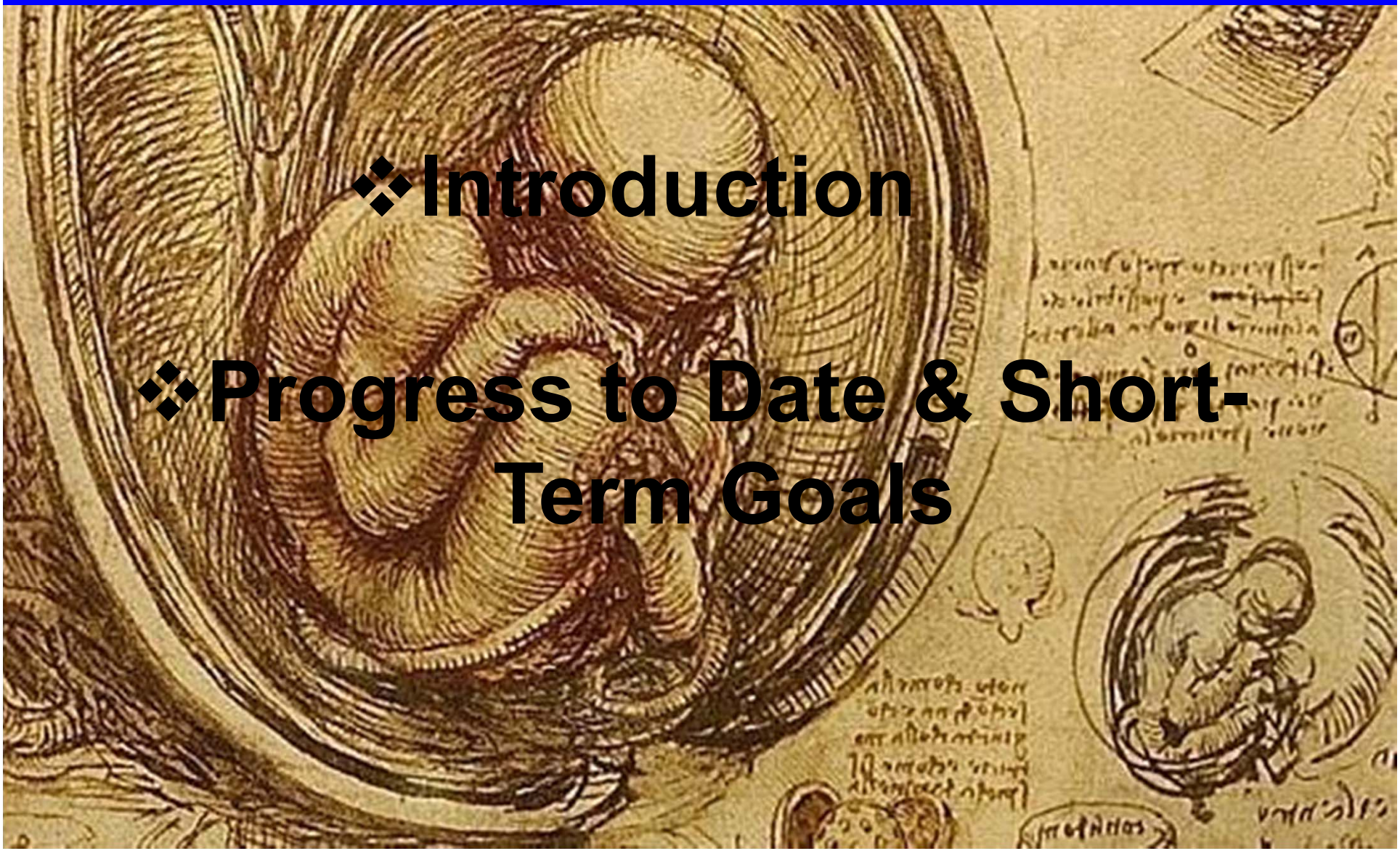
The Evolutionary Synthesis of Human Pregnancy



- Introduction
- Progress to Date & Short-Term Goals

The Central Mission

To synthesize the evolutionary portrait of human pregnancy to identify candidate genetic and environmental disruptors

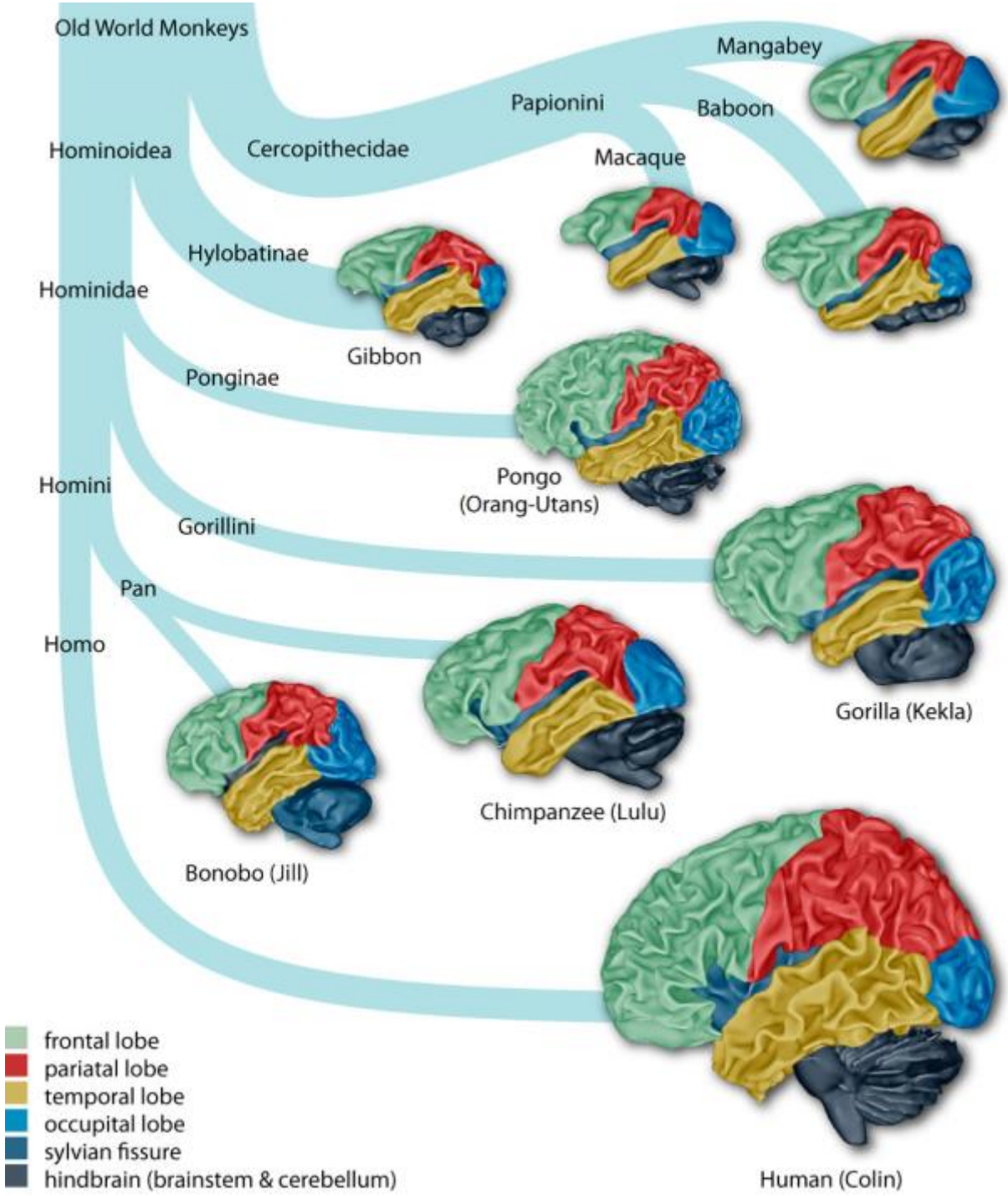
- **Encyclopedia**
- **Dynamic Portal for Data Retrieval & Analysis**
- **Predictive Algorithms**

Parturition is One of the Most Dramatically Evolved Human Traits

Australopithecus
3.2 mya
230 ml

Homo erectus
1.2 mya
315 ml

Homo sapiens
Present
350 ml


Old World Monkeys
Mangabey
Papionini
Baboon
Hominoidea
Cercopithecidae
Macaque
Hylobatinae
Hominidae
Gibbon
Ponginae
Pongo
(Orang-Utans)
Homini
Gorillini
Pan
Homo
Gorilla (Kekla)
Chimpanzee (Lulu)
Bonobo (Jill)
Human (Colin)
frontal lobe
pariatal lobe
temporal lobe
occupital lobe
sylvian fissure
hindbrain (brainstem & cerebellum)

Chimpanzee

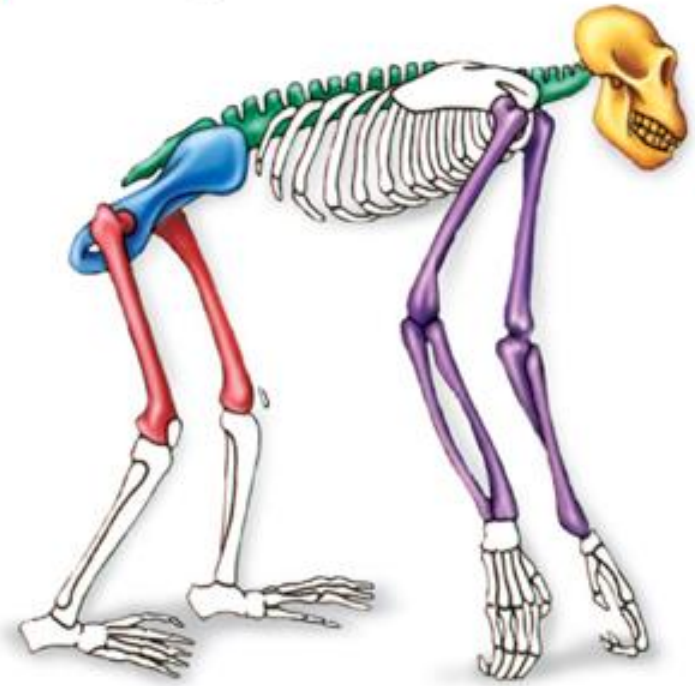
Skull attaches posteriorly

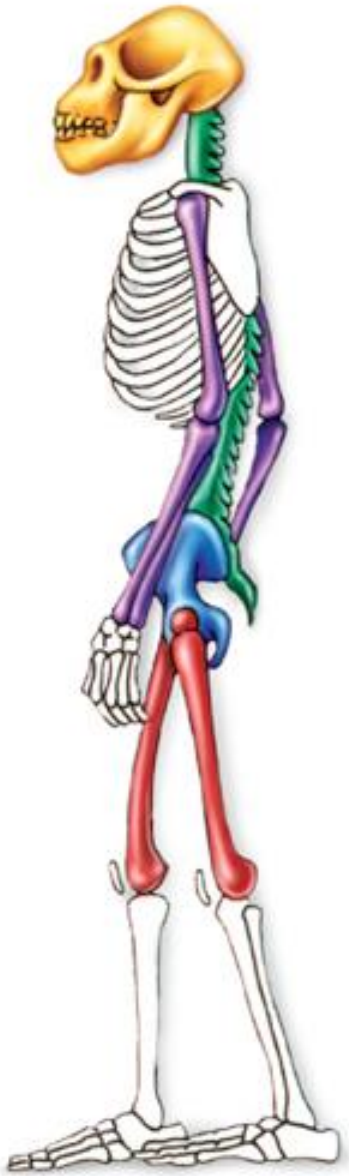
Spine slightly curved

Arms longer than legs and also used for walking

Long, narrow pelvis

Femur angled out





Australopithecine

Skull attaches inferiorly

Spine S-shaped

Arms shorter than legs and not used for walking

Bowl-shaped pelvis

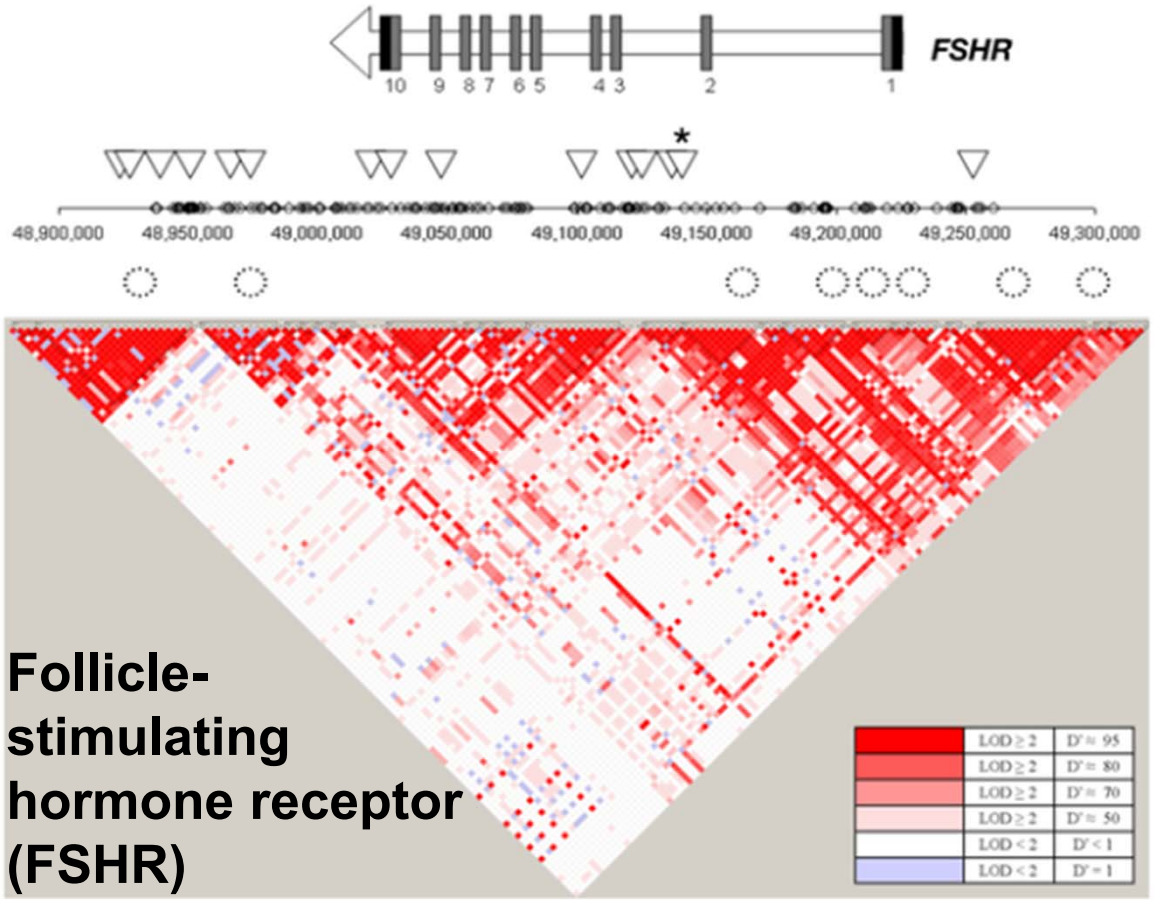
Femur angled in

The Promise of an Evolutionary Approach…

Circumscribe study group (e.g., primate species) → **Identify shared genes** → **Identify genes likely to have undergone adaptive evolution in humans**

↓

Test promising candidates for involvement in human birth timing



Follicle-stimulating hormone receptor (FSHR)

Plunkett et al. (2011) PLoS Genetics

Change Is Everywhere!

Many Genetic Changes

- SNPs
- Indels
- Gene gains / losses
- CNVs
- Domain gains / losses
- Methylation sites

Many Phenotypic Changes

- Gene Expression
- Protein Abundance
- Hormone Levels
- Placental morphology
- Physiology
- Immunity

Many Environments

- Stress
- Medical Care
- Geography
- Nutrition
- Microbiome

Many Analytical Tools

- dN/dS ratio
- MK test
- HKA test
- Tajima's D
- ω statistic

The 3 Key Challenges

❖ Synthesis

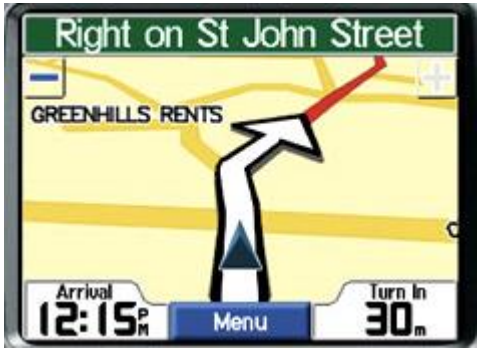
how do we integrate all this information?

❖ Accessibility

how do we make this synthesis available?

❖ Actionability

how do we make it useful?



The Aims of Theme 1

AIM 1. The Development of Artemis

To construct an integrated evolutionary encyclopedia for the deposition, analysis and prediction of genetic factors underlying pre-term birth

AIM 2. The Development of a Dynamic Web Portal and Associated Computational Tools

To enable investigators worldwide to intelligently search the encyclopedia and identify high-quality candidate genes

AIM 3. The Identification and Classification of the Candidate Genes and Genetic Networks that Influence Pre-Term Birth

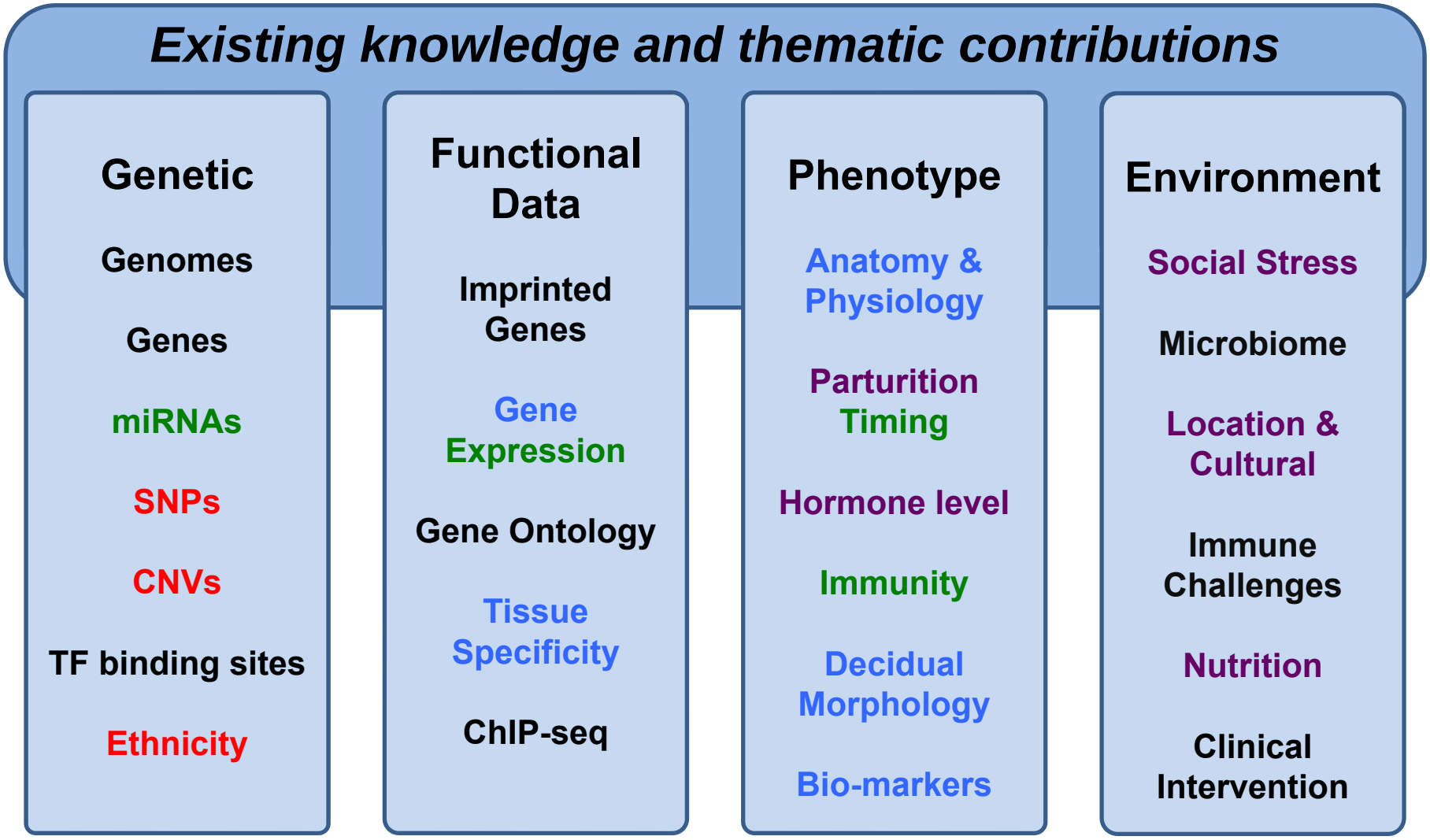
To perform cutting edge research on PTB using Artemis

Aim 1. Create the Map

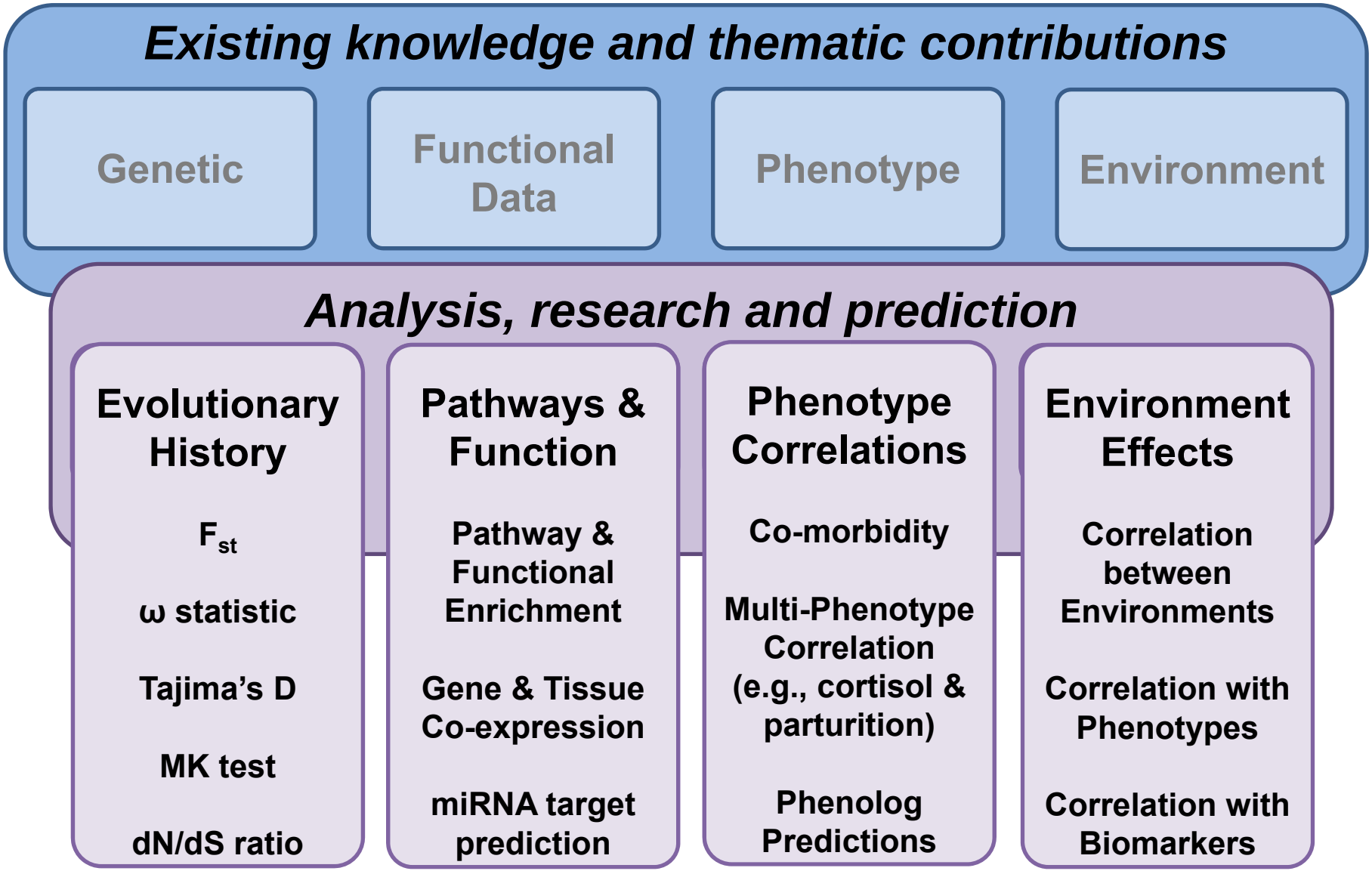
Existing knowledge and thematic contributions

Genetic	Functional Data	Phenotype	Environment
Genomes	Imprinted Genes	Anatomy & Physiology	Social Stress
Genes	Gene Expression	Parturition Timing	Microbiome
miRNAs	Gene Ontology	Hormone level	Location & Cultural
SNPs	Tissue Specificity	Immunity	Immune Challenges
CNVs	ChIP-seq	Decidual Morphology	Nutrition
TF binding sites		Bio-markers	Clinical Intervention
Ethnicity			

Aim 1. Create the Map



Aim 1. Create the Map



Aim 1. Create the Map

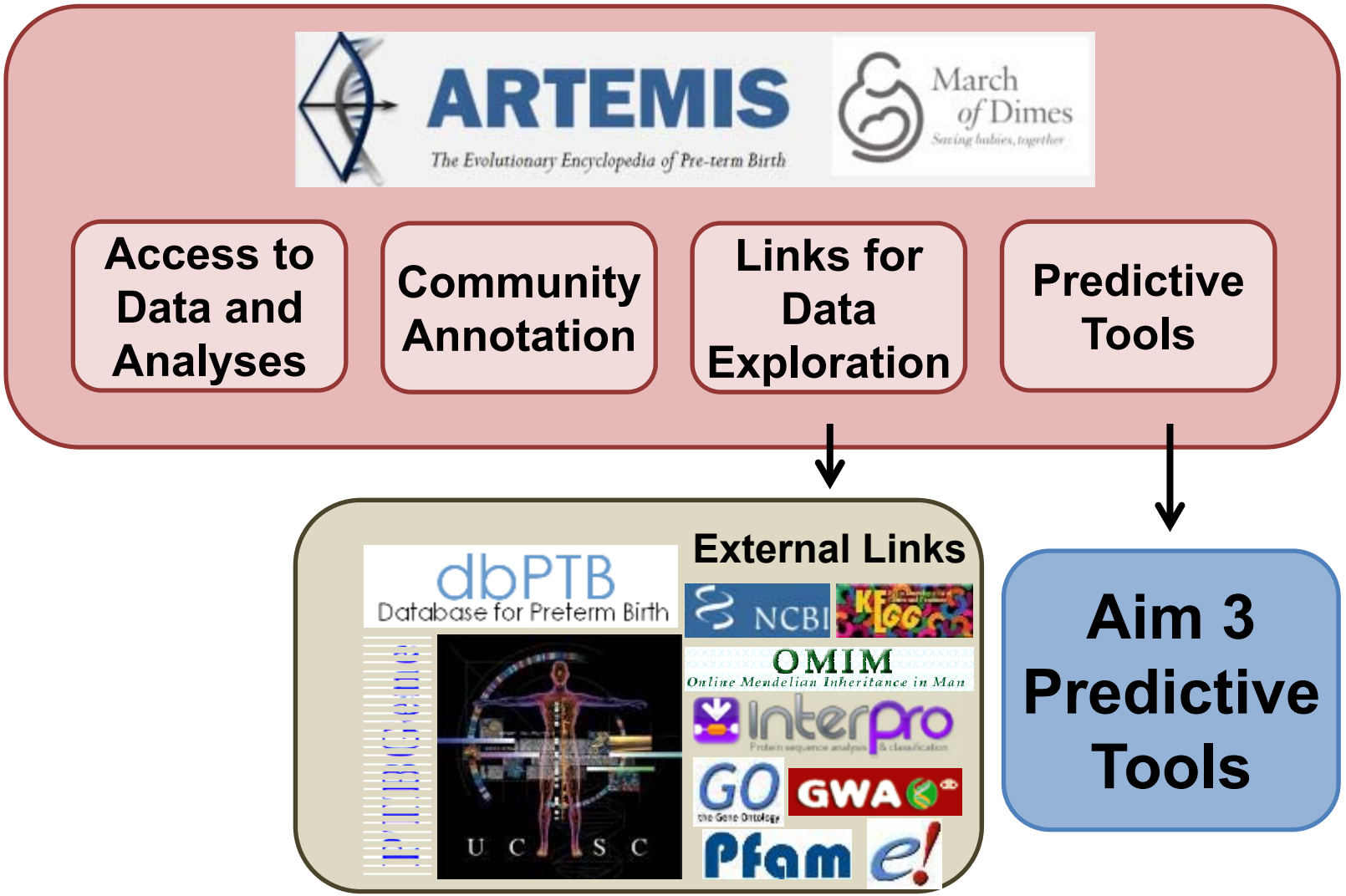
Existing knowledge and thematic contributions

- Genetic
- Functional Data
- Phenotype
- Environment

Analysis, research and prediction

- Evolutionary History
- Pathways & Function
- Phenotype Correlations
- Environment Effects

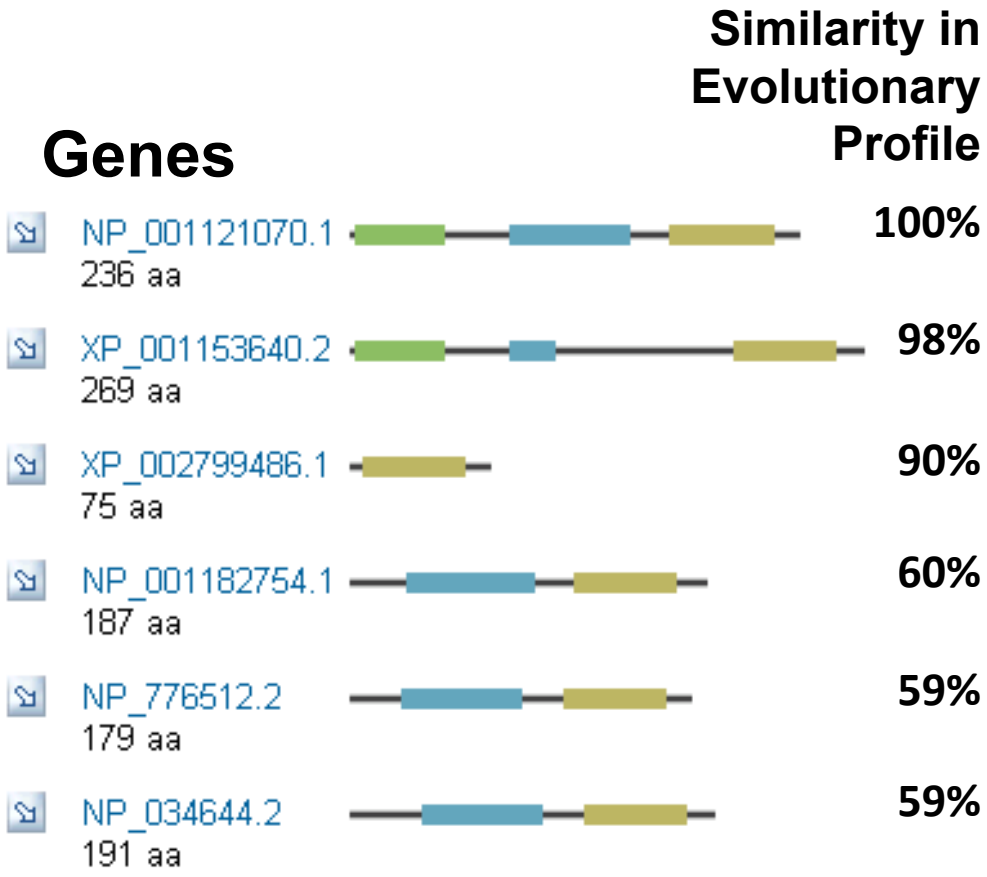
Aim 2. Make Access Simple



Aim 3. Give Clear Direction

GeneGoogle

Find genes with similar evolutionary and functional characteristics to a known parturition gene



Aim 3. Give Clear Direction

GeneGoogle

Find genes with similar evolutionary and functional characteristics to a known parturition gene

RateMyGenes

Identify and rank de novo candidate genes using user specified evolutionary and functional criteria

Infrastructure

VANDERBILT UNIVERSITY

ACCRE

Advanced Computing Center for Research & Education

Enabling Researcher-Driven Innovation and Exploration

Compute Cores	Diskspace	Memory
4,000+	500+ TB	128 GB

ARTEMIS Runs on GMOD

- **GMOD (Generic Model Organism Database)** Vast collection of open-source software tools for creating and managing genome-scale biological databases

CHADO

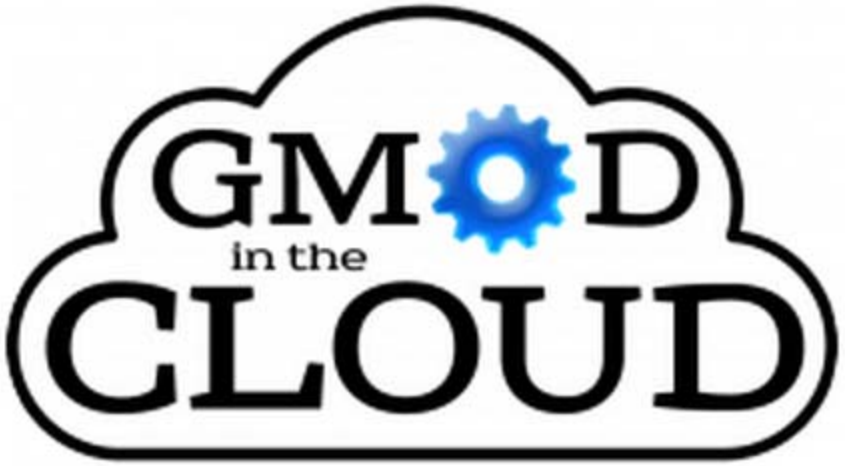
Chado: Biological database schema

- **A relational database schema capable of representing the general classes of data frequently encountered in modern biology such as sequence, sequence comparisons, phenotypes, genotypes, ontologies, publications, and phylogeny**

- **Designed to handle complex representations of biological knowledge; considered to be one of the most sophisticated relational schemas currently available in molecular biology**

Sequence	**Organism**
Expression	**Phylogeny**
Genetic	**Computational Analysis**
Phenotype	**Controlled Vocabulary**
Publication	**Stock**

MAKER: Genome annotation pipeline



GMOD in the Cloud toolset

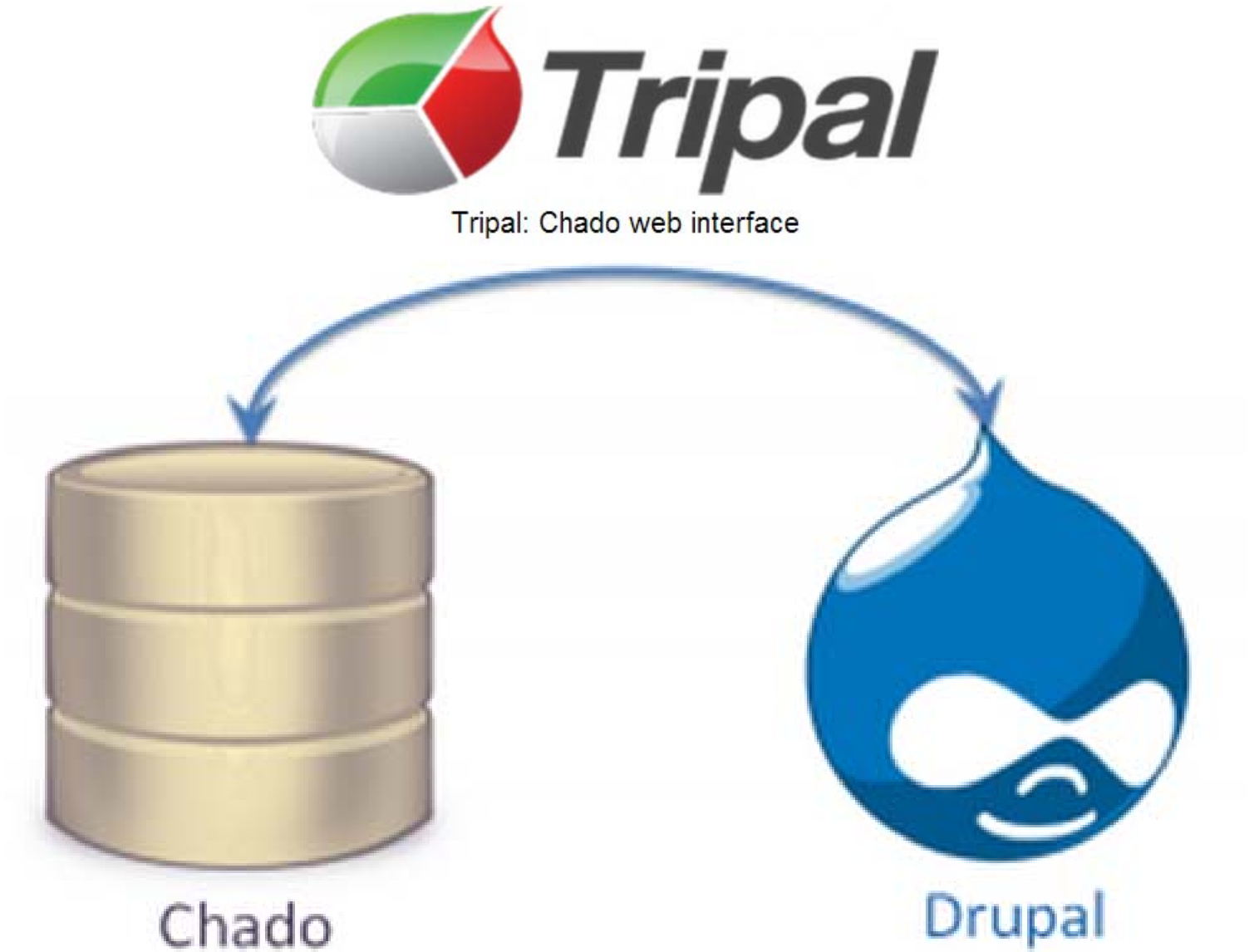
BioMart: Data mining system

JBrowse: Super-fast genome annotation viewer

Galaxy: Data analysis & integration

Pathway Tools: Metabolic, regulatory pathways

Back End Front End



Cacao Genome Database

Login

Home | Projects | Tools | Databases | Resources | Mailing List | Search | Download | About

The Genomics, Genetics and Breeding Resource for Cacao Improvement

A collaboration among MARS, USDA-ARS, IBM, NCGR, Clemson University, HudsonAlpha Institute for Biotechnology, Indiana University and Washington State University

Welcome to the Cacao Genome Project

Cacao production is important! Not only is it the basic ingredient in the world's favorite confection, **chocolate**, but it provides a livelihood for over 6.5 million farmers in Africa, South America and Asia and ranks as one of the top ten agriculture commodities in the world. Historically, cocoa production has been plagued by serious losses due to pests and diseases. The release of the cacao genome sequence will provide researchers with access to the latest genomic tools, enabling more efficient research and accelerating the breeding process, thereby expediting the release of superior cacao cultivars. The sequenced genotype, Matina 1-6, is representative of the genetic background most commonly found in the cacao producing countries, enabling results to be applied immediately and broadly to current commercial cultivars. Matina 1-6 is highly homozygous which greatly reduces the complexity of the sequence assembly process. While the sequence provided is a preliminary release, it already covers 92% of the genome, with approximately 35,000 genes. We will continue to refine the assembly and annotation, working toward a complete finished sequence. Updates will be made available via this website, to keep informed check back regularly or join CGD mailing list. If you have any questions, feedback or problems, please contact us through the contact link on the navigation bar, we promise to respond in a timely manner.

News

- CacaoCyc version 1.0 is available on CGD!
- Public and scientific interest in the chocolate genome continues to grow!
- Cacao Genome Database Computer Demo at PAG 2012
- Cacao Genome Sequencing and Breeders Workshop at PAG 2012
- The Cacao Genome Sequence is released 3 years ahead of schedule!
- More announcements ...

MARS incorporated | USDA ARS Agricultural Research Service | IBM | CUGI | HudsonAlpha Institute for Biotechnology | NCGR | PIPRA | Indiana University | Washington State University

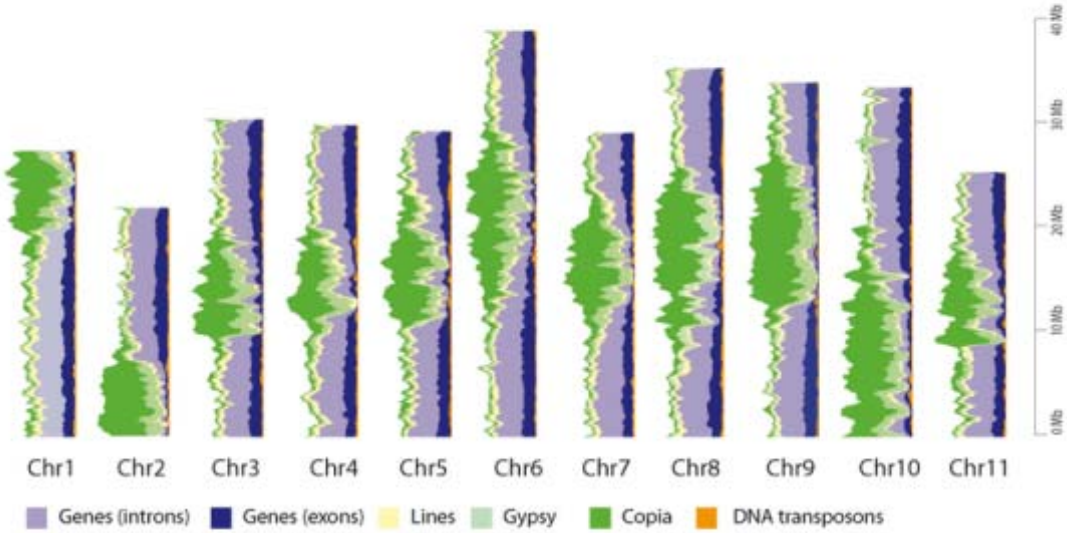
The Banana Genome Hub

Search

HOME | GENOME DETAILS | TOOLS | GBROWSE | BLAST | DOWNLOAD | DOCUMENTATION

The **Musa genome sequence** results from collaboration between Genoscope and Cirad (UMR AGAP) funded by ANR. The sequenced genotype is a doubled-haploid (2n=22, 1C=523 Mb) from the species Musa acuminata (A genome) subspecies malaccencis. The doubled-haploid (DH-Pahang) was produced at Cirad through anther culture of the wild diploid accession Pahang and spontaneous chromosomes doubling. The wild Pahang accession originated from Central Malaysia.

The sequence was analysed in collaboration with several teams in particular of the Global Musa Genomics Consortium and was published in : "The banana (Musa acuminata) genome and the evolution of monocotyledonous plants". D'Hont et al. 2012. Nature



Supported by:

The **Banana Genome Hub** centralises databases of genetic and genomic data for the musa acuminata crop Hub developed by Cirad and Bioversity and supported by the South Green Bioinformatics platform. Data available are the complete genome sequence along with gene structure, gene product information, metabolism, gene families, transcriptomics (ESTs, RNA-Seq), genetic markers (SSR, DArT, SNPs) and genetic maps.

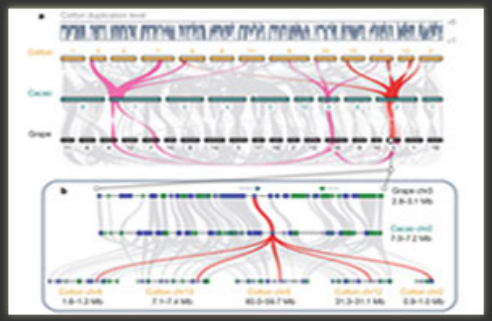
Sequence Contact:
dhont(@)cirad.fr
wincker(@)genoscope.cns.fr

Hub Contact:
droc(@)cirad.fr

CottonGen

a genomics, genetics and breeding resource for cotton

Publication



Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres

A phylogenetic and genomic study of plants of the cotton genus Gossypium provides insights into the role of polyploidy in the angiosperm evolution. Paterson, et al., 2012. Nature 492, 423 -427.

Welcome to CottonGen

CottonGen is a new cotton community genomics, genetics and breeding database being developed to enable basic, translational and applied research in cotton. It is being built using the open-source Tripal database infrastructure. CottonGen consolidates and expands the data from CottonDB and the Cotton Marker Database, providing enhaced tools for easy querying, visualizing and downloading research data. This project is funded by Cotton Incorporated, the USDA-ARS Crop Germplasm Research Unit at College Station, TX, the Southern Association of Agricultural Experiment Station Directors, Bayer CropScience, Dow/Phytogen, Monsanto and Washington State University.

What's New in CottonGen?

- The JGI *G. Raimondii* genome sequence (V2.0) and annotation (V2.1). Browse, search and download the genome sequence, predicted genes, homologs, markers, pathways and BLAST your sequences.
- Report a problem | Ask us a question | Post a job | Post a meeting or event | What's been added/fixed in CottonGen? | What are we working on? | Used CottonGen data or tools in your research? how to reference us!

News

- 2013 Cotton Breeders' Tour, Lubbock, Texas USA
- NCBI cotton sequences updated.
- Search site for Publications added.
- Search site for Quantitative Trait Loci (QTLs) added.
- CottonGen v1.0 released.
- News archive

Progress to Date

- ARTEMIS database structure built; tailoring of modules underway
- Computational pipelines for uploading large-scale datasets (e.g., genomes and annotations) in place; more on the way
- Uploading of genomic data and annotation has begun
- Computational pipelines for standard analyses (e.g., BLAST) in place; many more on the way

Challenges

- All established databases are model organism databases; ARTEMIS is an evolutionary database
- Conceptual design of ARTEMIS front end
- Integration of non-coding region data
- Integration of phenotypic data

Short Term Goals

- Upload mammal genomes and annotation into ARTEMIS
- Construct orthologs and gene families
- Begin uploading functional annotation and literature data
- Begin integration of non-coding region data
- Begin discussions on phenotypic data
- Complete conceptual design of ARTEMIS front end