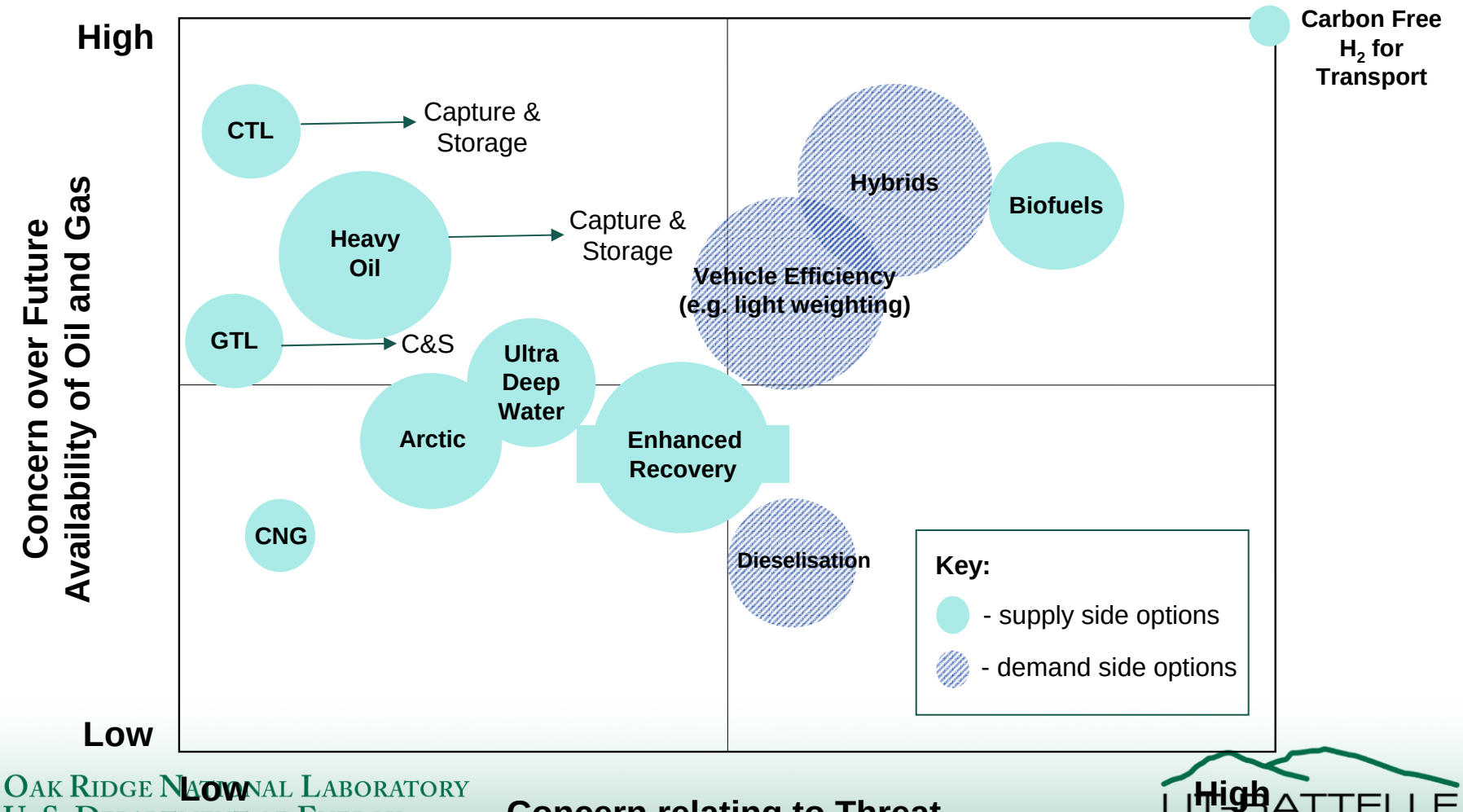
The DOE Biofuels Vision



Modified from Richard Bain, NREL

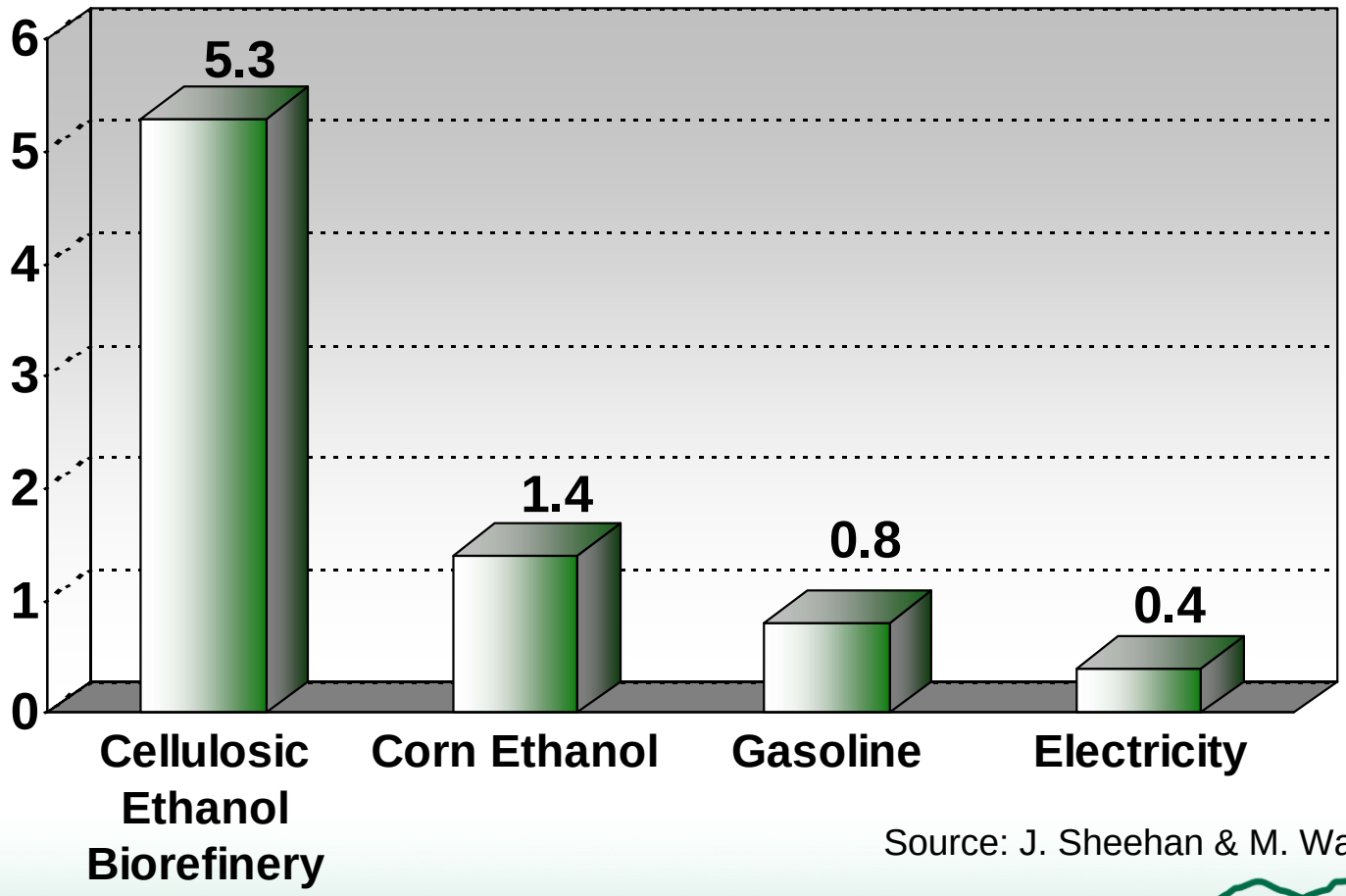
Technology Options for Transportation – from BP

transport sector



Fossil Energy Replacement Ratio

$$\text{Fossil Energy Ratio (FER)} = \frac{\text{Energy Delivered to Customer}}{\text{Fossil Energy Used}}$$



Source: J. Sheehan & M. Wang (2003)

Corn ethanol: Issues and facts

- New corn dry mill plant for ethanol
  - Average cost to build: $100 million
  - Production: ~ 70 million gal of ethanol from 26 million bushels corn
  - Production cost: 60 cents/gal
- TN corn output: 86 million bushels (of 12 billion bushels in U.S.)
- U.S. production of ethanol from corn: ~4.5 billion gallons in 2005; additional 2.5 billion gallon capacity planned by 2008
- Renewable Fuel Standard demand: 7.5 billion gallons to replace MTBE (methyl tertiary butyl ether) in gasoline by 2010
- Capacity to grow corn limits annual ethanol production to around 15-18 billion gallons/year without impacting food and feed

Corn production in Southeast (2005, MM bu)

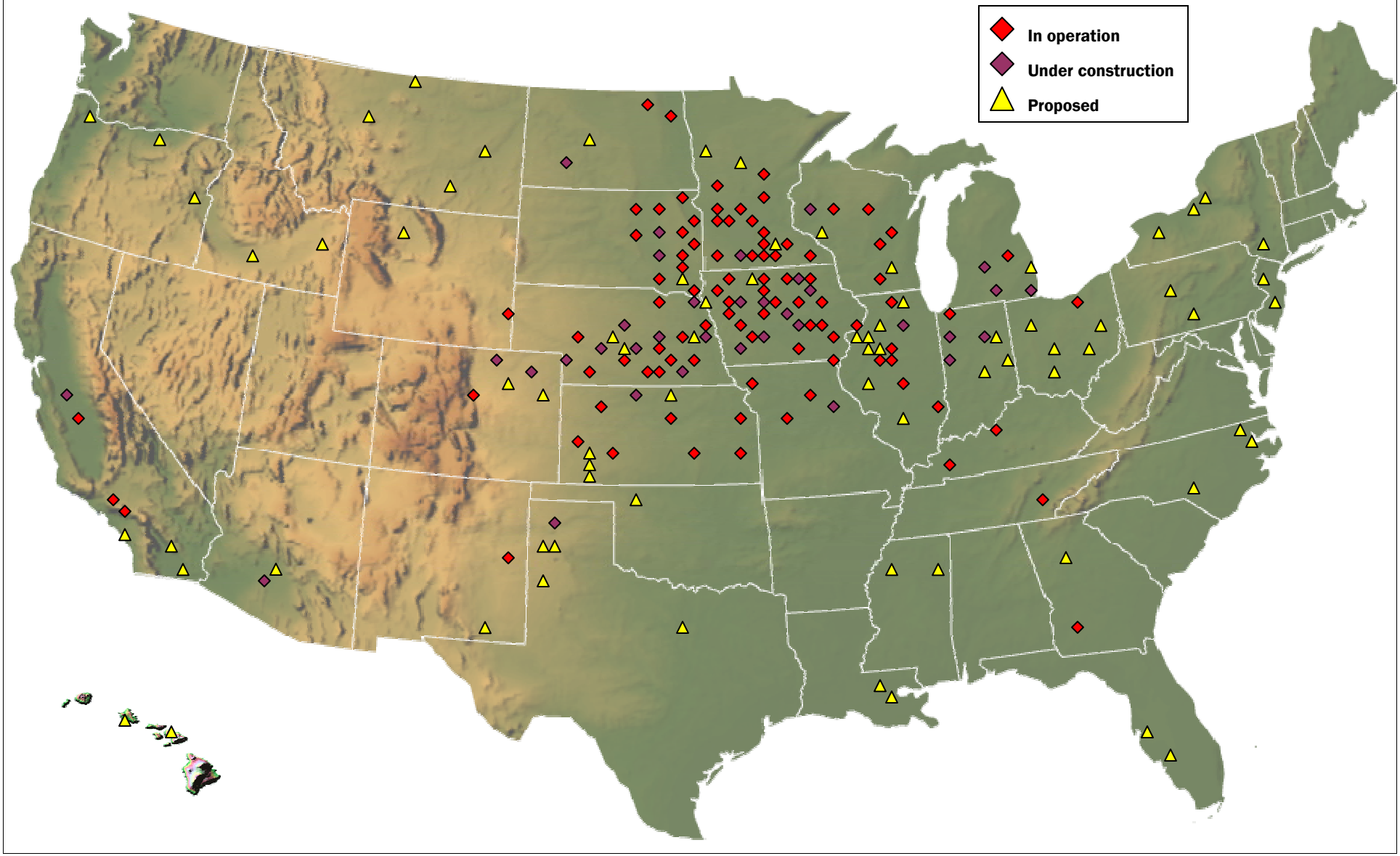
Alabama	24
Arkansas	30
Florida	3
Georgia	30
Kentucky	156
Louisiana	45
Mississippi	47
Missouri	330
North Carolina	84
South Carolina	33
Tennessee	77
Virginia	42
West Virginia	3
Total SE	904
Total US	*11,112*

Ref: NCGA World of Corn, 2005

NCGA
www.ncga.com

US Ethanol Plants

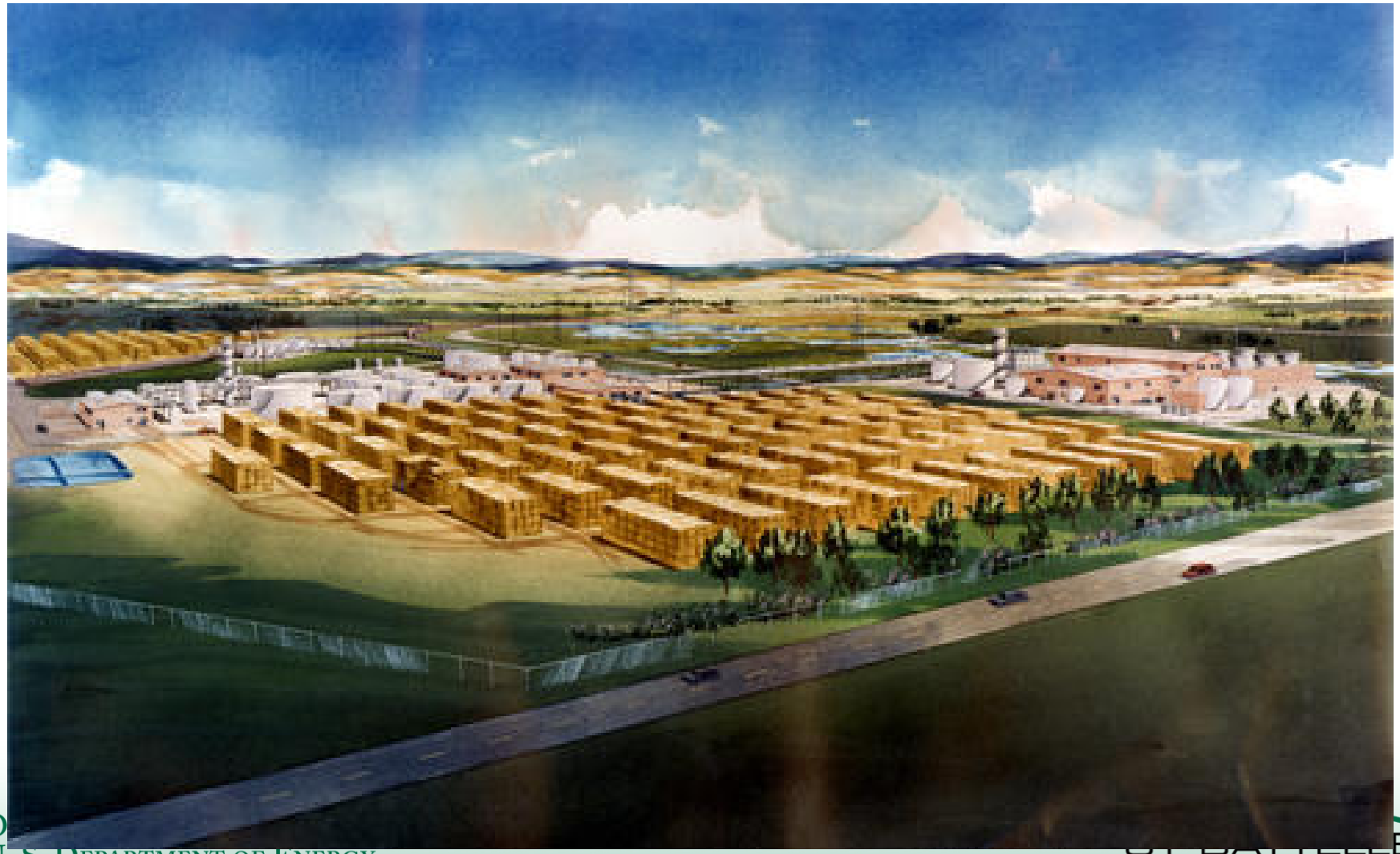
AS OF: March 2006



The business of cellulosic ethanol

- At >$2/gallon, cellulosic ethanol feasible, not economical
- R&D is expected to bring cost down to <$1/gallon
- Sources of cellulose more abundant supply than corn, better suited for Tennessee
- Cellulose-based ethanol delivers 5x more energy per fossil energy unit than corn
- Estimated construction cost to build production plant with comparable 70 million gallon/year capacity:
  - Today's cost: ~$200M
  - With technology improvements: $100M, comparable to corn plant
  - Several hundred tons of biomass processed daily
- Huge opportunity in the Southeast to have alternate use for forest products

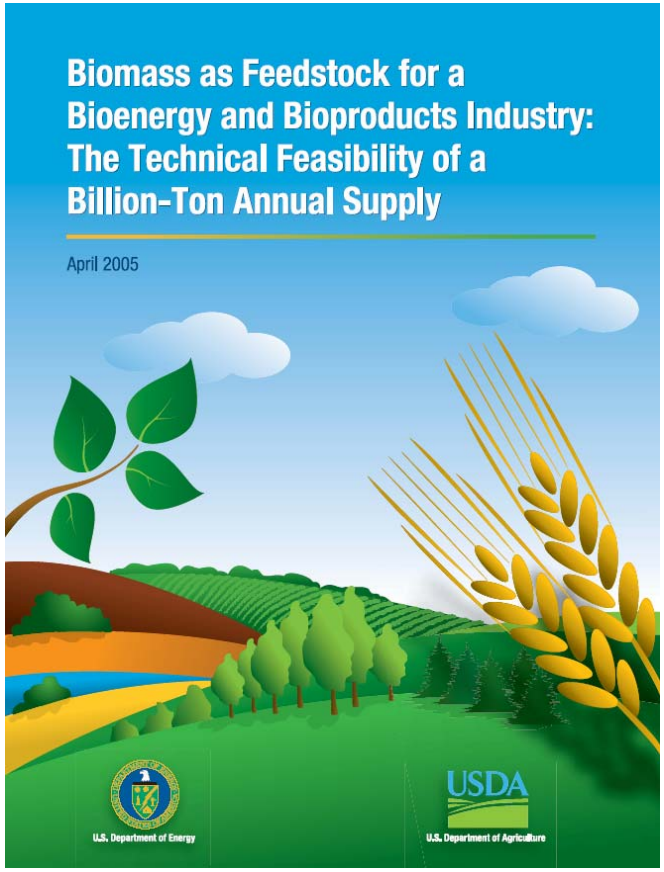
Artist vision of a biorefinery with biomass storage adjacent



Source: NREL Pix

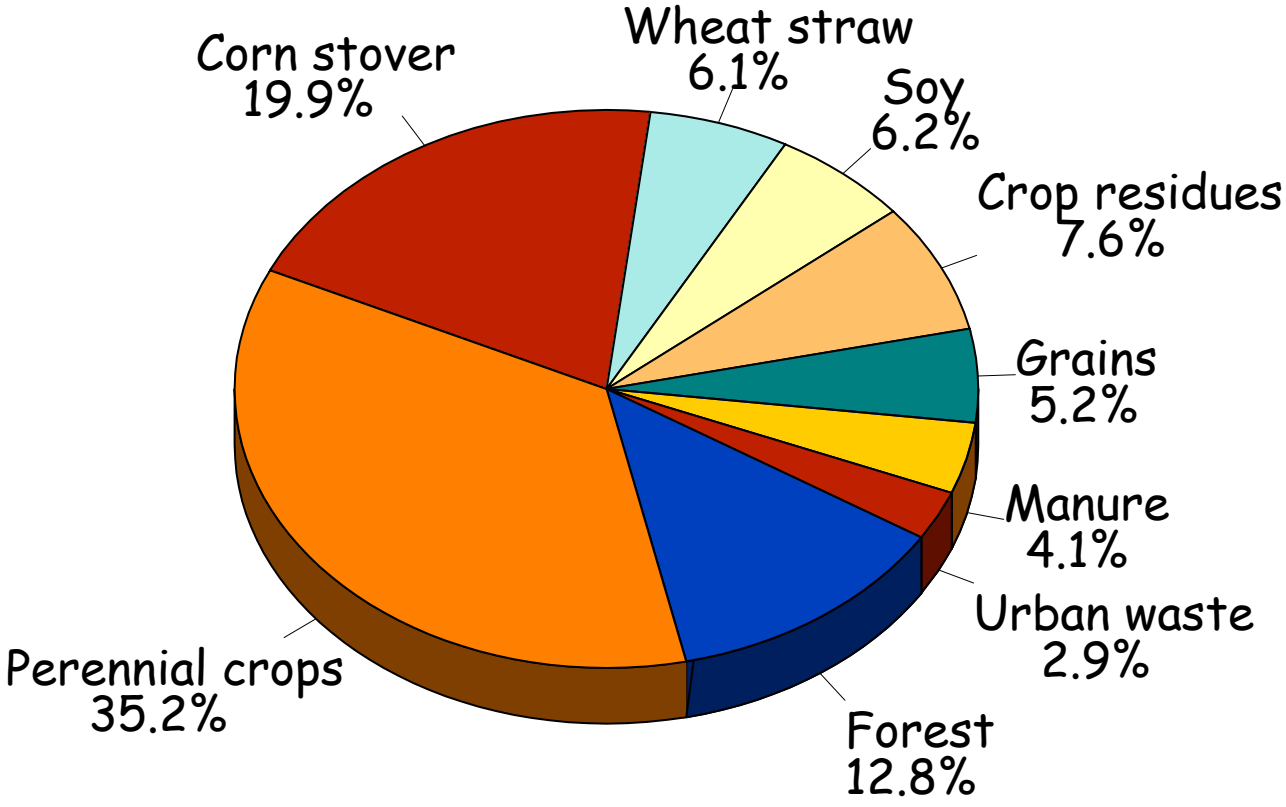
Are there sufficient amounts of biomass?

- **Yes, land resources of the U.S. can sustainably supply more than 1.3 billion dry tons annually and still continue to meet food, feed, and export demands**
- **Required changes are not unreasonable given current trends and time-frame for bio-industry scale-up and deployment**



Robert Perlack, Lynn Wright
Anthony Turhollow, Robin Graham (ORNL)
Bryce Stokes, Donald Erbach (USDA)

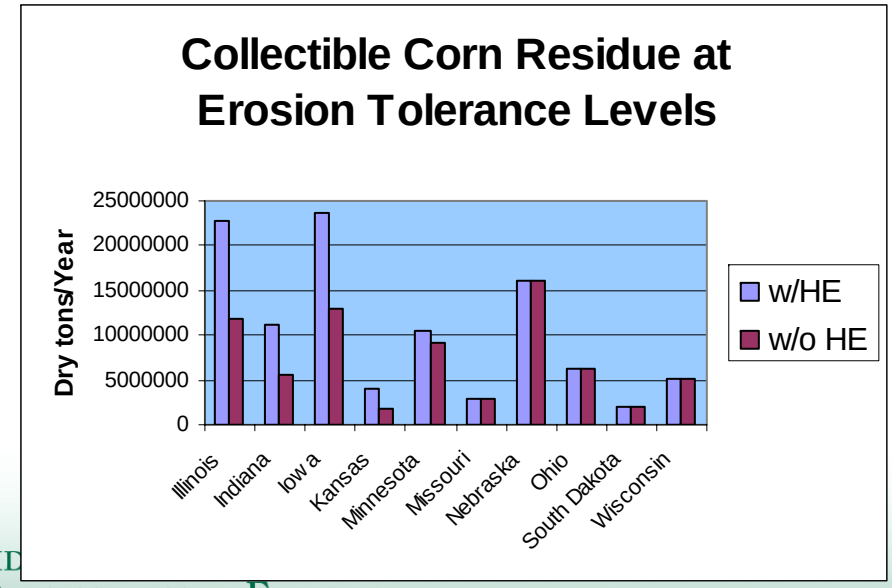
Potential U.S. biomass inventory = 1.3 billion tons

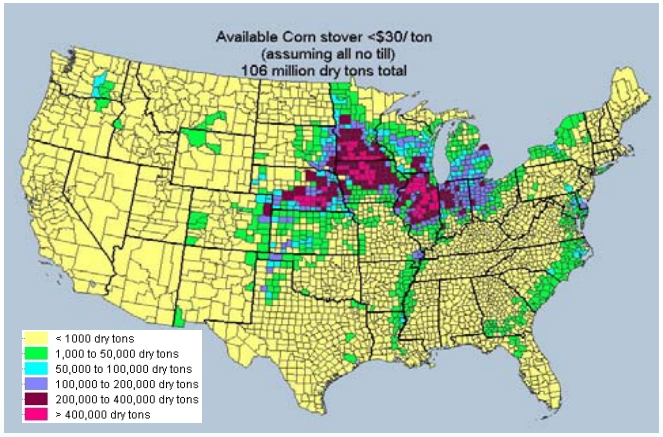


From: R. D. Perlack *et al.*, *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply* (DOE/GO-102005-2135, ORNL, 2005).

Biomass analysis capabilities at ORNL

- Biofeedstock supply logistics
- Agricultural residue sustainability
- Biomass Supply Forecasting analysis





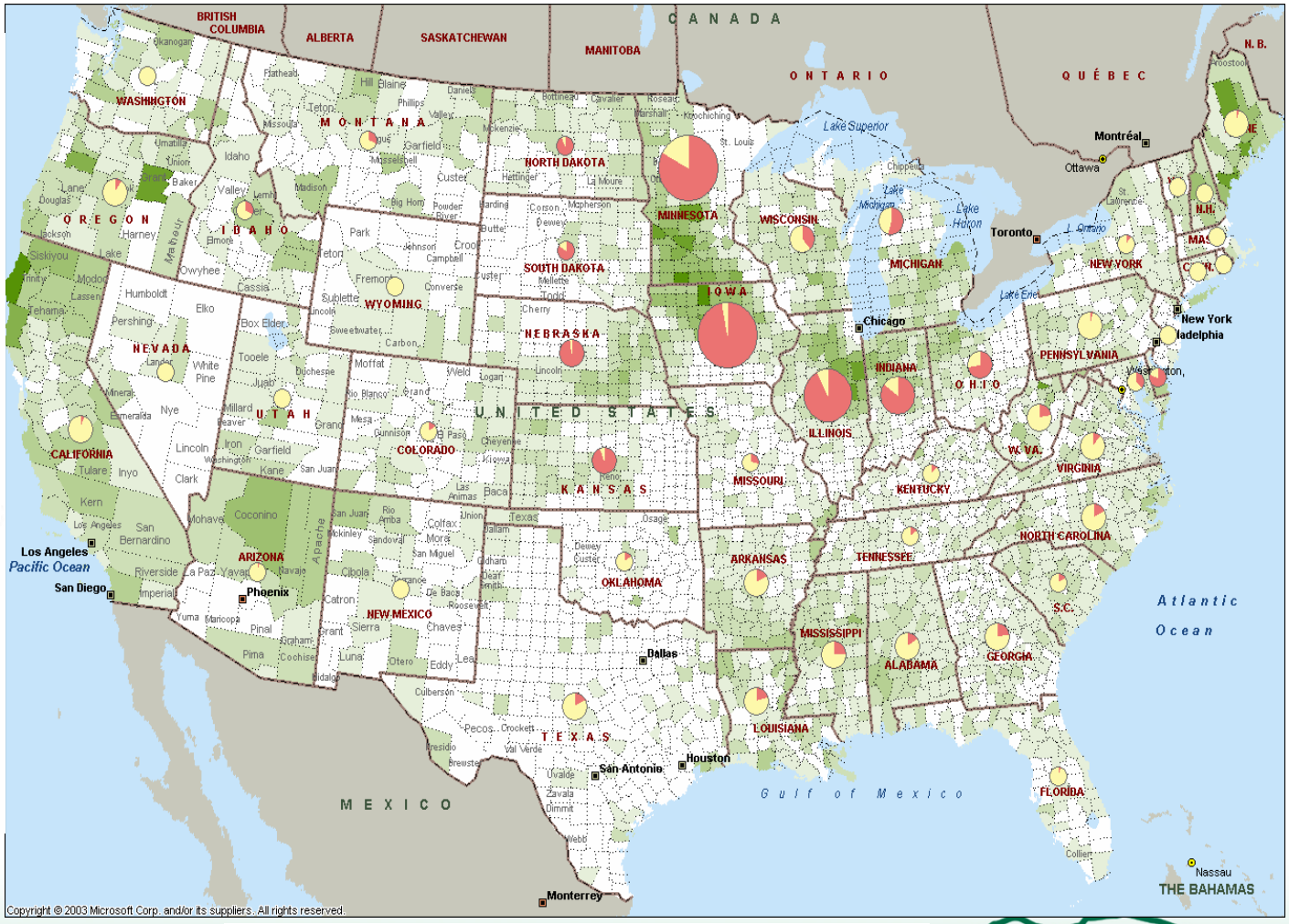
POTENTIAL BIOMASS RESOURCE AND REFINERY CAPACITY IN 2012

Total Number of 1 Million Dry Ton Capacity Refineries

- 14.4 - 18.0
- 10.8 - 14.3
- 7.2 - 10.7
- 3.6 - 7.1
- 0.0 - 3.5
- Biochemical Conversion Potential
- Thermochemical Conversion Potential

Total Residue in Dry Tons

- 750,001 to 900,000
- 600,001 to 750,000
- 450,001 to 600,000
- 300,001 to 450,000
- 150,001 to 300,000
- 75,001 to 150,000
- 25,001 to 75,000
- 0 to 25,000



Logging residues, crop residues, switchgrass

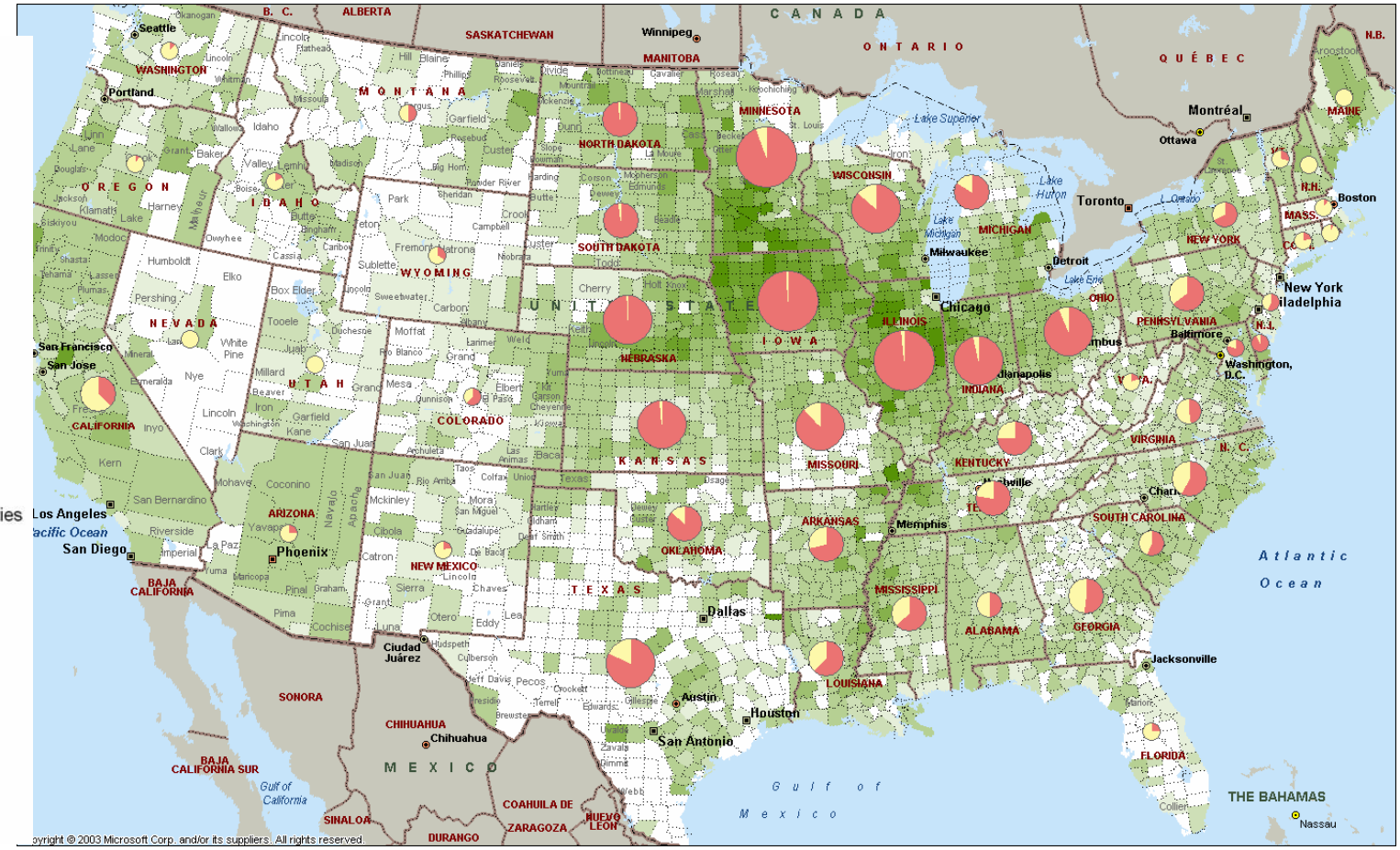
POTENTIAL BIOMASS RESOURCE AND REFINERY CAPACITY IN 2030

1-Million Ton Refineries by State

- 50 - 90
- 20 - 49
- 10 - 19
- 5 - 9
- 0 - 4
- Biochemical Refineries
- Thermochemical Refineries

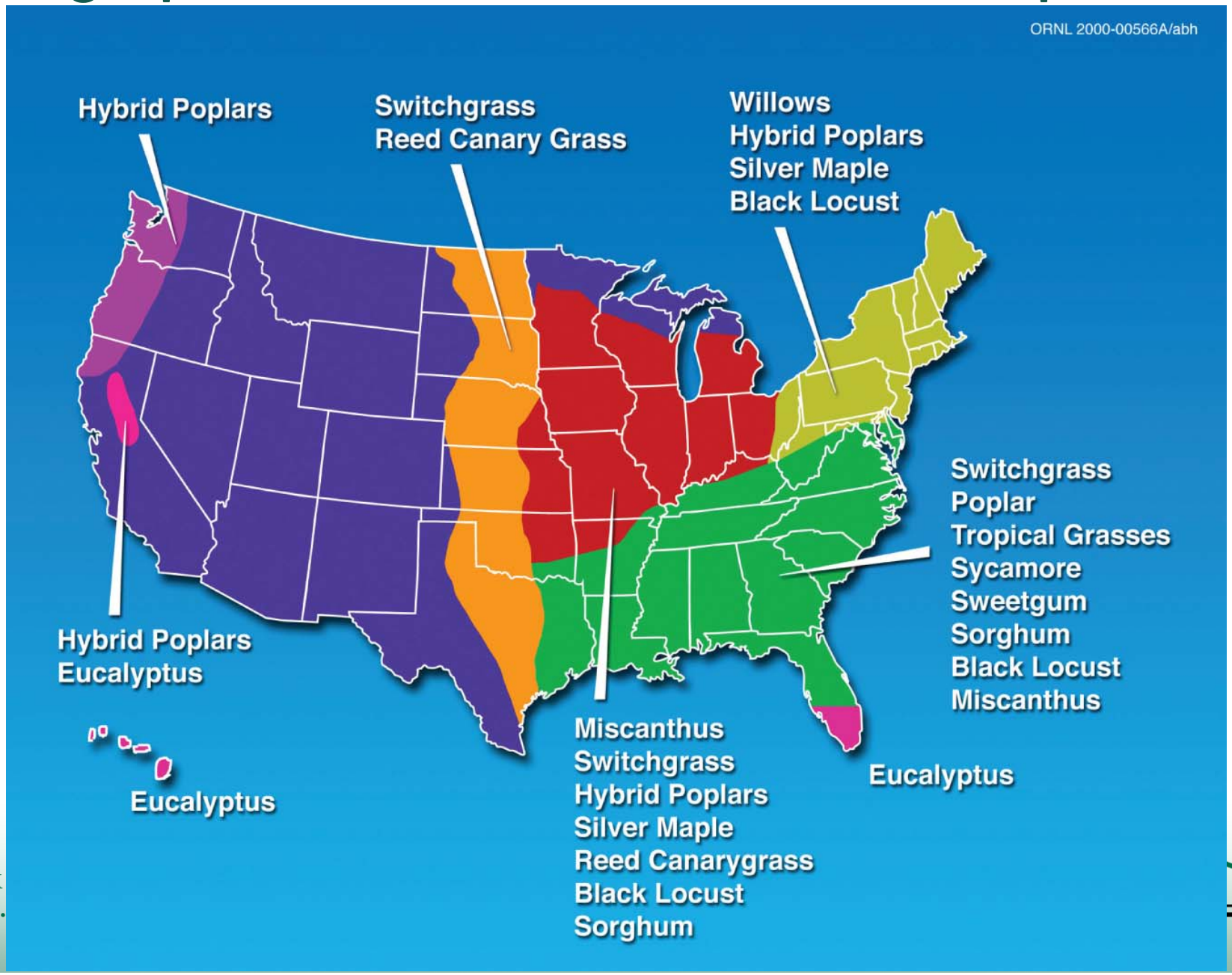
Biomass Total by County

- 1,500,001 to 2,500,000
- 1,000,001 to 1,500,000
- 700,001 to 1,000,000
- 400,001 to 700,000
- 100,001 to 400,000
- 50,001 to 100,000
- 25,001 to 50,000
- 0 to 25,000

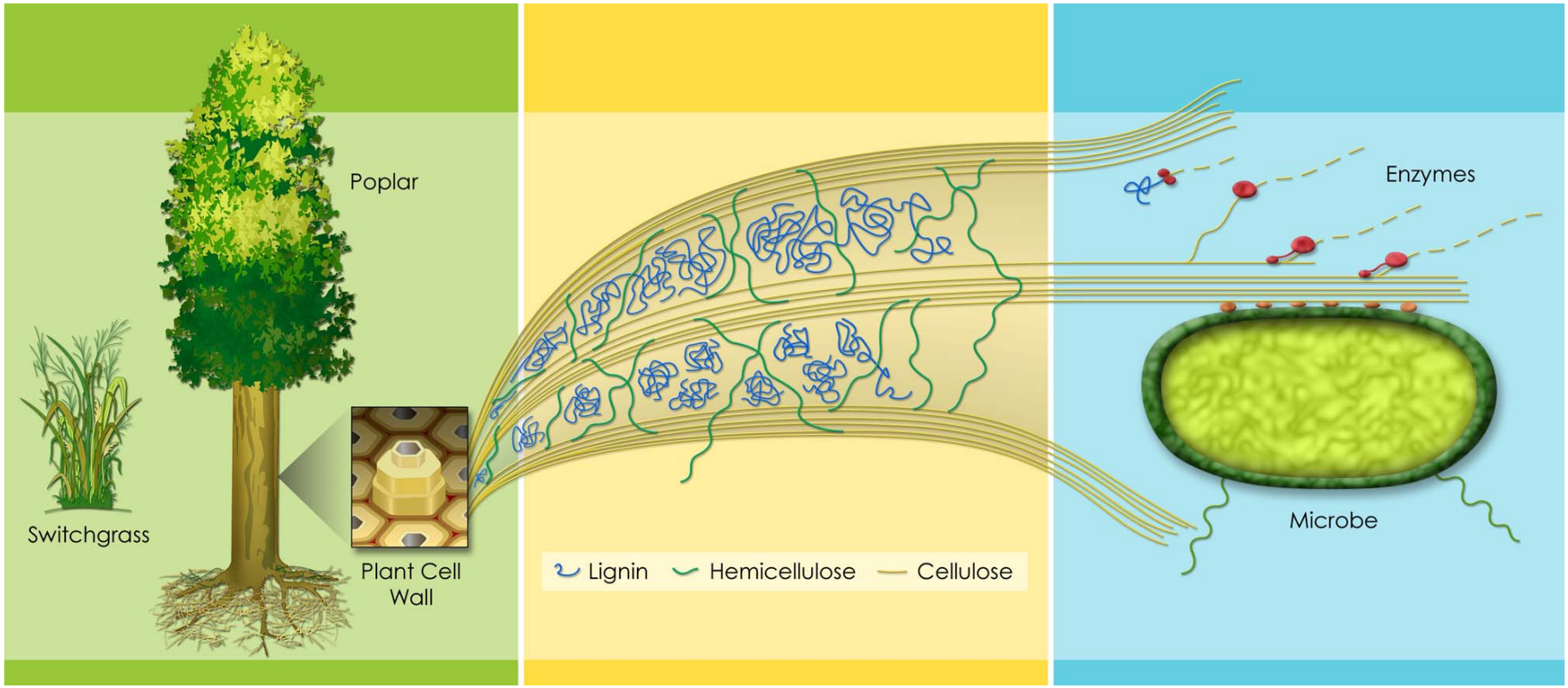


Forest residues (all), crop residues, switchgrass

Geographic distribution of biomass crops



Making fuel from plant biomass



The challenges: Lignocellulosic biomass is complex and heterogeneous

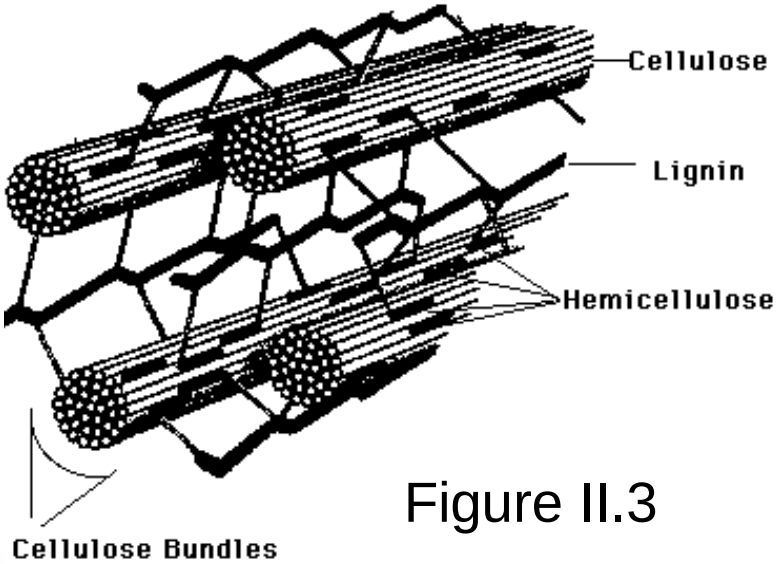


Figure II.3

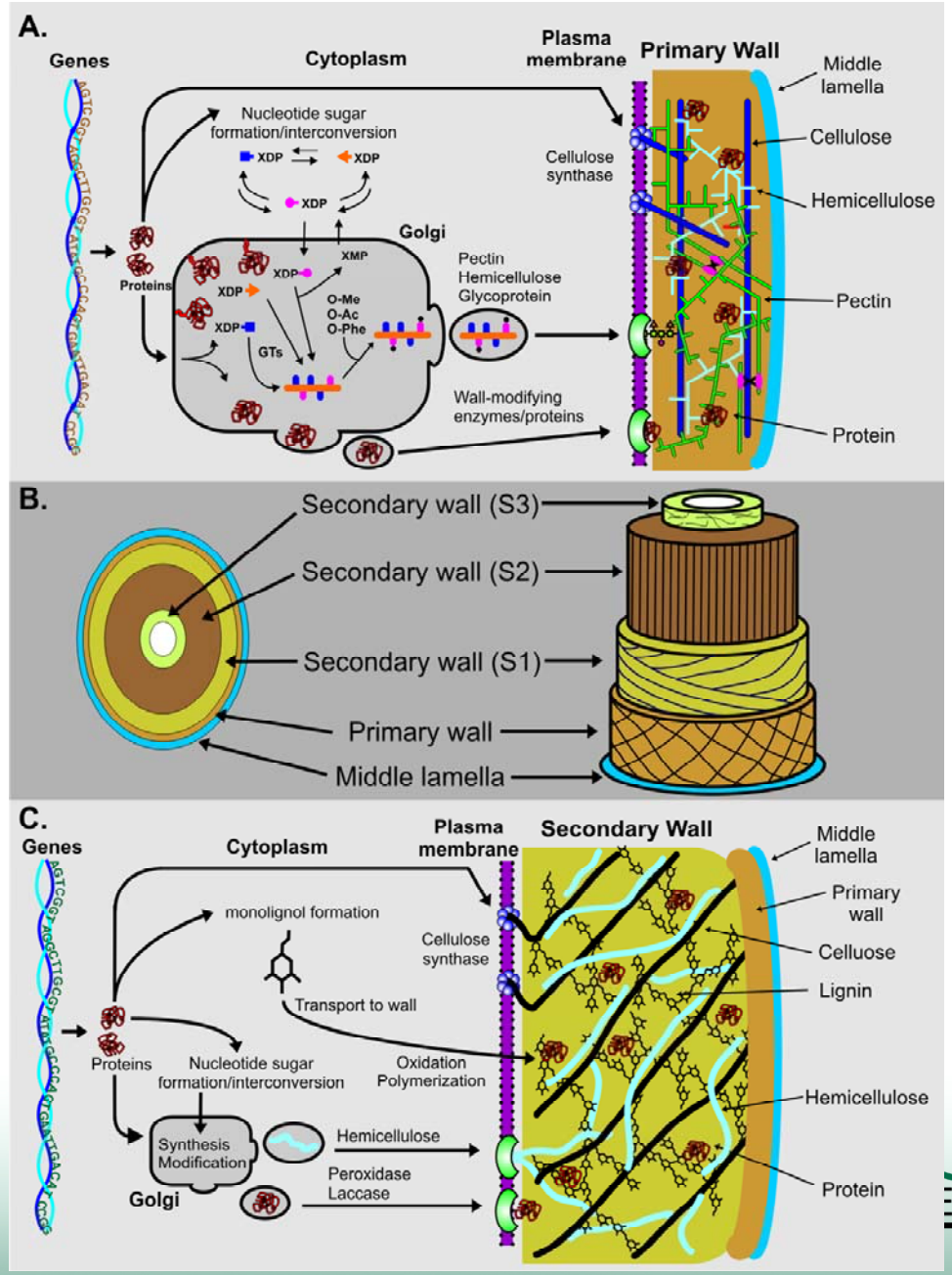
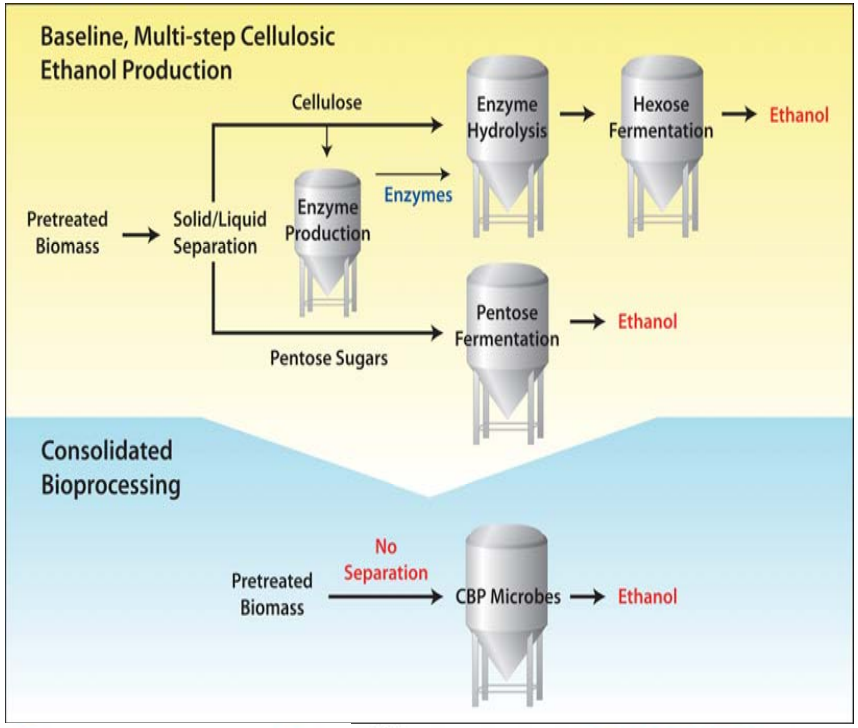
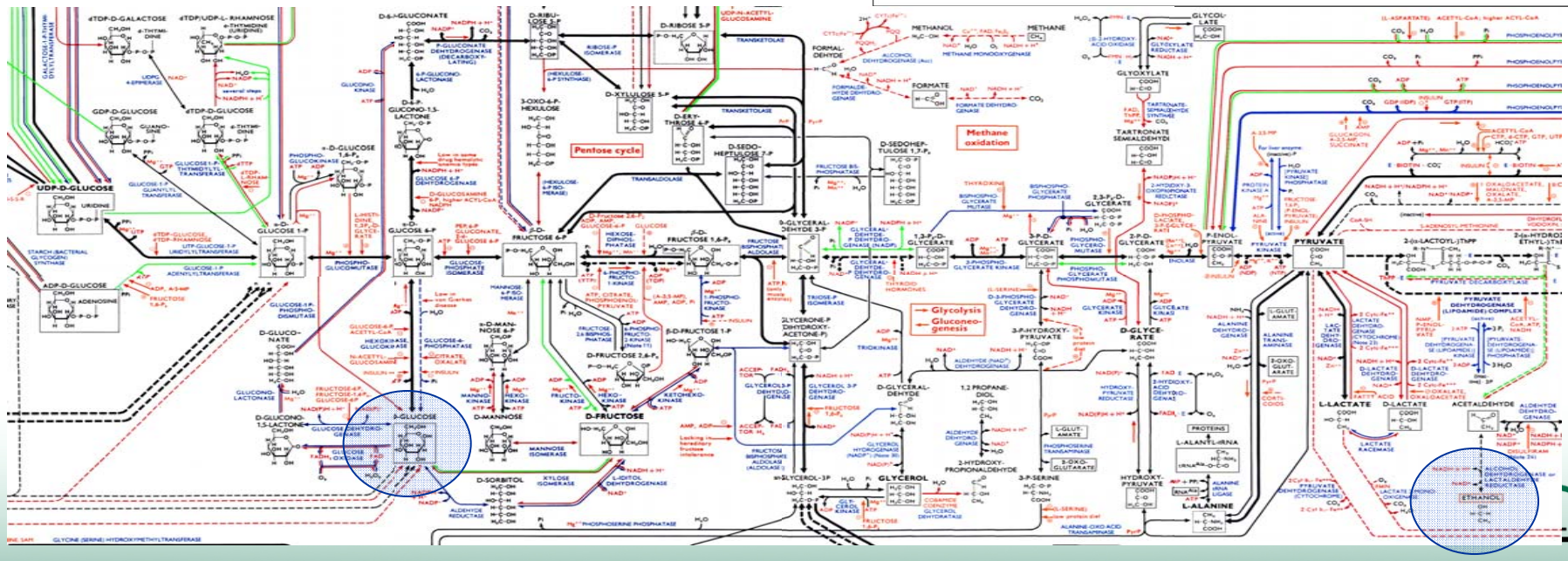


Fig. II.2. Biosynthesis of primary and secondary walls: from genes to polymers. A.

The challenges: Lignocellulosic biomass is difficult to breakdown and ferment



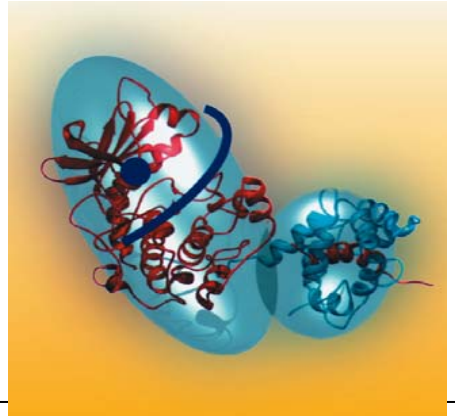


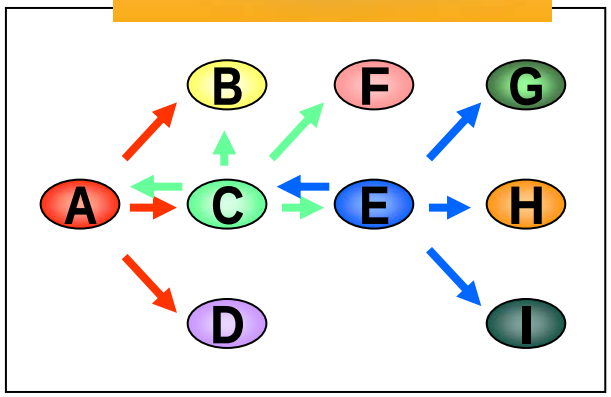
Systems Biology:

Philosophical approach to consider biology as an integrated complex of microbial and metazoan systems

Experimental approach to generate and analyze HTP data

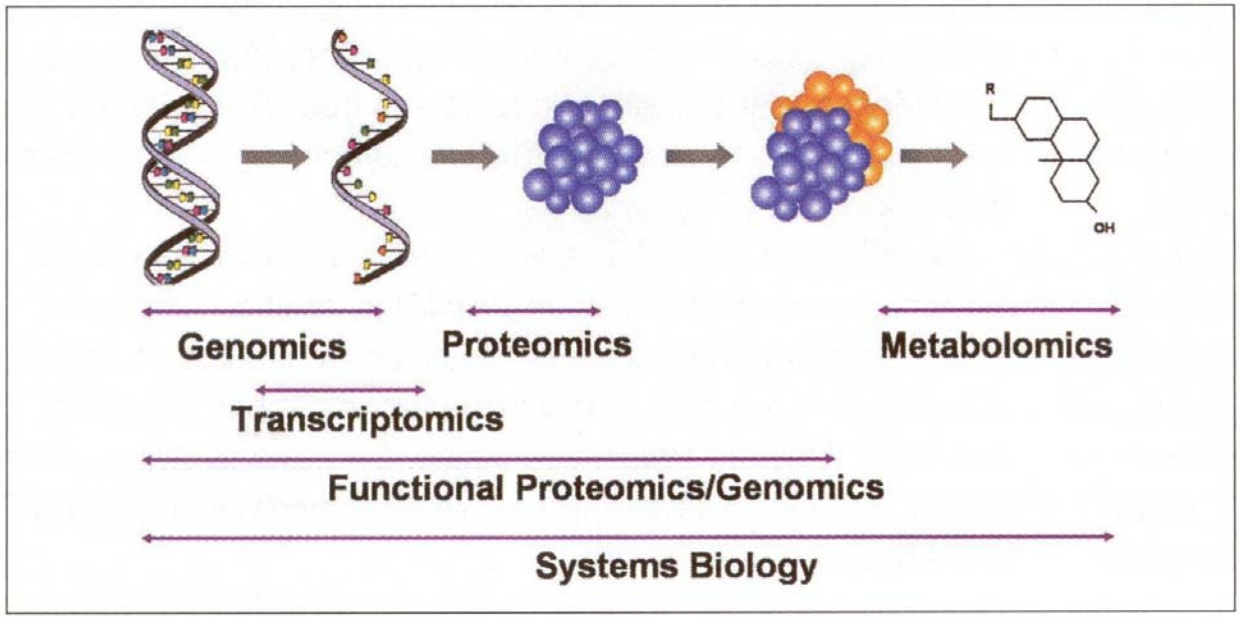
ORNL's goal is to apply systems biology to the missions of sustainable bioenergy





ORNL (and others) has “omic” capabilities for Systems Biology

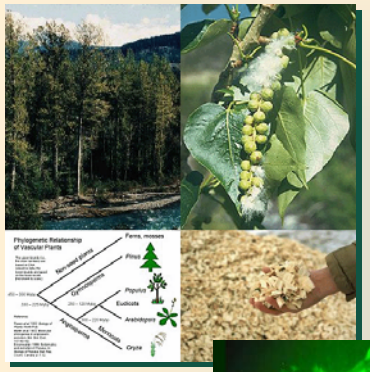
- Genomics
- Transcriptomics
- Proteomics
- Interactomics
- Metabolomics

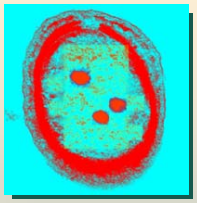


- Together these can provide a deeper picture of how a microbe or plant is functioning
- This can help identify where improvements need to be made

Components of ORNL Biology Program

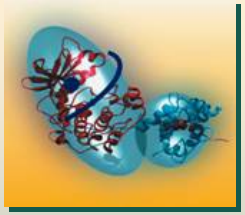
Plant Sciences



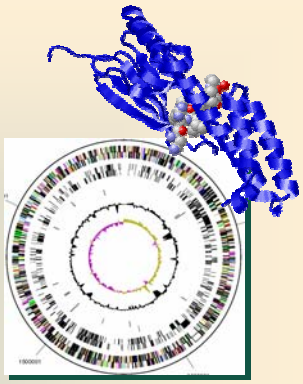


Microbiology

Mass spectrometry

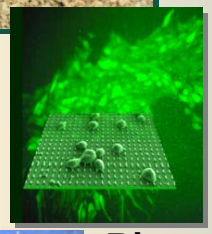


Neutrons for structural biology



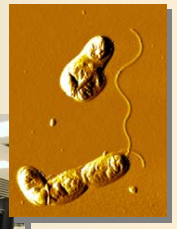
Bioinformatics

Understand molecular interactions in cells and communities



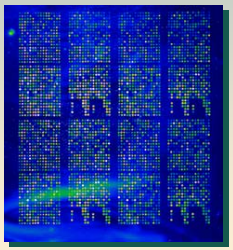
Biotechnology

Imaging



Joint Genome Institute

Mouse genetics and comparative genomics



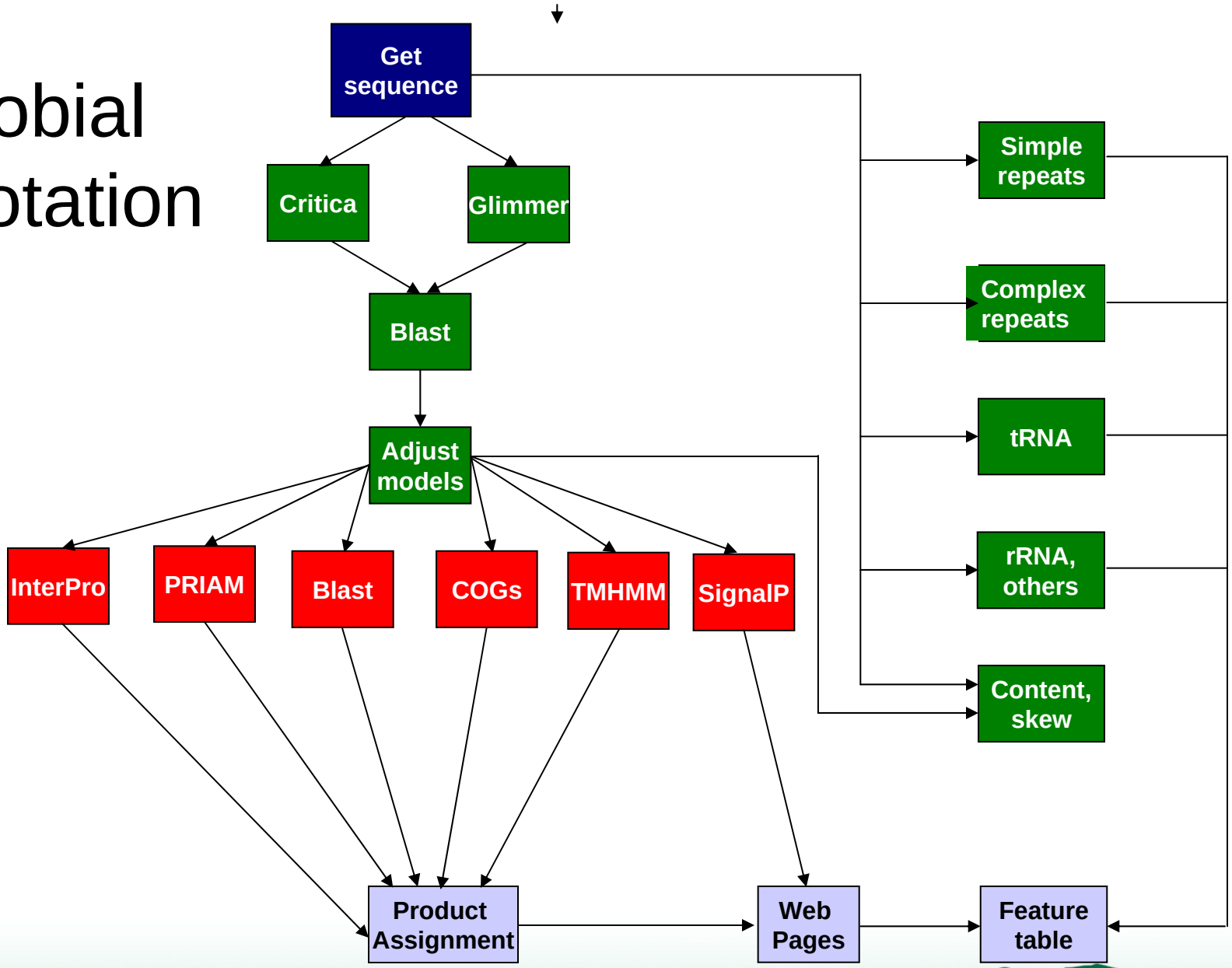
Expression profiling

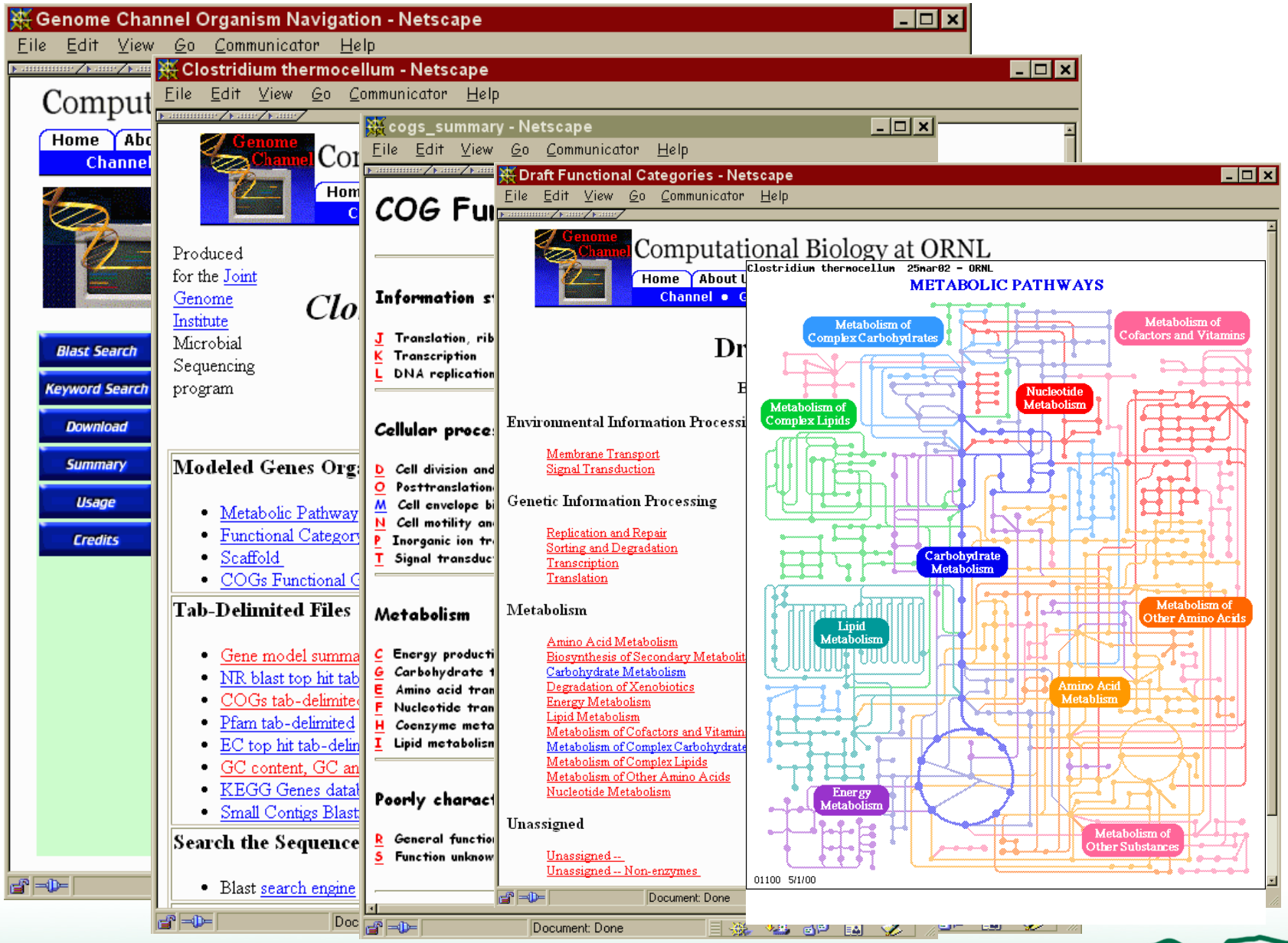
Computational biology

Genomics

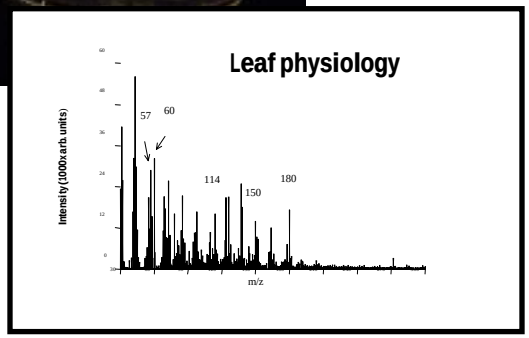
- Genome sequence helps provide baseline of what an organisms capabilities are
- ORNL can help provide access to sequencing capability to JGI through off-cycle Lab Science program
- ORNL provides primary annotation to all microbes sequenced at JGI
- This can help OBP's industrial partners

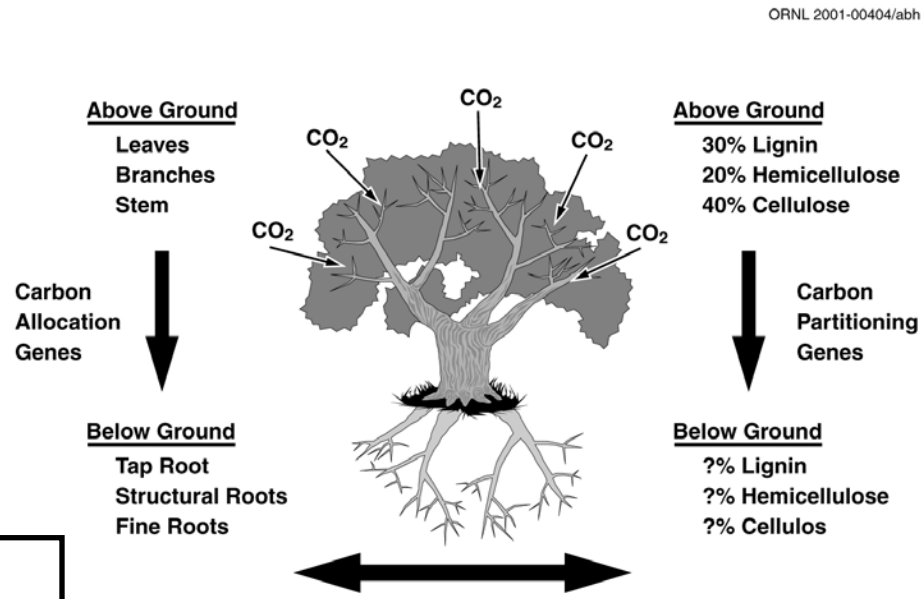
Microbial Annotation




Genome Channel Organism Navigation - Netscape
Clostridium thermocellum - Netscape
cogs_summary - Netscape
Draft Functional Categories - Netscape
File Edit View Go Communicator Help
Home
Channel
Blast Search
Keyword Search
Download
Summary
Usage
Credits
Produced for the Joint Genome Institute Microbial Sequencing program
Modeled Genes Org
Metabolic Pathway
Functional Category
Scaffold
COGs Functional C
Tab-Delimited Files
Gene model summa
NR blast top hit tab
COGs tab-delimite
Pfam tab-delimited
EC top hit tab-delin
GC content, GC an
KEGG Genes datal
Small Contigs Blast
Search the Sequence
Blast search engine
COG Fu
Information s
J Translation, rib
K Transcription
L DNA replication
Cellular proce
D Cell division and
O Posttranslation
M Cell envelope bi
N Cell motility an
P Inorganic ion tr
T Signal transduc
Metabolism
C Energy producti
G Carbohydrate t
E Amino acid tran
F Nucleotide tran
H Coenzyme meta
I Lipid metabolisn
Poorly charact
R General functio
S Function unknow
Document: Done
Computational Biology at ORNL
Environmental Information Processi
Membrane Transport
Signal Transduction
Genetic Information Processing
Replication and Repair
Sorting and Degradation
Transcription
Translation
Metabolism
Amino Acid Metabolism
Biosynthesis of Secondary Metabolit
Carbohydrate Metabolism
Degradation of Xenobiotics
Energy Metabolism
Lipid Metabolism
Metabolism of Cofactors and Vitamin
Metabolism of Complex Carbohydrate
Metabolism of Complex Lipids
Metabolism of Other Amino Acids
Nucleotide Metabolism
Unassigned
Unassigned --
Unassigned -- Non-enzymes
Clostridium thermocellum 25mar02 - ORNL
METABOLIC PATHWAYS
Metabolism of Complex Carbohydrates
Metabolism of Cofactors and Vitamins
Nucleotide Metabolism
Metabolism of Complex Lipids
Carbohydrate Metabolism
Metabolism of Other Amino Acids
Lipid Metabolism
Amino Acid Metablism
Energy Metabolism
Metabolism of Other Substances
01100 5/1/00

ORNL is a world leader in woody plant genomics – *designing crops for energy & C sequestration*





High-throughput Phenotyping

Plant Traits Needed to Address all Parts of Equation:

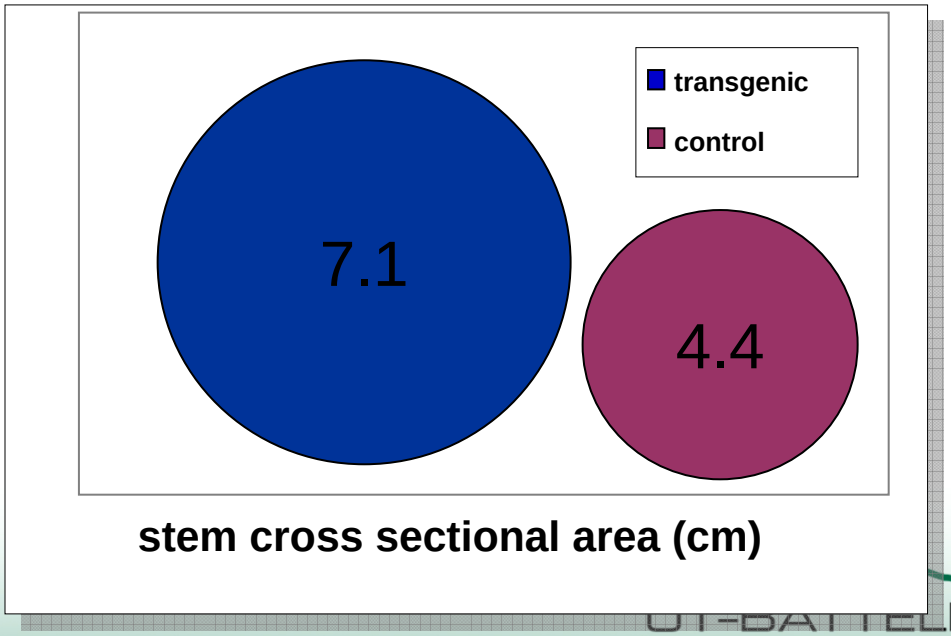
Targeted Goal	Traits & Technologies
Broaden Planting Range	• Tolerance to chronic and acute drought • High salt tolerance • Tolerance to heat shock • improvement in seedling growth under cold conditions
Increase tons per acre	• Increase in biomass in arabidopsis in the greenhouse (5X) • Increase in rice in the field (3X) • Increase in CO2 uptake (30%)
Dollars per acre	• Significant reduction in required nitrogen • Improvement in photosyn efficiency on low nitrogen (20%) • increase in root biomass
Gallons per ton	• Decreased lignin • Increased cellulose/ hemicellulose

Source: Modified from Ceres Corp.

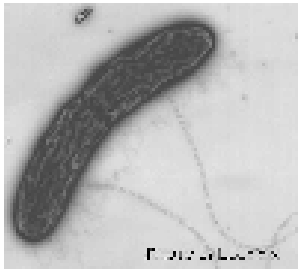
Populus - early results from genome sequence availability (Jerry Tuskan, ORNL)

90-day-old *Populus* cuttings

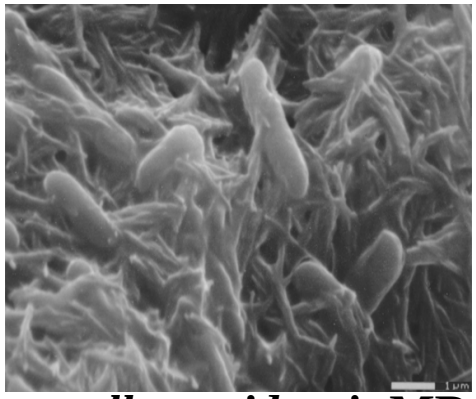
Using Poplar tree genome, the expression of one gene (IAA16.3) was altered. This resulted in enhanced radial growth of IAA16.3 transgenics vs. controls



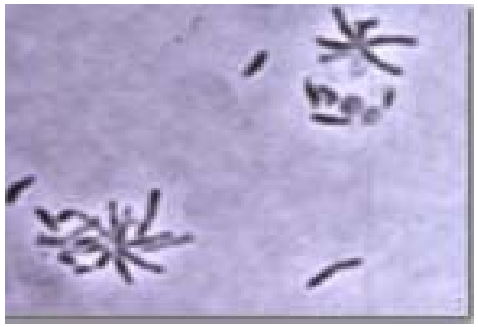
Whole genome microarrays available



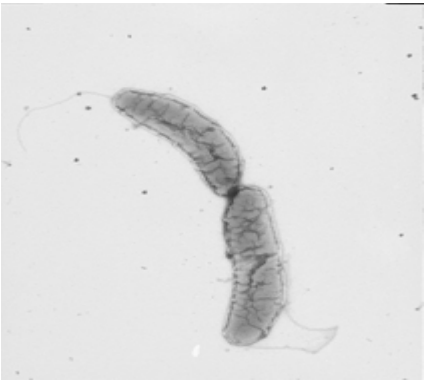
Geobacter metallireducens:
Metal-reducing bacterium



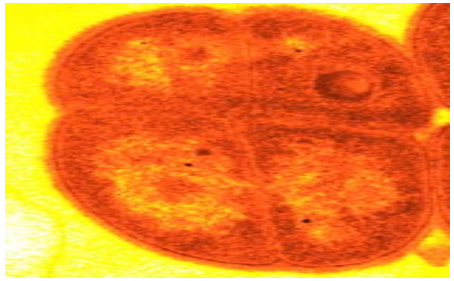
***Shewanella oneidensis* MR-1:**
Metal-reducing bacterium



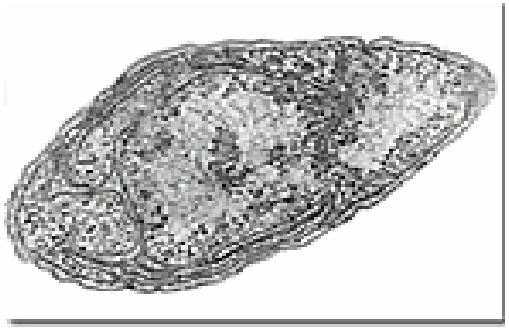
***Rhodopseudomonas palustris*:**
Photosynthetic bacterium



***Desulfovibrio vulgaris*:**
Sulfate-reducing bacterium

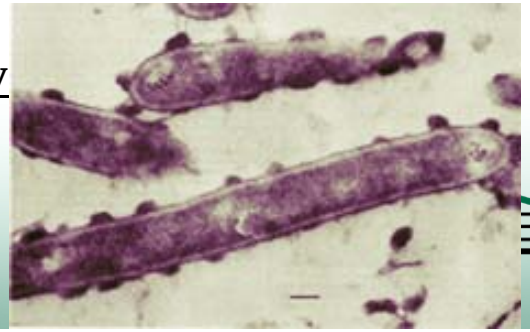


***Deinococcus radiodurans* R1:**
Radiation-resistant bacterium



***Nitrosomonas europaea*:**
Ammonium-oxidizing bacterium

***Clostridium thermocellum* now ready**
cellulolytic fermenting bacterium



Transcriptomics

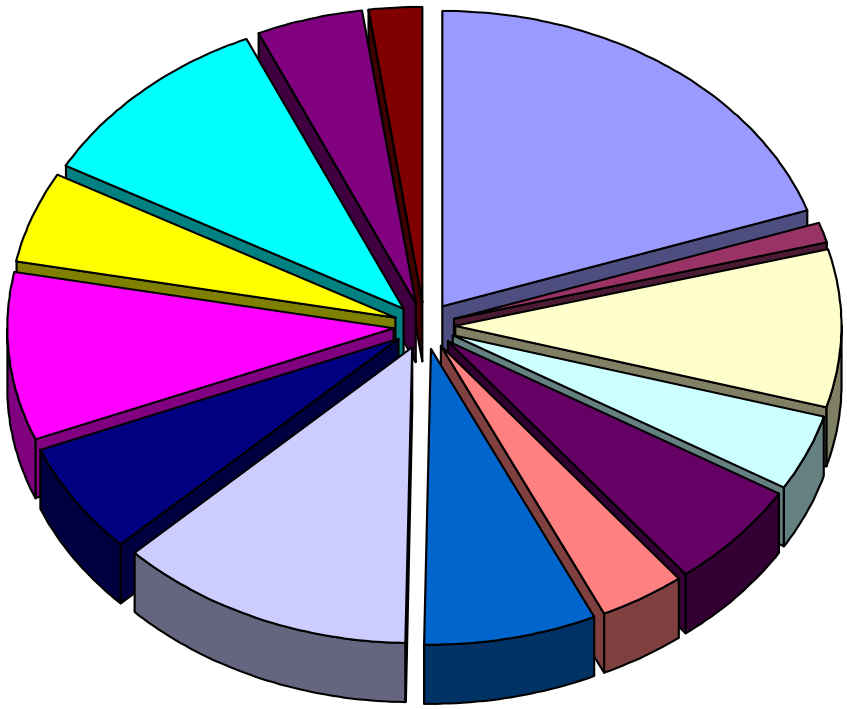
- Tells what genes the microbe is trying to express under what conditions
- Typically requires a sequenced genome
- ORNL has capability to rapidly prepare new arrays (for quantities to small for commercial interest)
- New approaches (functional genomics) allow testing of mixed cultures and unknowns in a Functional Genome Array (FGA)

MS Proteomics

- Can identify what proteins are actually present
- When and where proteins are present
- Detailed analysis can identify how proteins are altered to change activity and location (post translational modifications)
- Example: Strader performed top-down and bottom-up analysis of isolated R. palustris ribosome
  - Identified 52 of known 54 subunits
  - Identified PTMs (methylation, acetylation, glycosylation) and location on many of the subunits
  - Improper PTMs are why many genetically engineered proteins do not work properly the first time

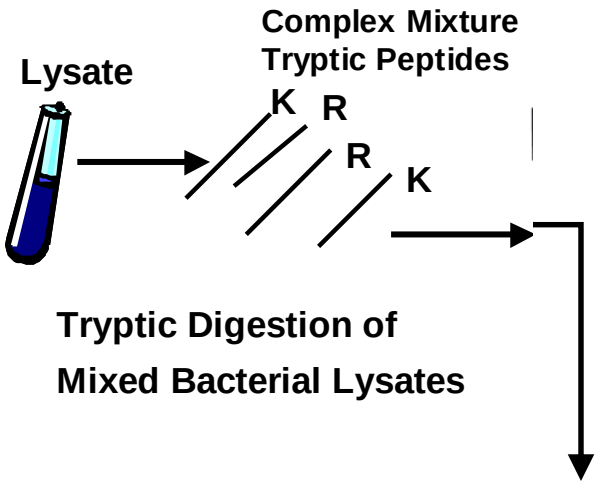
Proteomics: Functional Categories of *R. palustris* Proteins

***R. palustis* proteome**
1695 Total Proteins



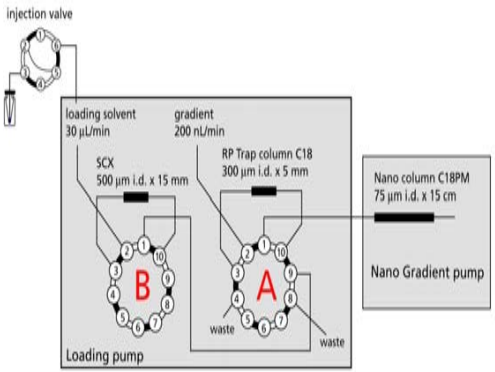
- Unknowns and Unclassified
- Replication and Repair
- Energy Metabolism
- Carbon and Carbohydrate Metabolism
- Lipid Metabolism
- Transcription
- Translation
- Cellular Processes
- Amino Acid Metabolism
- General function prediction
- Metabolism of Cofactors and Vitamins
- Transport
- Signal Transduction
- Purine and Pyrimidine Metabolism

Proteome Analysis with Mass Spectrometry



Tryptic Digestion of Mixed Bacterial Lysates

Autosampler makes injections on to the 2D system from fractions in microtiter plates



Automated 2D HPLC
1st Dimension: SCX
2nd Dimension: Nano RP C18

ES-Ion Trap MS
Data-Dependent MS/MS

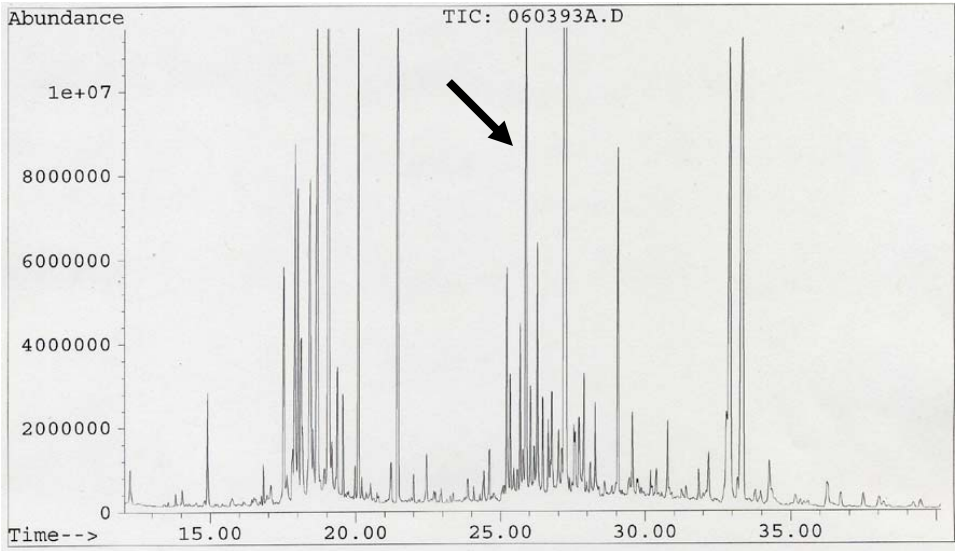
Interactomics

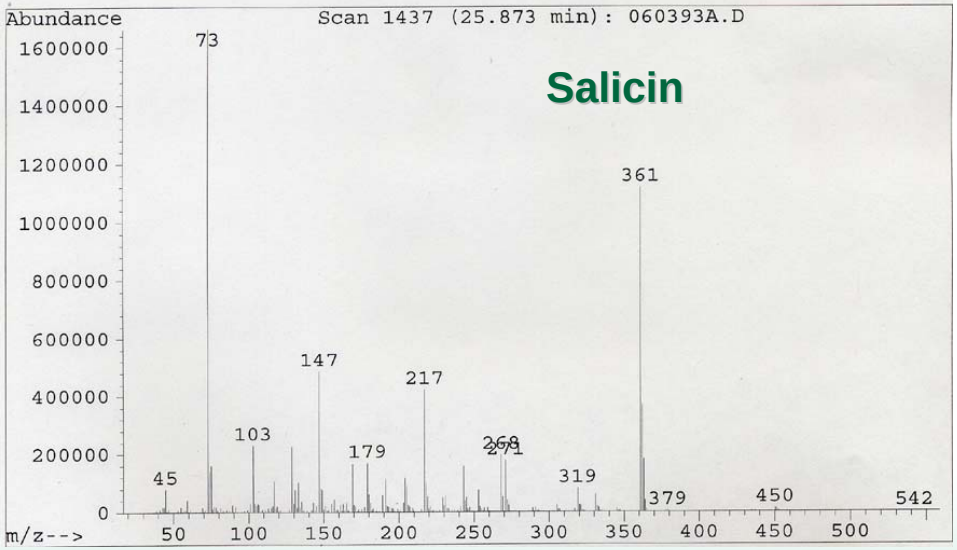
- Identifies required units of a protein complex
  - So that all pieces can be used together
- Identifies transitory protein interactions
  - These may be critical regulators or cofactors

Metabolomics

- LC-MS/MS techniques for plant and microbial metabolites
- Identifies what is actually being made
- Challenge is the “steady state” levels are often low
- Metabolic flux requires modeling

Metabolic Variants





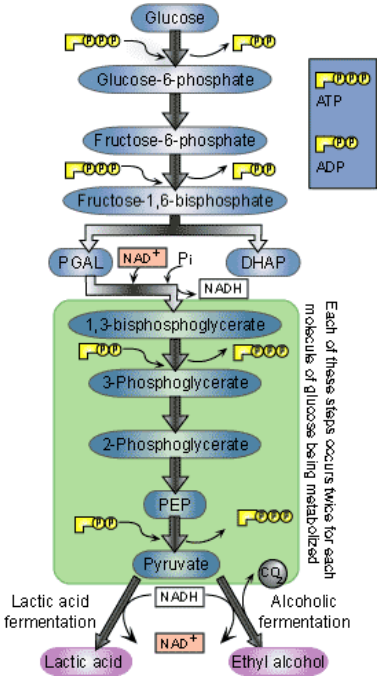
Segregating Population

F2 Progeny

Charged metabolites targeted by PACE-ES-MS

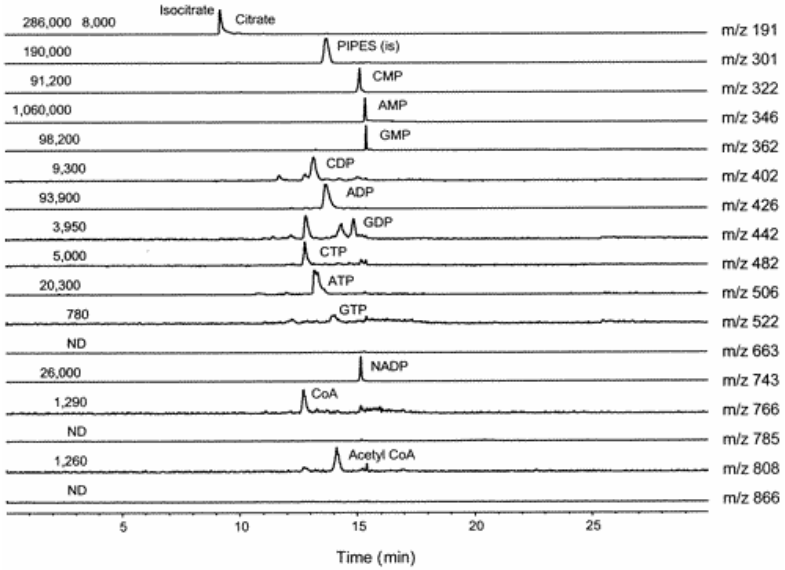
Beckman Coulter PACE system

- acetate
- lactate
- pyruvate
- acetyl phosphate
- acetyl-CoA
- ferredoxin (reduced)
- ferredoxin (oxidized)
- NAD
- NADH
- NADP
- NADPH
- ATP
- ADP
- AMP
- glucose-6P
- glucose-1P
- fructose-6P
- fructose-1P
- fructose-1,6-P2
- glyceraldehyde-3P
- glycerate-1,3P2
- glycerate-3P
- glycerate-2P
- phosphoenolpyruvate
- ribulose-1,5-P2
- inositol phosphates



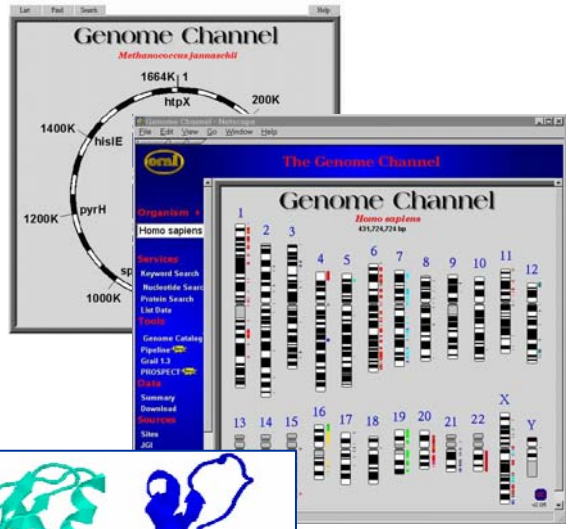
API Qstar XL

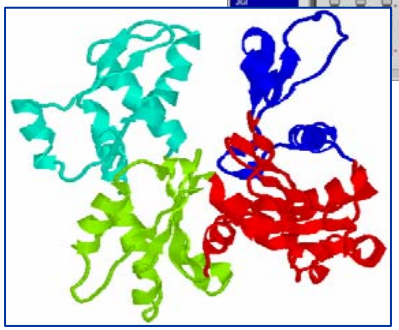
Hybrid Qq-Time-of-Flight-MS System



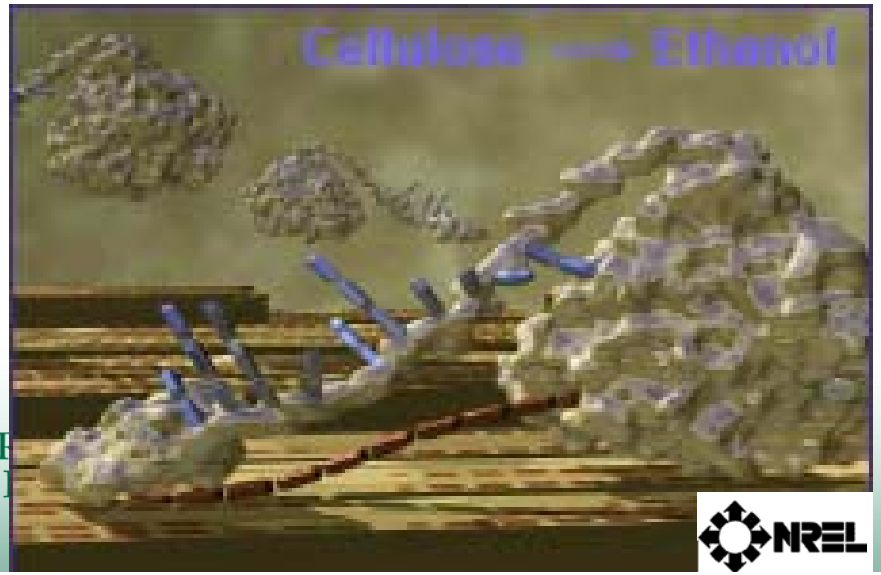
Computational Biology

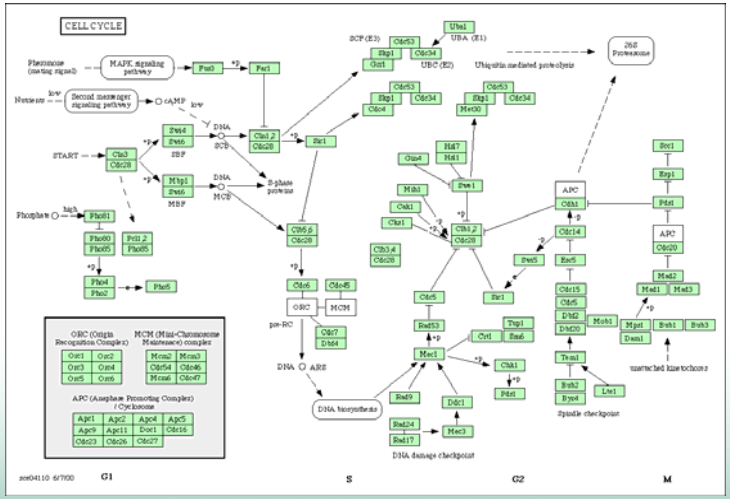
- Annotation DNA sequence analysis, gene finding, regulatory regions, comparative analyses for JGI Microbes
- Systems analysis of mRNA and protein expression patterns
  - Meta-omic databases
- Protein structure prediction
- Protein complexes
- High-performance computing
  - Petaflop computing



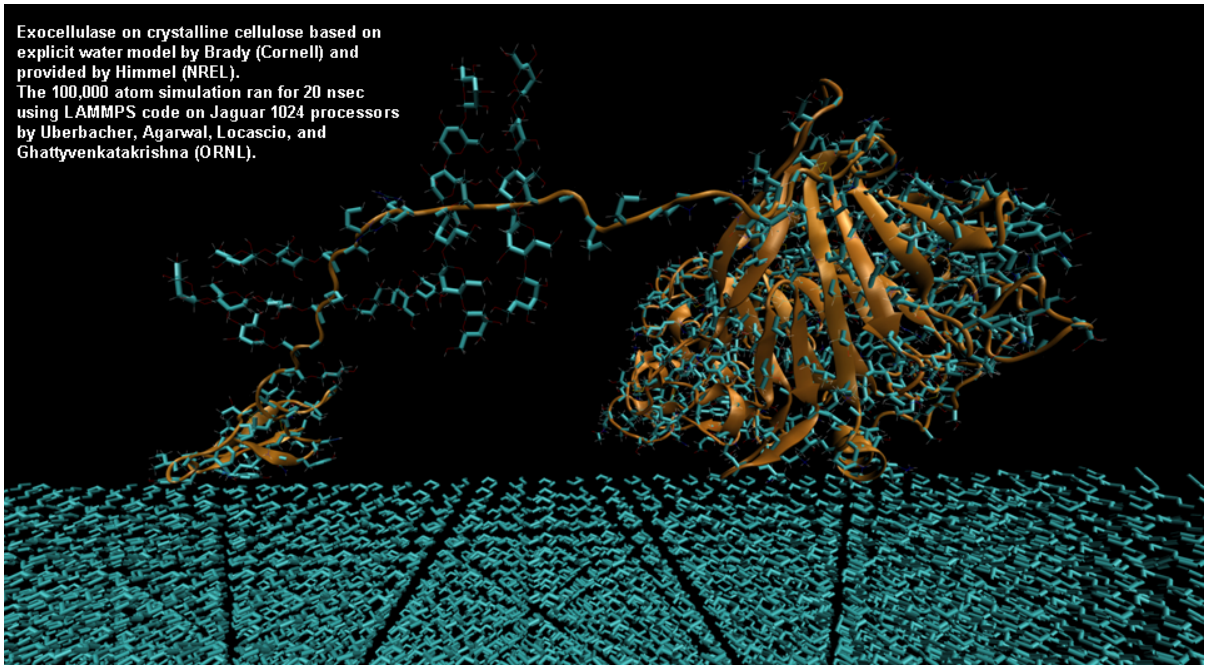


yeast





Modeling cellulase on cellulose



- 100,000 atom model of enzyme cellulase on crystalline cellulose in water
- Computation for 50 nS on Jaguar by Uberbacher et al.
- Based on explicit model of Brady (Cornell) and Himmel (NREL) – leads of new project.
- Shows cellulose binding domain and active site interaction with surface

Neutrons and Biology

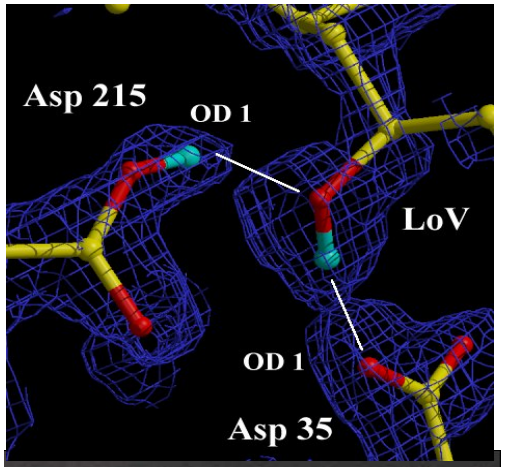
ORNL will provide world-leading instruments for neutron scattering at HFIR and at SNS

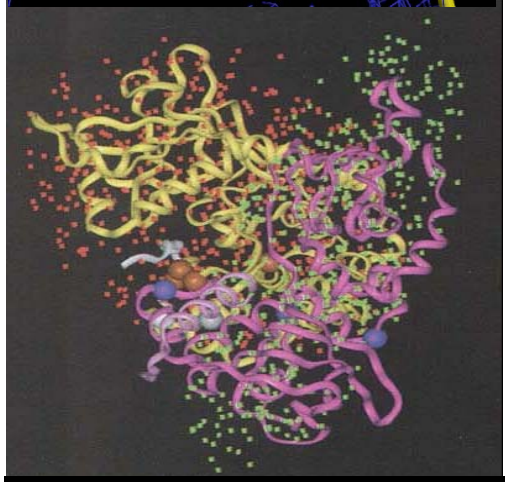
Neutrons are excellent probes for Hydrogen –and can discriminate between hydrogen and deuterium

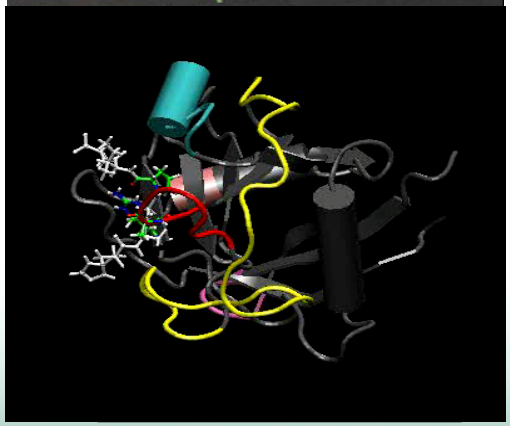
Function: H/D in enzyme mechanism; proton shuttling & transfer

Structure: H/D Labeled protein in complex systems

Dynamics: Specific H-Labeling in deuterated systems







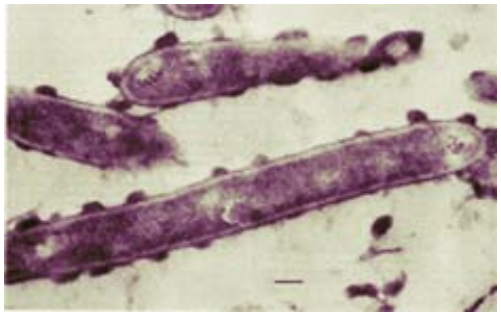
Goal is to extend Systems Biology into understanding organism in its environment and its communities

- Community analyses
  - Metatranscriptome, metaproteome
- Microbial Communities
- Plant-Microbe interactions
- Integrated “Omic” analysis

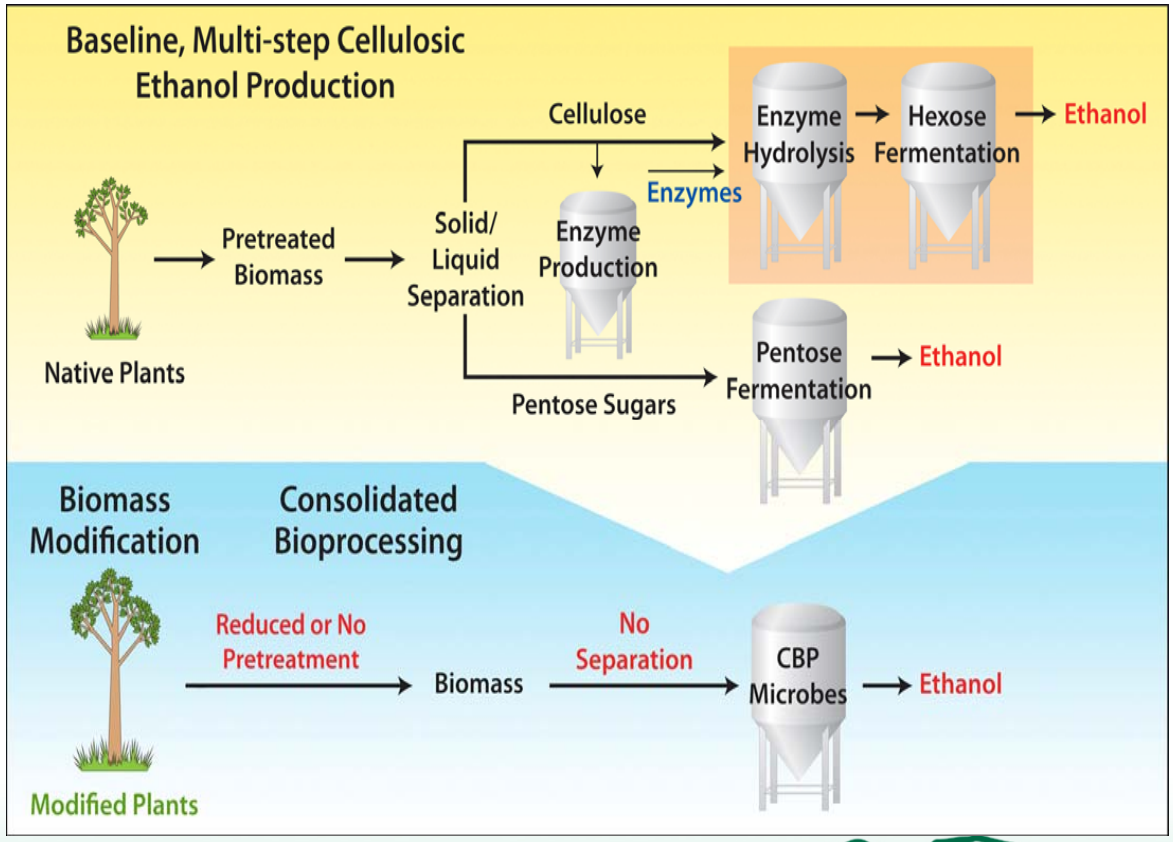
Bioenergy and Microbial Systems Biology

ORNL is performing systems biology to discover regulation in Consolidated Bioprocessing (CBP)

- Genome annotation
- Transcriptome
- Controlled cultivation
- Metabolites
- Proteomics
- in collaboration with Dartmouth/Mascoma



Clostridium thermocellum
Ethanologenic cellulolytic bacterium

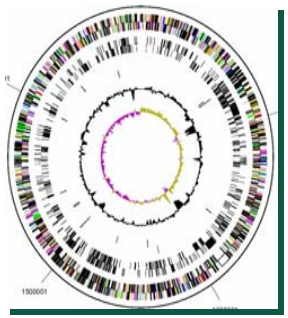


‘Omics technologies applied to help design better ethanologens - where are the biochemical limitations?

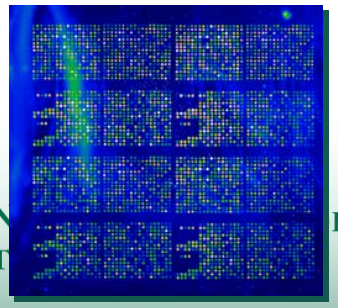
Biochemical pathway

Cellulose → Ethanol & Byproducts

Genome annotation gives possible genes, proteins, etc.



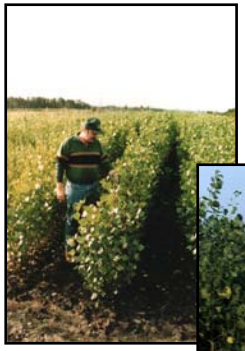
Transcriptome - Gene expression measured by microarray analysis



Proteome- Protein expression measured by mass spectrometry

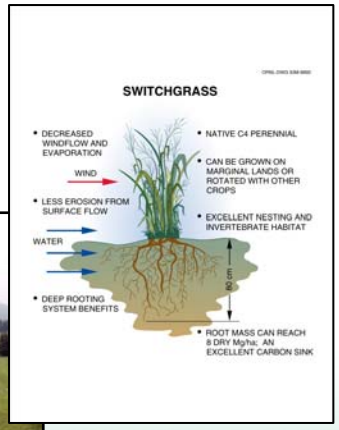
Metabolome - Cellulose hydrolysis fragments and fermentation products measured by mass spectrometry

Bioenergy and Plant Genomics: Expanding the nation's renewable energy resources



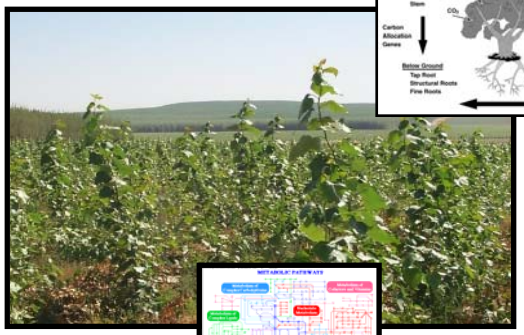
Conventional Forestry

Yesterday

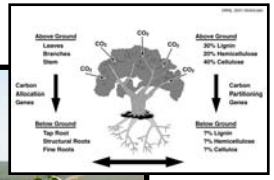


Short rotation hardwoods

Today



Carbon allocation



Metabolic Profiling

Whole Genome Microarrays

Accelerated Domestication

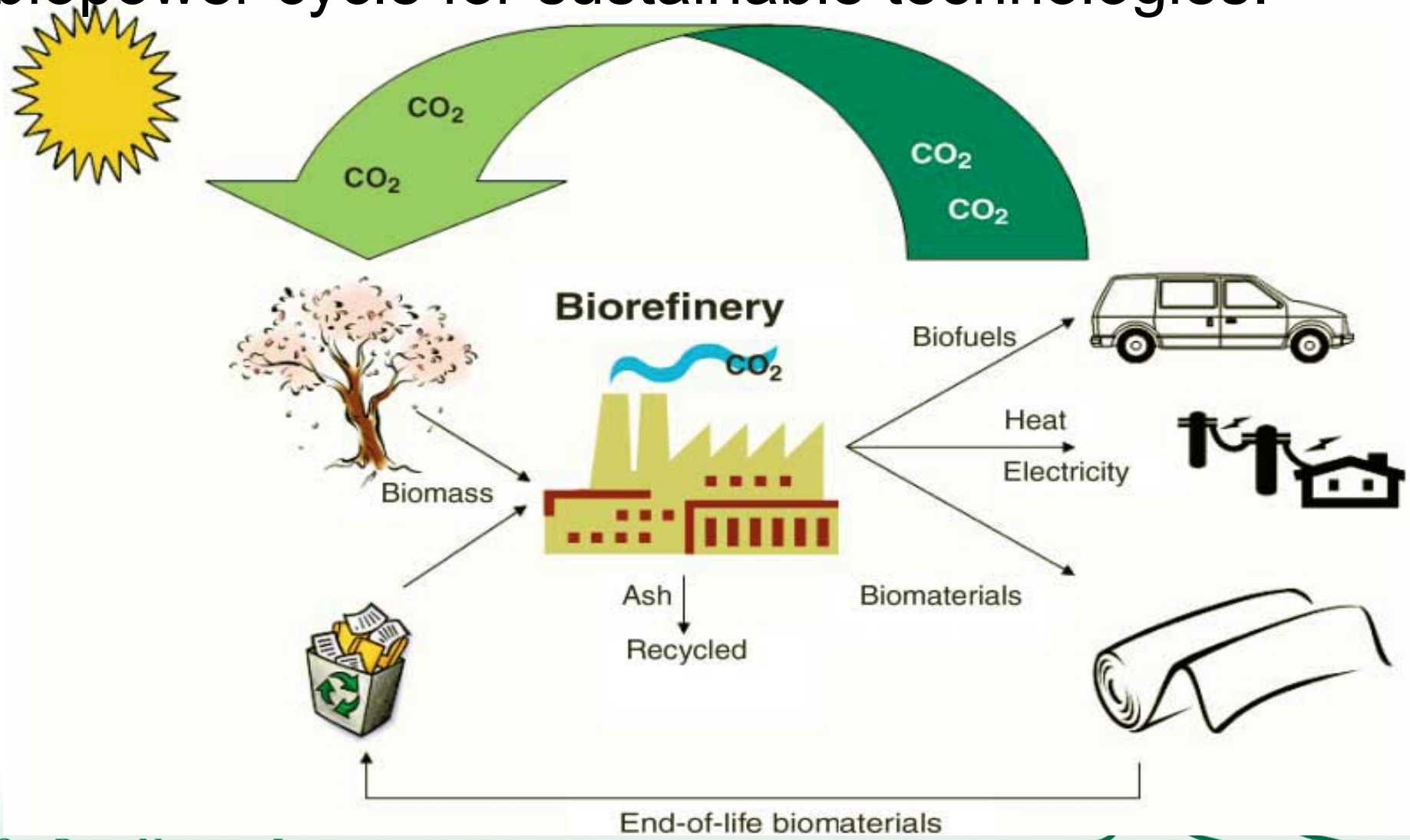
Tomorrow

High yield wood crops

Putting Genomes to work For Energy Security

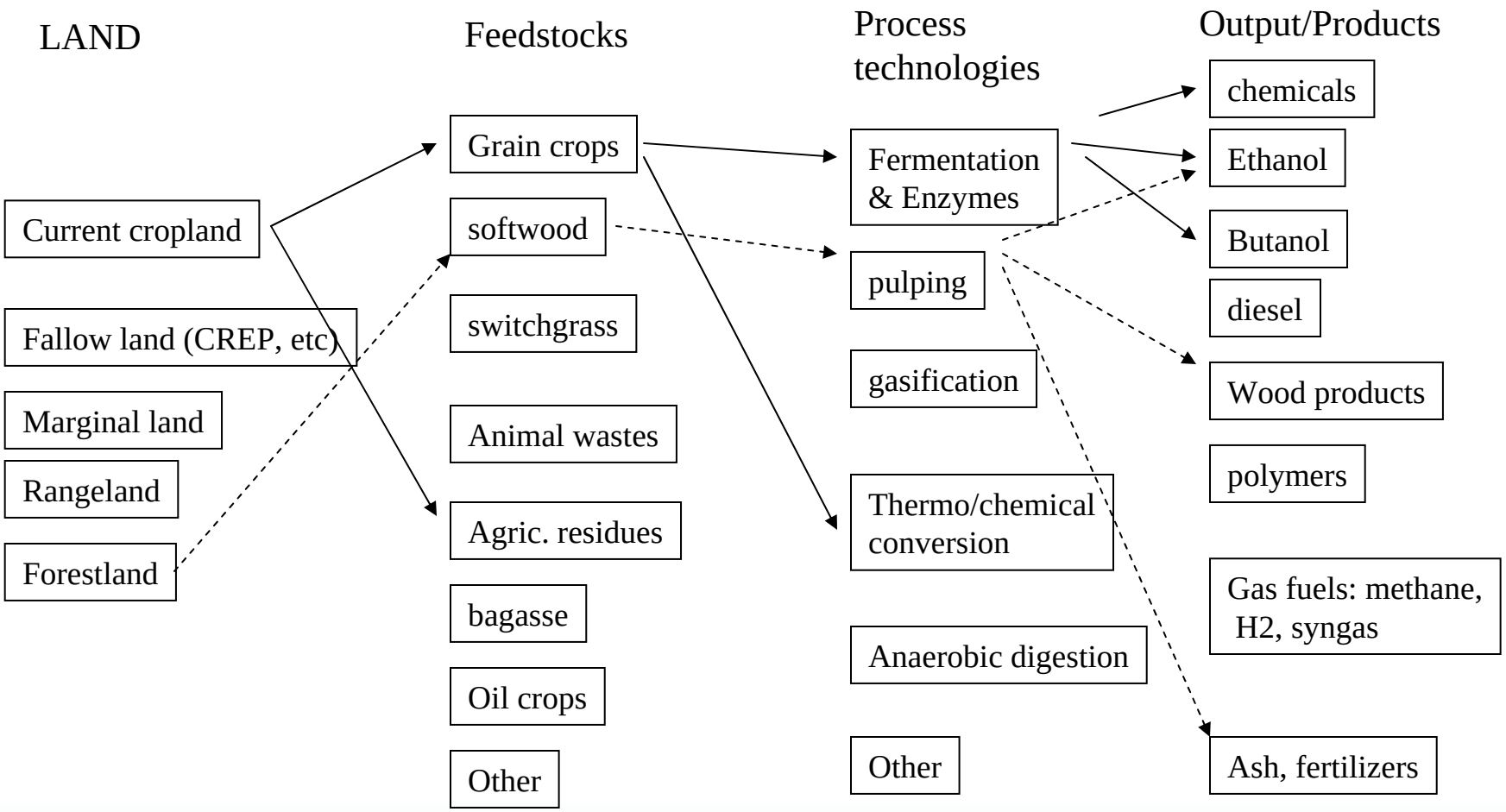
UT-BATTELLE

The fully integrated agro-biofuel-biomaterial-biopower cycle for sustainable technologies.

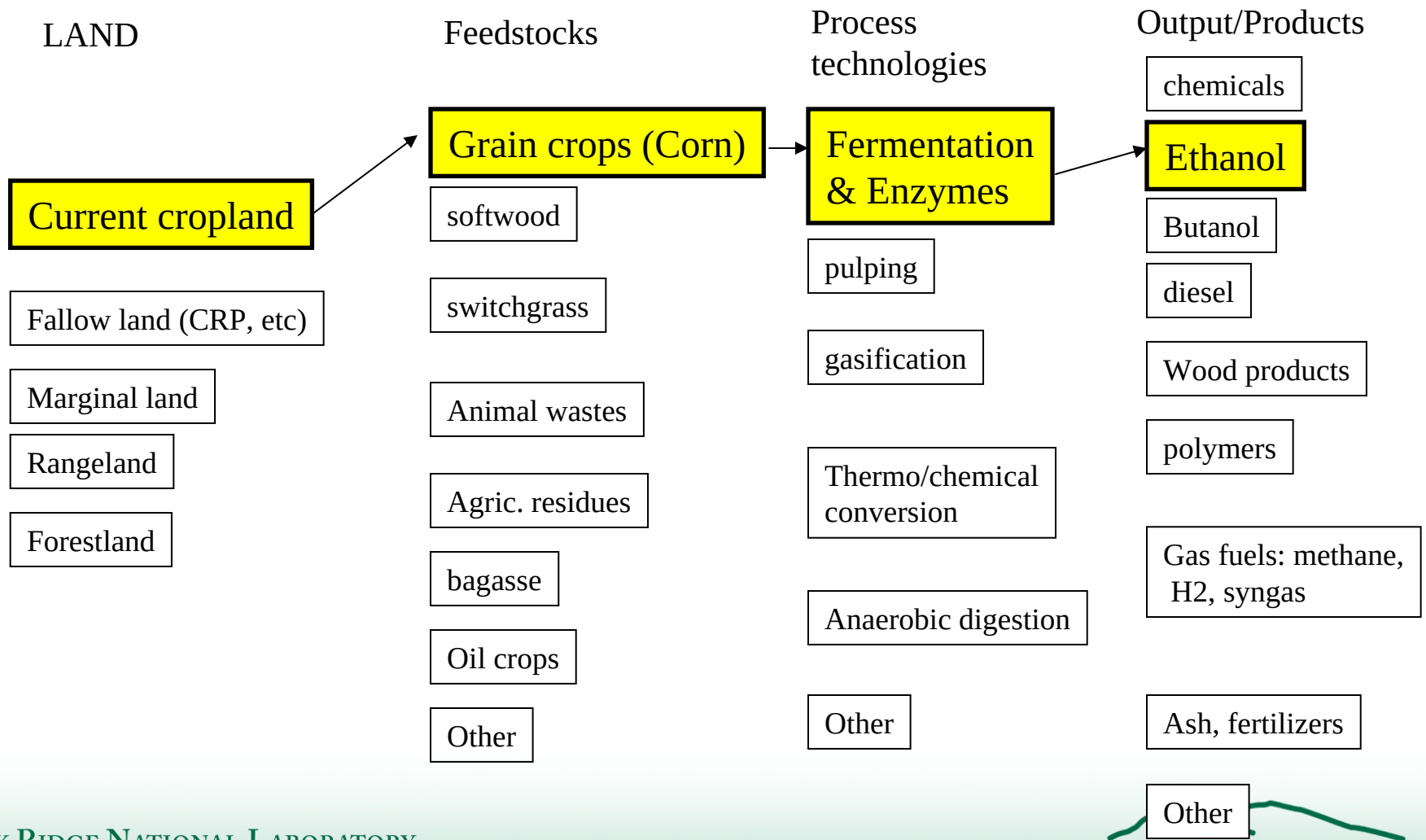


Ragauskas et al. 2006, *SCIENCE* **311**:484-489

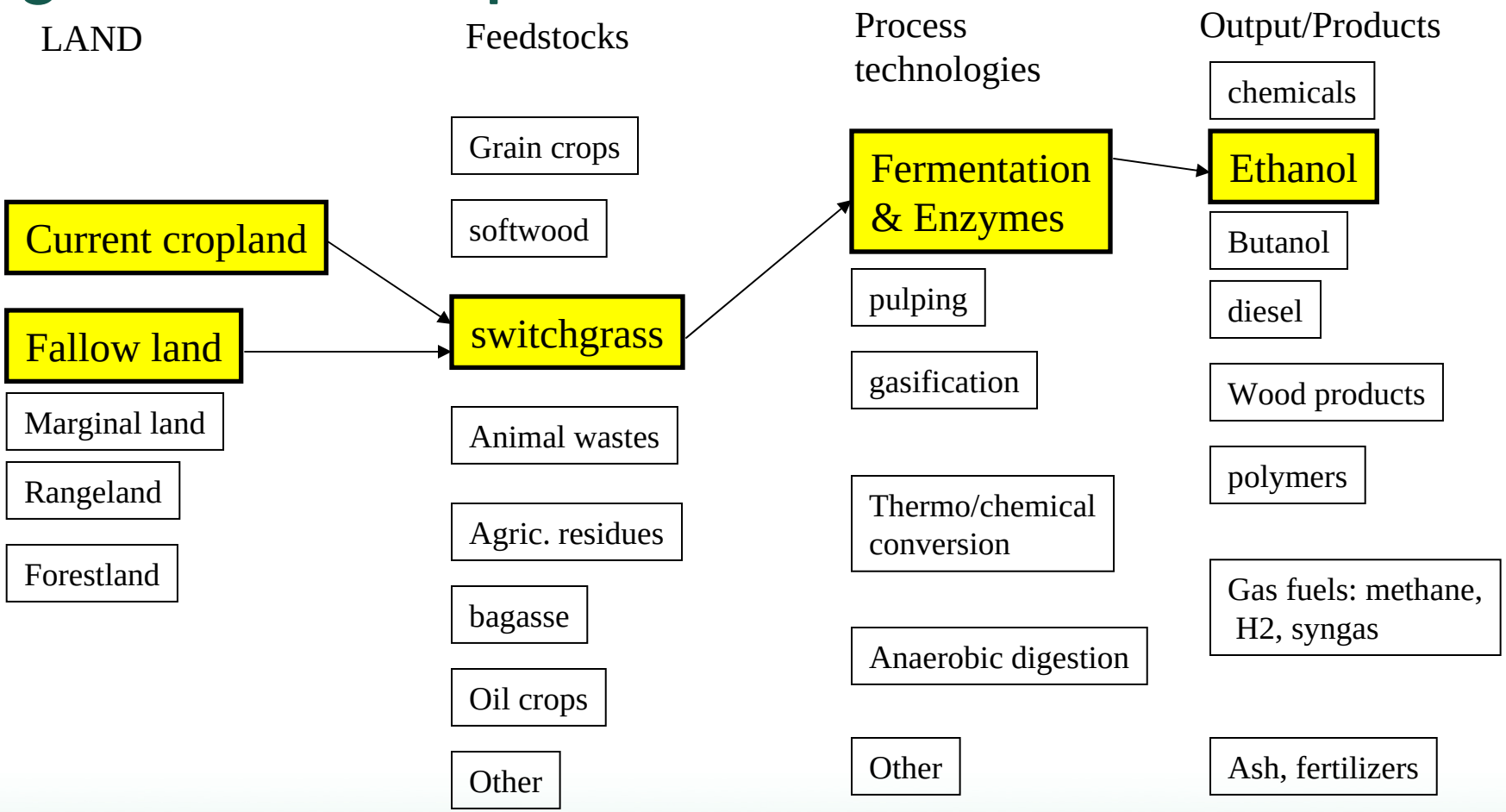
Biomass Utilization is a multi-factorial problem (multiple choice)



Biomass Utilization: current Corn Bioethanol



Biomass Utilization: Gov. Bredesen's announcement of switchgrass to bioethanol might follow this path



BioEnergy Science Center:
a DOE Bioenergy Research Center

Martin Keller, Ph.D.
BESC Director

July 2007

http://www.bioenergycenter.org/

BESC
BioEnergy Science Center

ORNL will lead a DOE Bioenergy Research Center

- **Funded for $125M over 5 y**
  - **Awarded June 26**
- **Start in Sept. 07**
- **Two other centers at**
  - **U Wisconsin**
  - **www.wisconsinbioenergy.com**
  - **LBNL**
  - **jbei.lbl.gov**

BESC
BioEnergy Science Center

BESC
BioEnergy Science Center

The BESC team has been assembled to overcome recalcitrance

Joint Institute for Biological Sciences

Alternative Fuels User Facility

Complex Carbohydrate Research Center

- Oak Ridge National Laboratory
- University of Georgia
- University of Tennessee
- National Renewable Energy Laboratory
- Georgia Tech
- Samuel Roberts Noble Foundation
- Dartmouth
- ArborGen
- Diversa
- Mascoma
- Individuals from U California-Riverside, Cornell, Washington State, U Minnesota, NCSU, Brookhaven National Laboratory, Virginia Tech

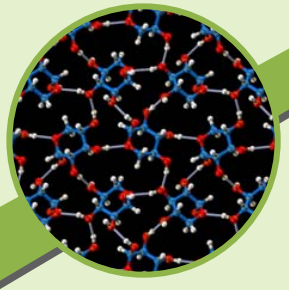
BESC
BioEnergy Science Center

The fundamental science of biomass recalcitrance is poorly understood

Overcoming recalcitrance is the single coherent overarching theme for the BESC

Recalcitrance: Resistance to breakdown into sugars

Cellulosic biomass



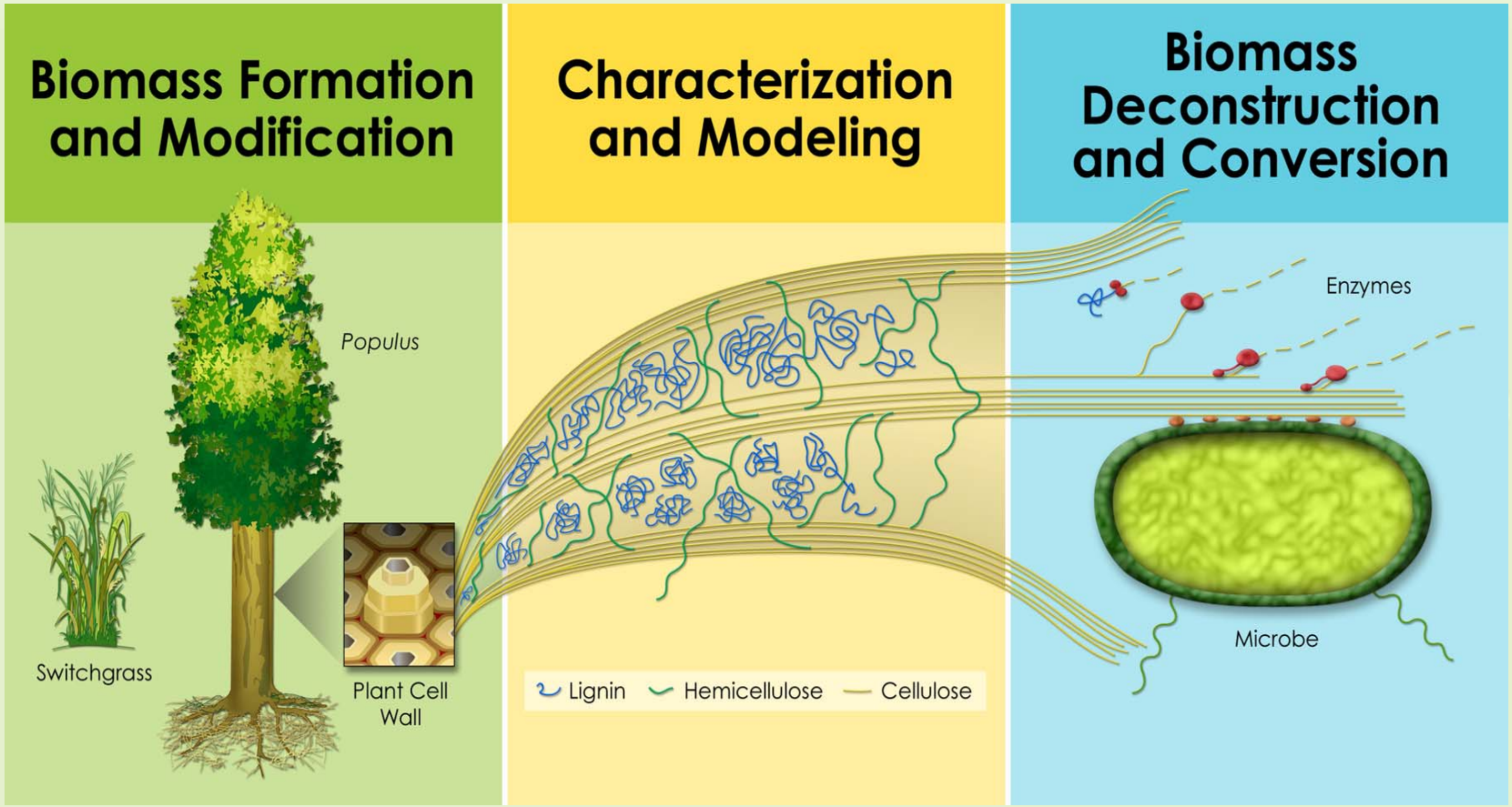
Sugars

Fuel(s)

- A large-scale, integrated, interdisciplinary approach is needed to overcome this problem
  - Current research efforts are limited in scope
  - BESC will launch a broad and comprehensive attack on a scale well beyond any efforts to date
- Without advances, a cellulosic biofuels industry is unlikely to emerge
- Knowledge gained will benefit other biofuels and biofeedstocks

BESC
BioEnergy Science Center

Three linked scientific focus areas will enable BESC to understand and overcome biomass recalcitrance



BESC
BioEnergy Science Center

BESC leverages substantial firm commitments and investments

- State of Tennessee, through UT: >$24.6M
  - JIBS construction: $11.6M
  - Research equipment: $3M
  - Joint UT-ORNL research: $10M
  - 3 Governor's Chairs: $3M–$5M
- State of Tennessee: $48M
  - $40M for 5M-gal/year pilot switchgrass-to-ethanol facility, located <40 miles from ORNL, to be operational in 2009
  - $8M in agricultural price supports
- Mascoma: $5M to support CBP research at Mascoma and Dartmouth
- Georgia Research Alliance: $6M
  - $3M for equipment,
  - $1.5M for two Eminent Scholar hires (matched by $1.5M from UGa and Georgia Tech)
- Virginia Tech: $0.5M in cost share support
- Oklahoma Bioenergy Center
  - Update to Governor Henry's letter
  - >$10M at the Noble Foundation as part of this $40M center with University of Oklahoma and Oklahoma State

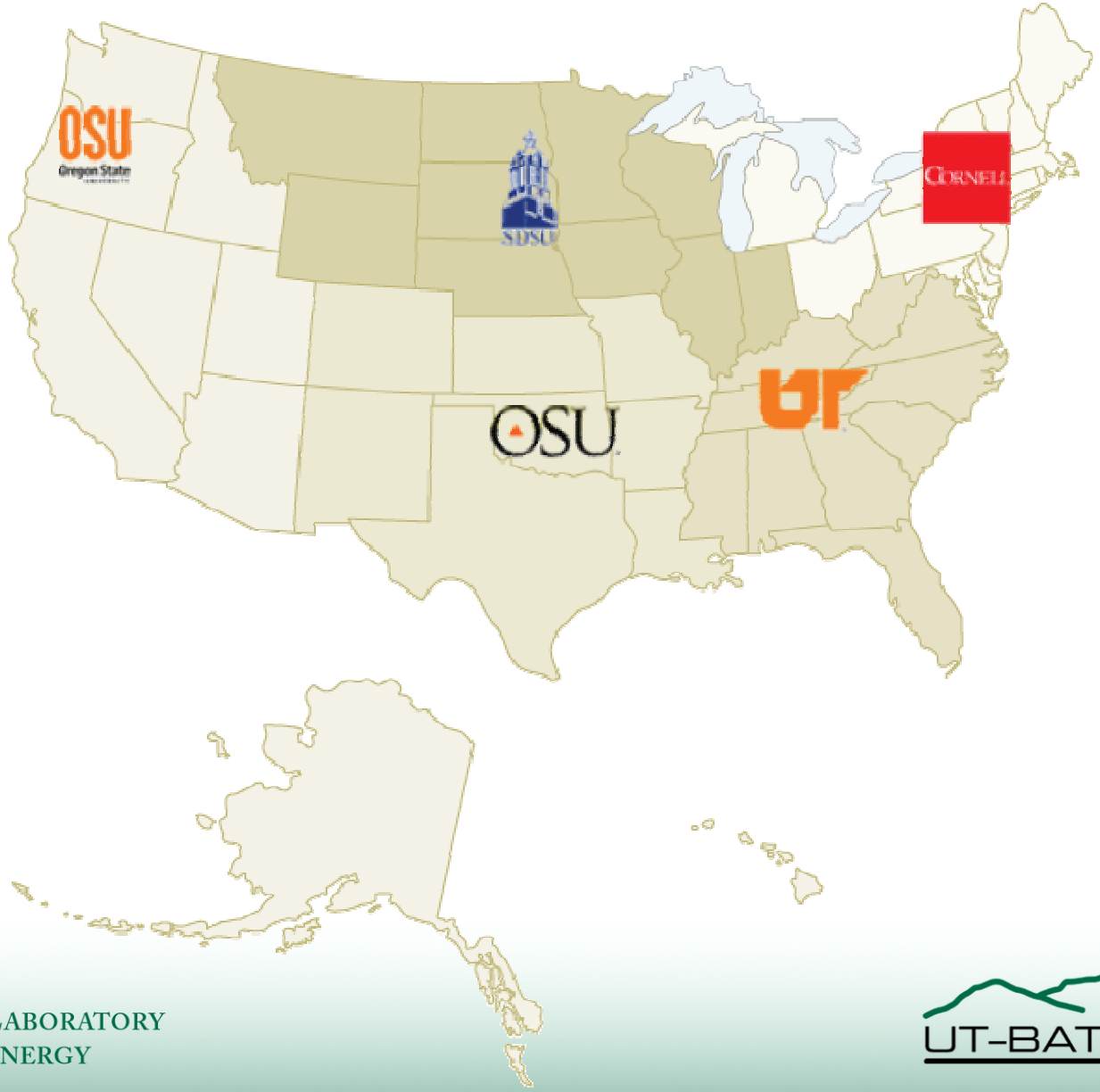
Total: More than $90M

BESC
BioEnergy Science Center

Tennesse Biofuels Initiative

- January 31: Governor Bredesen proposed a significant investment in biofuels
  - More than $70M for feedstock production, R&D, pilot production, fuel distribution
  - $40M for 5M-gal/year pilot plant
- Leverages R&D at ORNL and UT
- Builds on state biomass resources:
  - Corn stover
  - Switchgrass
  - Soybeans
  - Tree residue
  - Poplar

Sun Grant Centers



DOE Selects Six Cellulosic Ethanol Plants for Up to $385 Million in Federal Funding (2/07)

Company	Site	Feed	Process to ethanol
Abengoa	KS	Corn stover, ag residue	Dilute acid/enzyme/ferment for easy sugars; then gasify the residuals to syngas and power shared with dry mill
Iogen with Shell	ID	Ag residue: wheat straw	Dilute acid, enzyme, ferment
Range Fuels (Kergy), Merrick, CH2MHill, Khosla	GA	Wood residues	Thermochemical (gasify or pyrolysis) then catalysis to liquids (ethanol and methanol)
Broin with DuPont, Novozyme, NREL	IA	Corn fiber, cobs, stalks	Dilute acid, fermentation Shared with dry mill
Alico, BRI, etc.	FL	Yard waste, citrus waste, energy cane	Gasify followed by fermentation
BlueFire with Waste Mgmt, Petrodiamond	So. CA	Yard waste, wood and green waste,	Conc. acid to sugars, then fementation

Summary

- Significant opportunity exists for increased penetration of biofuels and for establishing a biofuels industry in the state
- Strong scientific base from which to solve significant challenges in cost-effective production of cellulosic ethanol and other biofuels
- Significant federal R&D investments are being competed by DOE Office of Science, and by DOE EERE
- Partnership of States/academia/ORNL/industry with farmers is a critical factor in our collective success

Panel Consensus from 27th Symposium on Biotechnology for Fuels and Chemicals

- We need commercialization and policy support for these current and near-term opportunities to grow the industry from their current base.
- Equally important, we need R&D to increase the impact, the efficiency, the sustainability, and the ability to add value to markets and farmers while preparing for possible long-term limits on land use.
- Corn ethanol is the starting point for this domestic industry and can expand somewhat further; <u>however, lignocellulosic ethanol is ready to be developed in niche markets now and substantially improved for widespread use.</u>