The group

Basal genera: Dione iuno, Dryas iulia and Podotricha telesiphe

Euiedes lybia, E. tales and E. isabella

Laparus doris (blue and red form) and Neruda aoede

H. wallacei, H. burneyi and H. xanthocles

H. melpomene amaryllis, H. m. aglaope and H. cydno chioneus

H. hecale felix, H. elevatus and H. numata tarapotensis

H. erato favorinus, H. e. emma and H. telesiphe telesiphe
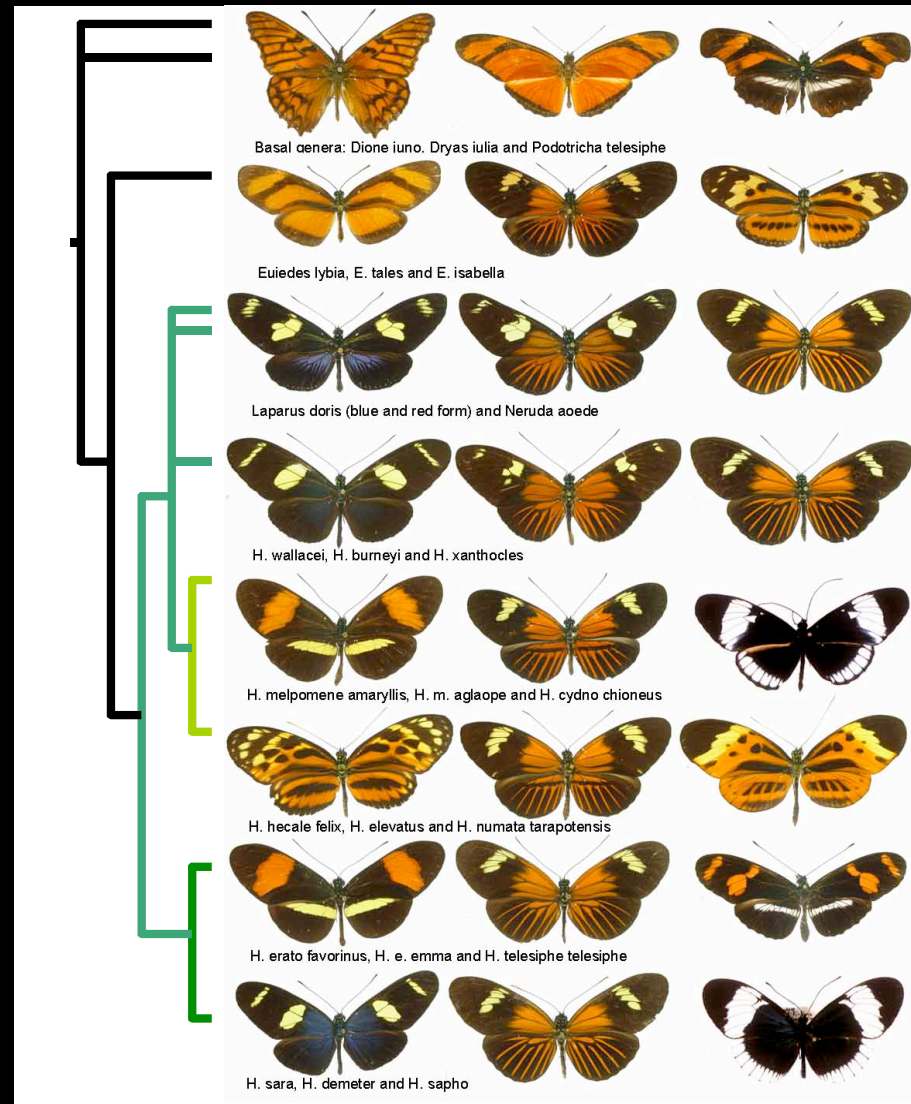
H. sara, H. demeter and H. sapho

Well known adaptive radiation and model system in evolutionary biology

The Challenge

Spatial information

(historical samples, coordinate information, habitat characteristics)

↕

Phenotypic Diversity

[EAV terms, wing phenotypes (quantitative descriptions, images), pedigrees, other morphological traits]
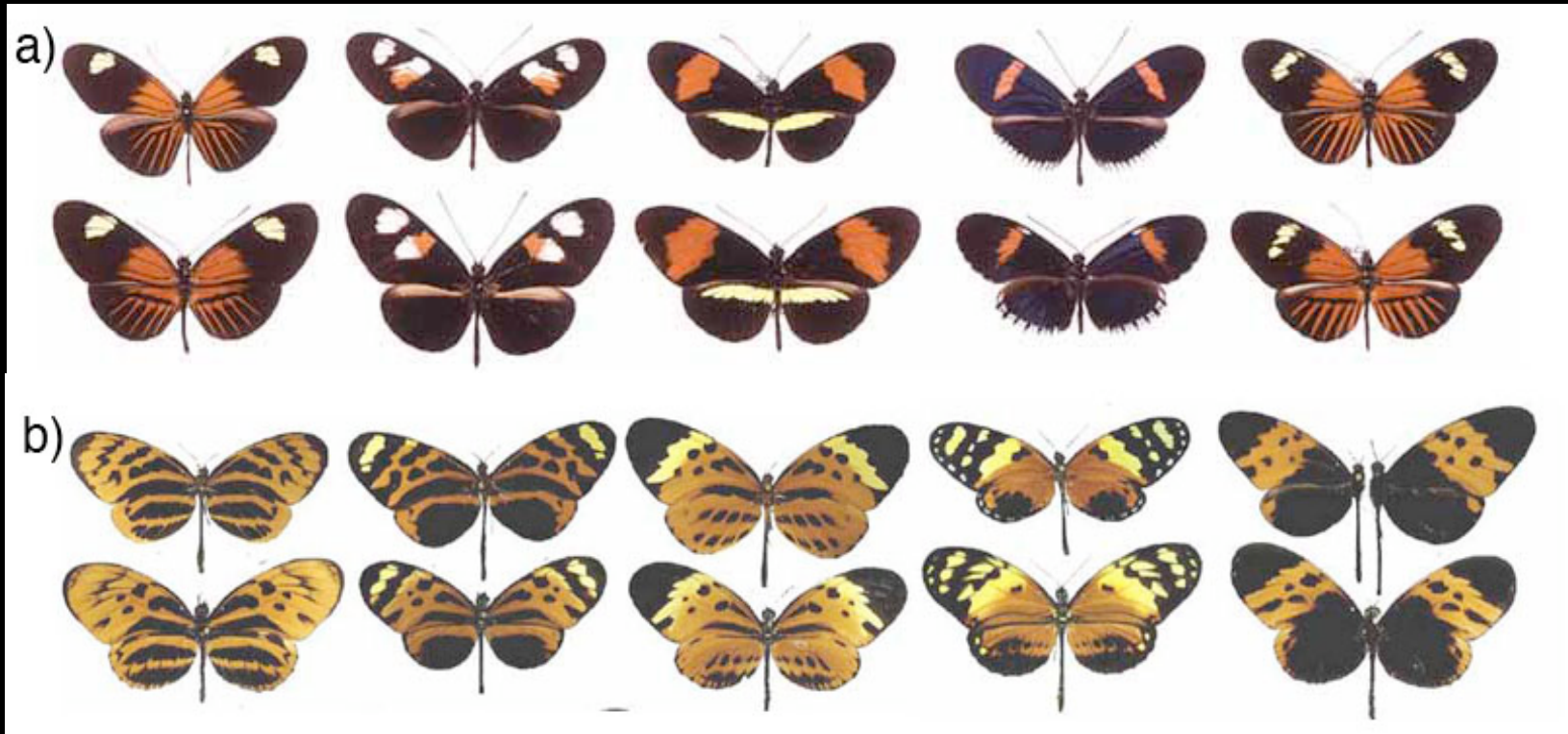
↕

Genomic/Genetic Diversity

(linkage/QTL maps, array data, *Ests*, SNP, BAC sequence)
(more BAC sequence, genome)

How to link information on genomic and phenotypic diversity

Across taxa

Morphological similarity is not always a good predictor of genomic similarity (and vica versa)!

What emerged for the mini-GMOD meeting in January

Diversity rules: Most of the MODs are tackling similar issues.

We can play too: The evolutionary community can help.

We should play: It is simply makes good sense to utilize existing resources and to help develop new ones.

Our task

Explore the utility of GMOD tools for non-genome enabled organisms.

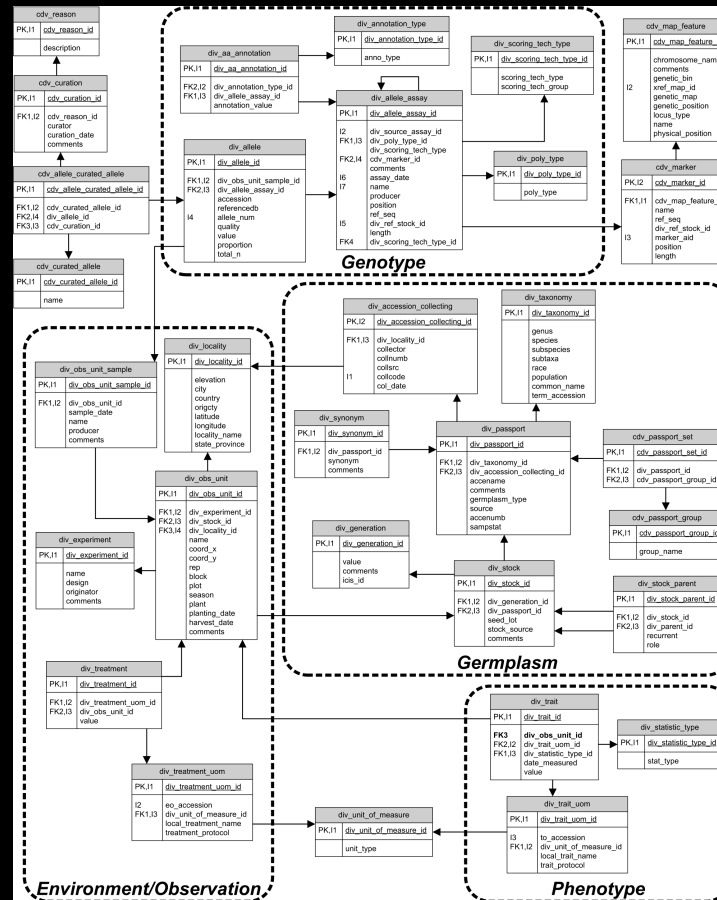**Understand Chado (both its strengths and weaknesses).**

Our task (needs analysis)

We understood Chado at a very general level but needed to understand it at a very detailed level.

We understood our schema needs at a very general level but needed to understand it at a very detailed level.
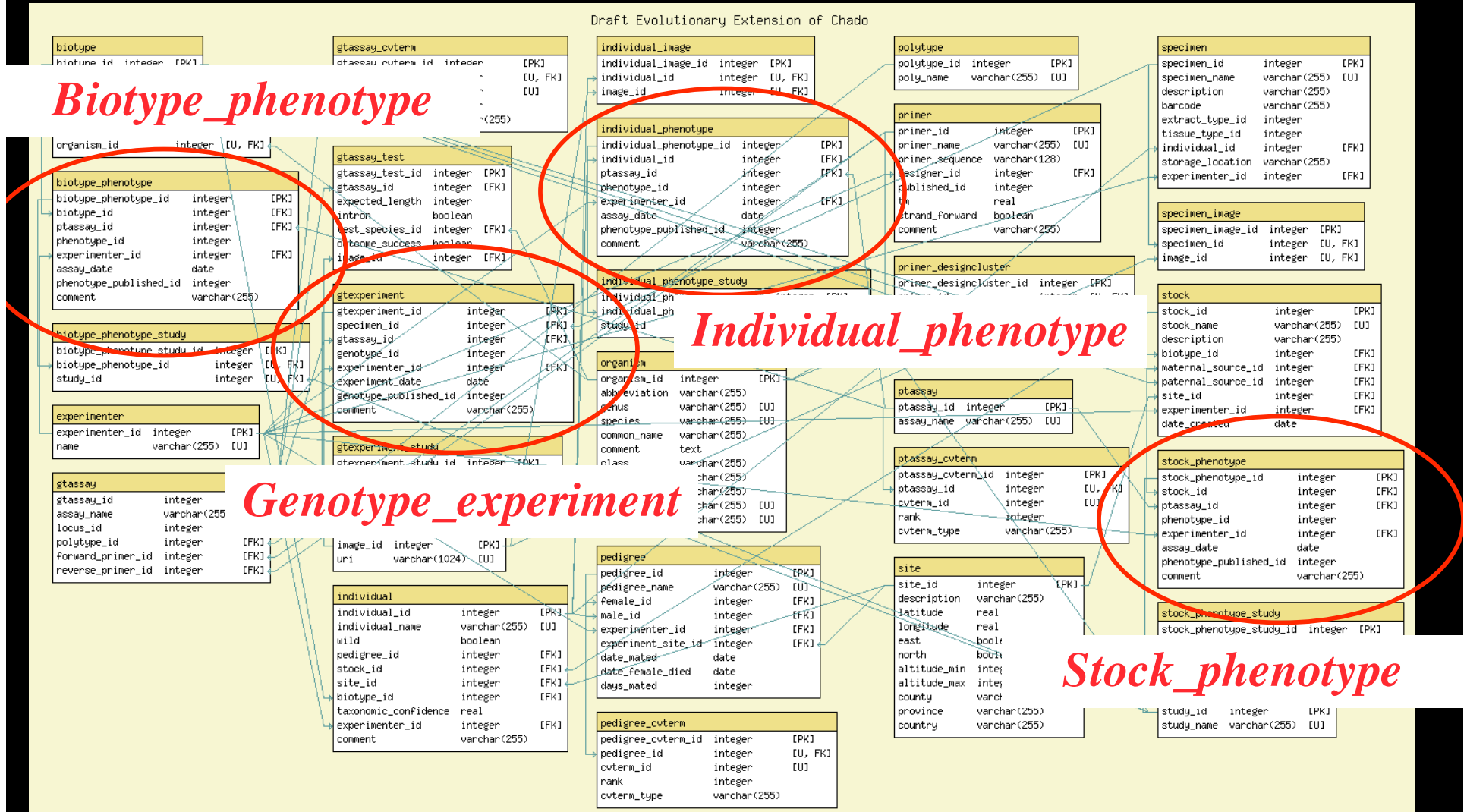
Our working model evolutionary schema

Draft Evolutionary Extension of Chado

90% chadoized (?)

90% complete (?)

biotype
biotype_id	integer	[PK]

biotype_organism
biotype_organism_id	integer	[PK]
biotype_id	integer	[U, FK]
organism_id	integer	[U, FK]

biotype_phenotype
biotype_phenotype_id	integer	[PK]
biotype_id	integer	[FK]
ptassay_id	integer	[FK]
phenotype_id	integer	
experimenter_id	integer	[FK]
assay_date	date	
phenotype_published_id	integer	
comment	varchar(255)	

biotype_phenotype_study
biotype_phenotype_study_id	integer	[PK]
biotype_phenotype_id	integer	[U, FK]
study_id	integer	[U, FK]

experimenter
experimenter_id	integer	[PK]
name	varchar(255)	[U]

gtassay
gtassay_id	integer	[PK]
assay_name	varchar(255)	[U]
locus_id	integer	
polytype_id	integer	[FK]
forward_primer_id	integer	[FK]
reverse_primer_id	integer	[FK]

gtassay_cvterm
gtassay_cvterm_id	integer	[PK]
gtassay_id	integer	[U, FK]

expected_length	integer	
intron	boolean	
test_species_id	integer	[FK]
outcome_success	boolean	
image_id	integer	[FK]

genotype_published_id	integer	
comment	varchar(255)	

gtexperiment_study
gtexperiment_study_id	integer	[PK]
gtexperiment_id	integer	[U, FK]
study_id	integer	[U, FK]

image
image_id	integer	[PK]
uri	varchar(1024)	[U]

individual
individual_id	integer	[PK]
individual_name	varchar(255)	[U]
wild	boolean	
pedigree_id	integer	[FK]
stock_id	integer	[FK]
site_id	integer	[FK]
biotype_id	integer	[FK]
taxonomic_confidence	real	
experimenter_id	integer	[FK]
comment	varchar(255)	

individual_image
individual_image_id	integer	[PK]
individual_id	integer	[U, FK]

experimenter_id	integer	[FK]
assay_date	date	
phenotype_published_id	integer	
comment	varchar(255)	

abbreviation	varchar(255)	
genus	varchar(255)	[U]
species	varchar(255)	[U]
common_name	varchar(255)	
comment	text	
class	varchar(255)	
order	varchar(255)	
family	varchar(255)	
subfamily	varchar(255)	[U]
subspecies	varchar(255)	[U]

pedigree
pedigree_id	integer	[PK]
pedigree_name	varchar(255)	[U]
female_id	integer	[FK]
male_id	integer	[FK]
experimenter_id	integer	[FK]
experiment_site_id	integer	[FK]
date_mated	date	
date_female_died	date	
days_mated	integer	

pedigree_cvterm
pedigree_cvterm_id	integer	[PK]
pedigree_id	integer	[U, FK]
cvterm_id	integer	[U]
rank	integer	
cvterm_type	varchar(255)	

polytype
polytype_id	integer	[PK]
poly_name	varchar(255)	[U]

tm	real	
strand_forward	boolean	
comment	varchar(255)	

primer_designcluster

ptassay
ptassay_id	integer	[PK]
assay_name	varchar(255)	[U]

ptassay_cvterm
ptassay_cvterm_id	integer	[PK]
ptassay_id	integer	[U, FK]
cvterm_id	integer	[U]
rank	integer	
cvterm_type	varchar(255)	

site
site_id	integer	[PK]
description	varchar(255)	
latitude	real	
longitude	real	
east	boolean	
north	boolean	
altitude_min	integer	
altitude_max	integer	
county	varchar(255)	
province	varchar(255)	
country	varchar(255)	

specimen
specimen_id	integer	[PK]
specimen_name	varchar(255)	[U]
description	varchar(255)	
barcode	varchar(255)	
extract_type_id	integer	
tissue_type_id	integer	
individual_id	integer	[FK]
storage_location	varchar(255)	
experimenter_id	integer	[FK]

specimen_image
specimen_image_id	integer	[PK]
specimen_id	integer	[U, FK]
image_id	integer	[U, FK]

stock
stock_id	integer	[PK]
stock_name	varchar(255)	[U]
description	varchar(255)	
biotype_id	integer	[FK]
maternal_source_id	integer	[FK]
paternal_source_id	integer	[FK]
site_id	integer	[FK]
experimenter_id	integer	[FK]
date_created	date	

stock_phenotype
stock_phenotype_id	integer	[PK]
stock_id	integer	[FK]
ptassay_id	integer	[FK]
phenotype_id	integer	
experimenter_id	integer	[FK]
assay_date	date	
phenotype_published_id	integer	
comment	varchar(255)	

stock_phenotype_study
stock_phenotype_study_id	integer	[PK]
stock_phenotype_id	integer	[U, FK]
study_id	integer	[U, FK]

study
study_id	integer	[PK]
study_name	varchar(255)	[U]

Legend
[FK] Foreign Key
[U]   Unique constraint
[PK] Primary key

Created by SQL::Translator 0.

Why not just go with GDPDM?

Interpreting a schema can be a big job for a biologist that doesn't understand relational databases and a computer scientist that doesn't understand biology.
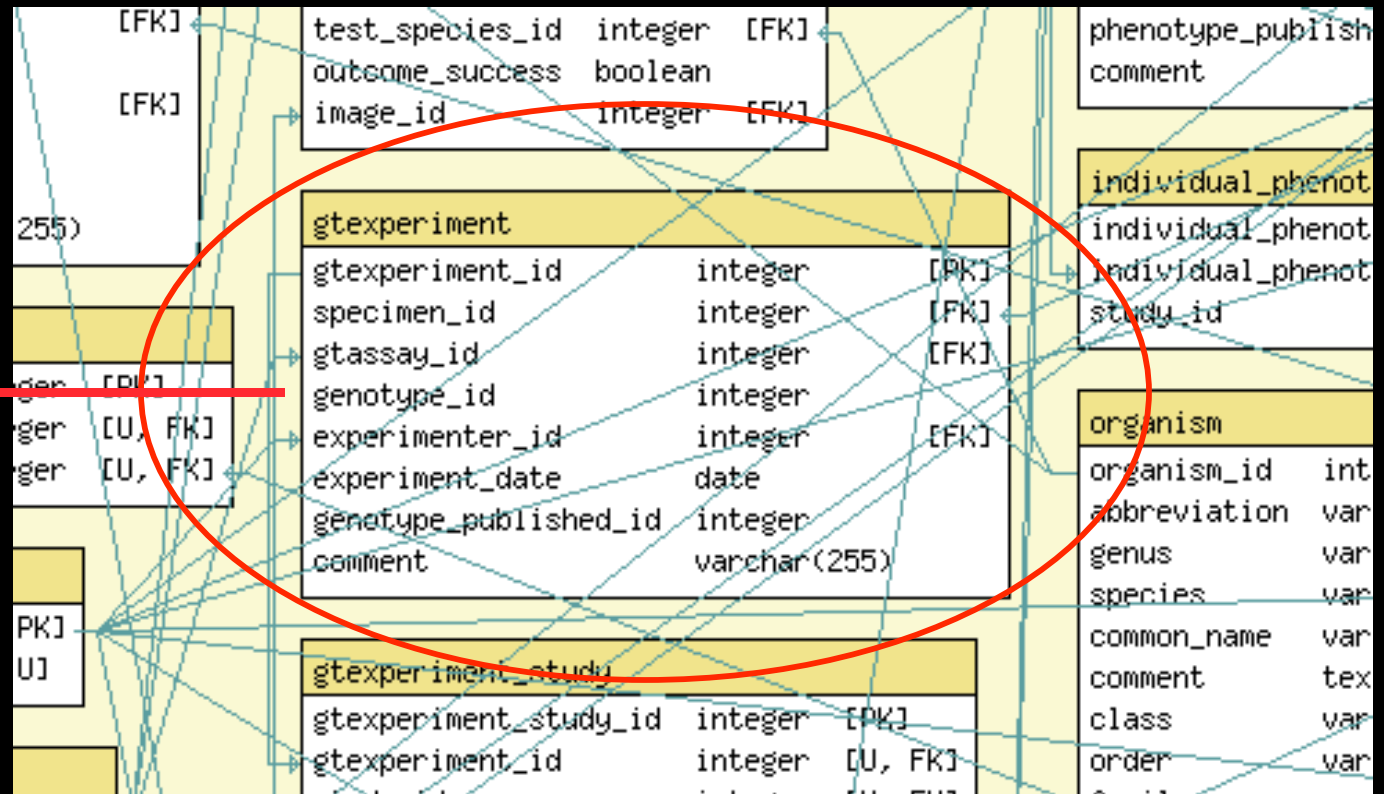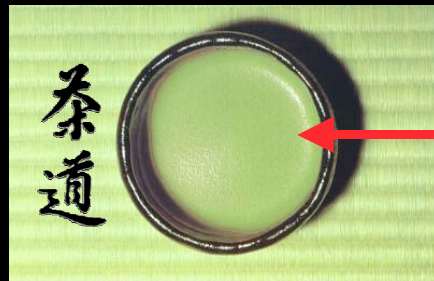
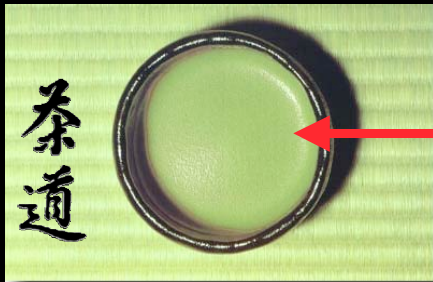# *Focal Points*

Draft Evolutionary Extension of Chado

Major join tables that make associations genotype and phenotype associations

Genotype_experiment
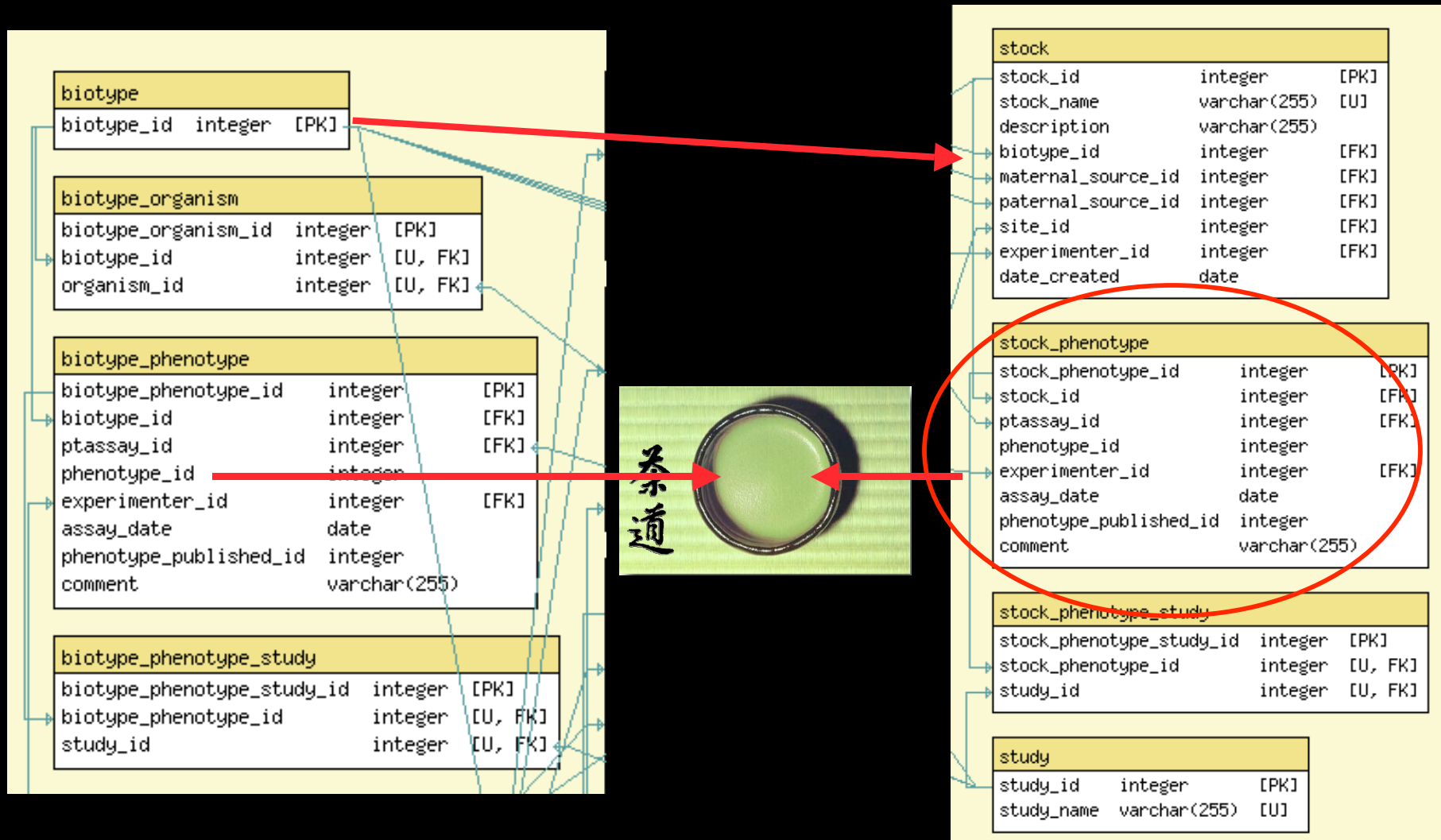

Snap shot of genotype experiment

Individual_phenotype

individual_phenotype		
individual_phenotype_id	integer	[PK]
individual_id	integer	[FK]
ptassay_id	integer	[FK]
phenotype_id	integer	
experimenter_id	integer	[FK]
assay_date	date	
phenotype_published_id	integer	
comment	varchar(255)	

Record of phenotypic information on an individual

Stock and biotype phenotype

Deal with different aspects of phenotype and phenotype experiments

What next?

What next?

Is this an efficient strategy?

Or should well stick with individual, taxon specific schema and develop middleware and mapping tools to integrate them?