China National GeneBank 国家基因库

华大生命科学研究院 BGI·Research

Starting a comparative genome study from CNGBdb

"Raising comparative genomic research idea in the big data era"

Jieyu (Jerry) WANG

王洁雨

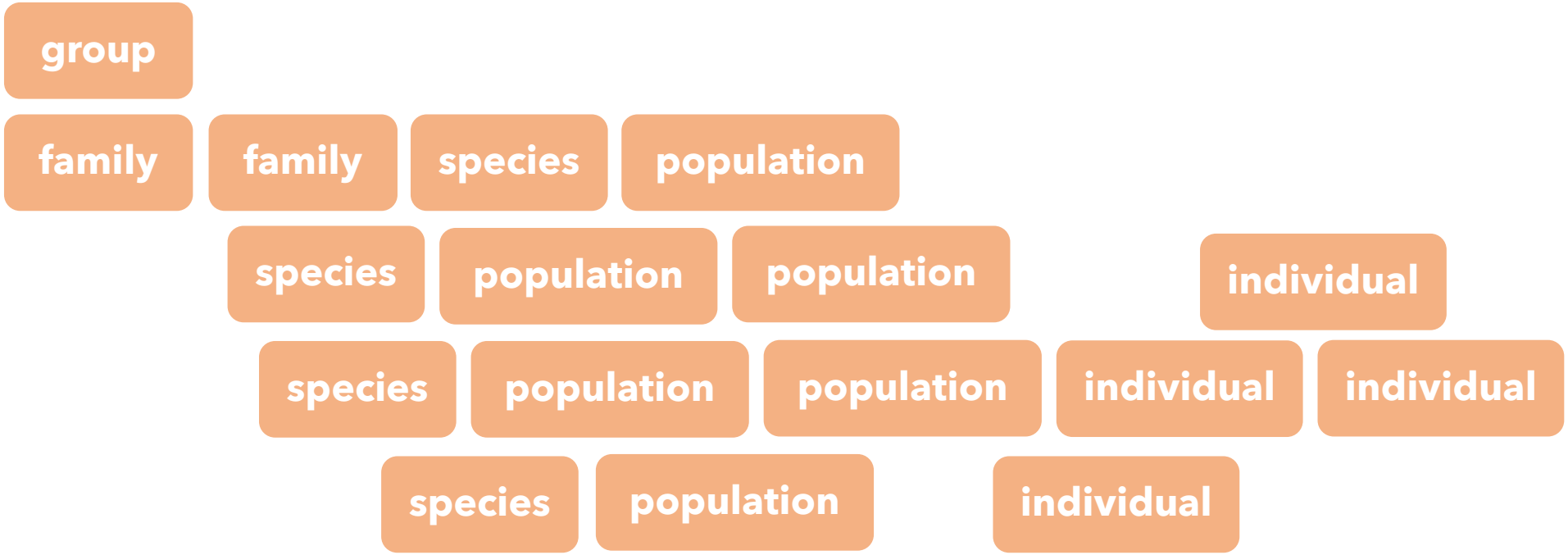
2023.9

what is compare?

“to examine or look for the difference between two or more things”

—— Cambridge Dictionary

what is compare?



“to examine or look for the difference between two or more things”

—— Cambridge Dictionary

before genome era

PCR

sanger

sequence

repeats

compare

genes

Phylogenetic reconstruction

Element evolution

in big data era

NGS

long reads

genome

Hi-C

chromosome level

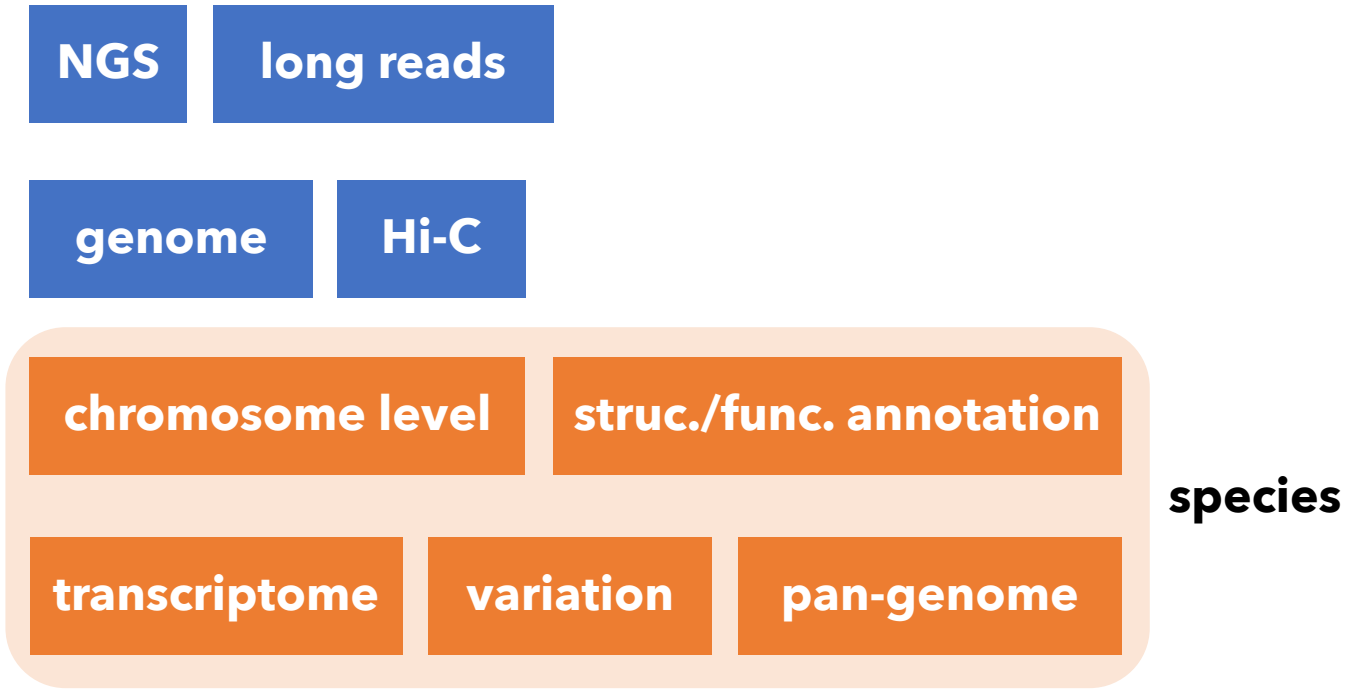
struc./func. annotation

transcriptome

variation

pan-genome

in big data era



in big data era

chromosome level | struc./func. annotation

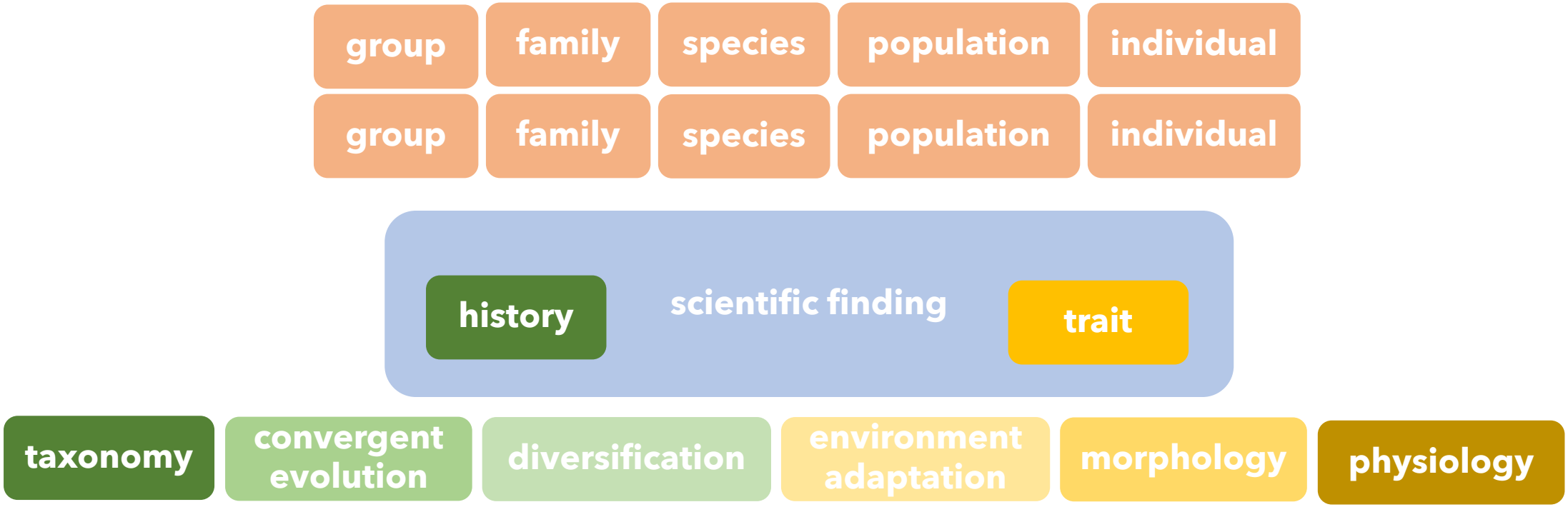
transcriptome | variation | pan-genome

species

species | species | species | species | species | species

scientific finding

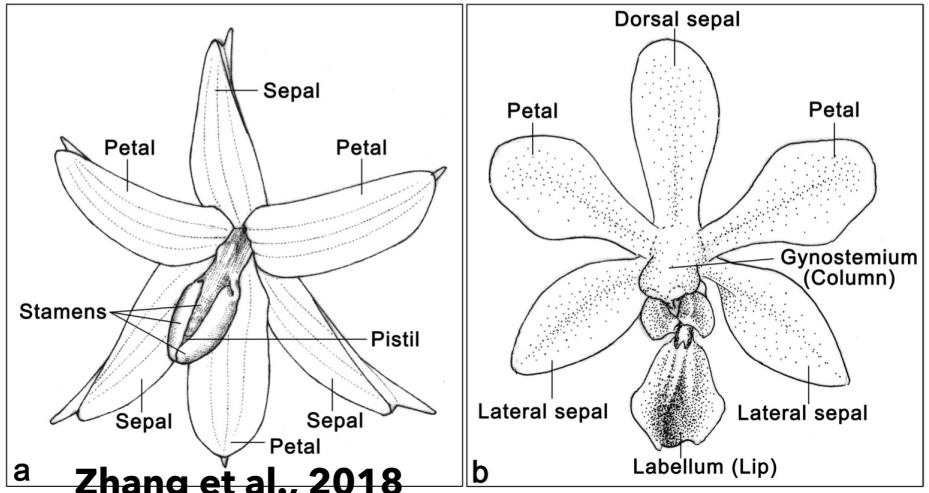
in big data era



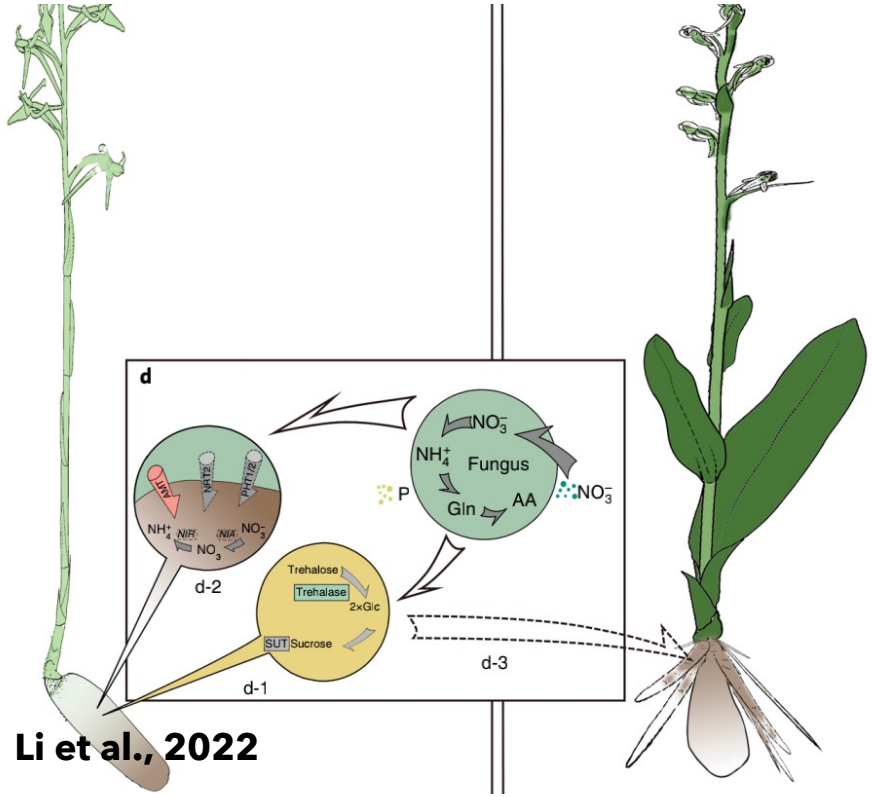
"to examine or look for the difference between two or more things"

—— Cambridge Dictionary

find the similarity and difference



Zhang et al., 2018



Li et al., 2022

Moon et al., 2023

what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

whole genome duplication

calibration

event

biogeography

gene family evolution

function evolution

adaptation

diversification

what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

whole genome duplication

calibration

event

biogeography

gene family evolution

function evolution

adaptation

diversification

phylogenetic reconstruction

single-copy | SNPs | plastid

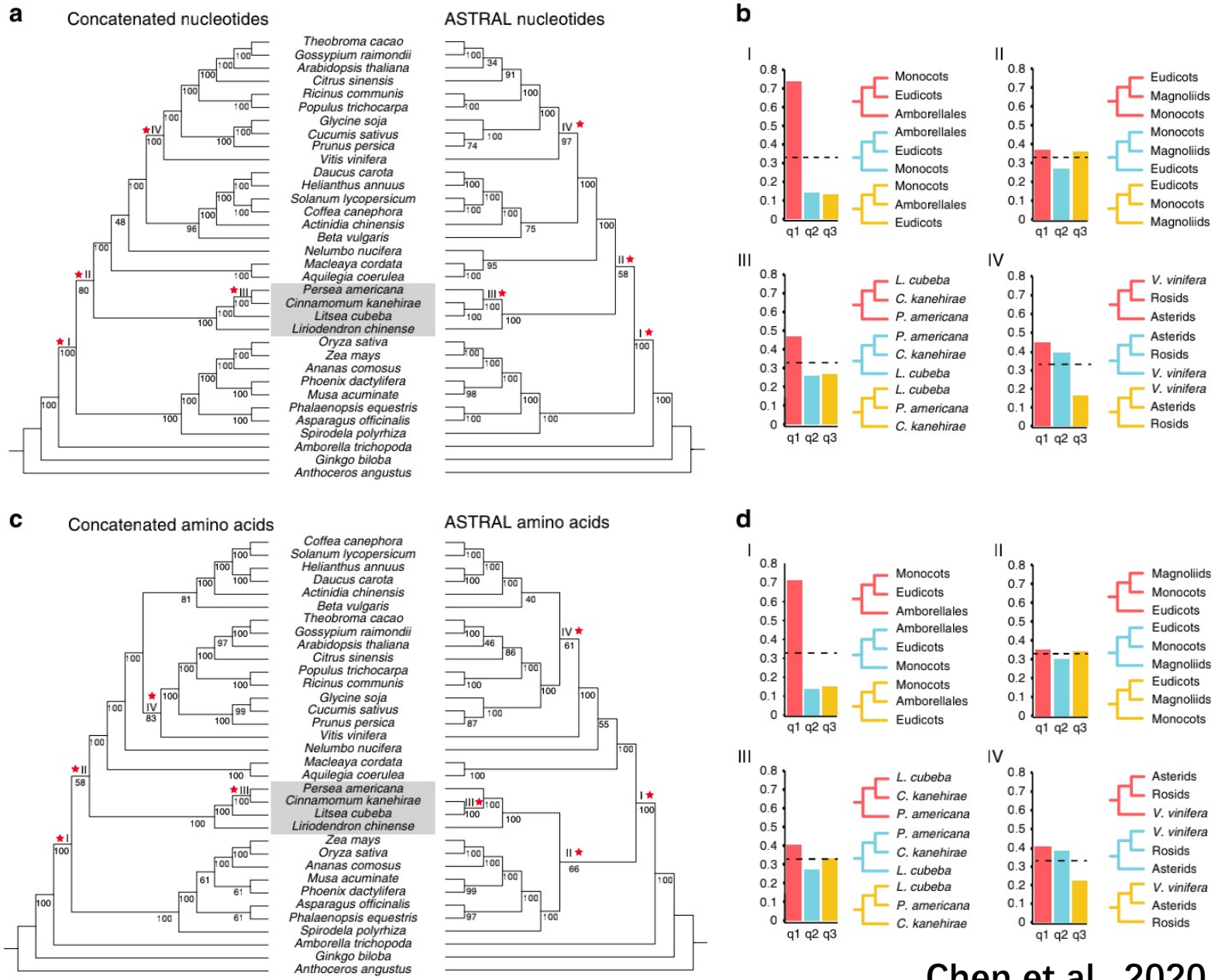
concatenated

multi-species coalescent

whole genome info.

MAFFT | clustalx | MUSCLE

astral | raxml | iqtree



Chen et al., 2020

what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

whole genome duplication

calibration

event

biogeography

gene family evolution

function evolution

adaptation

diversification

biogeography

phylogenetic tree

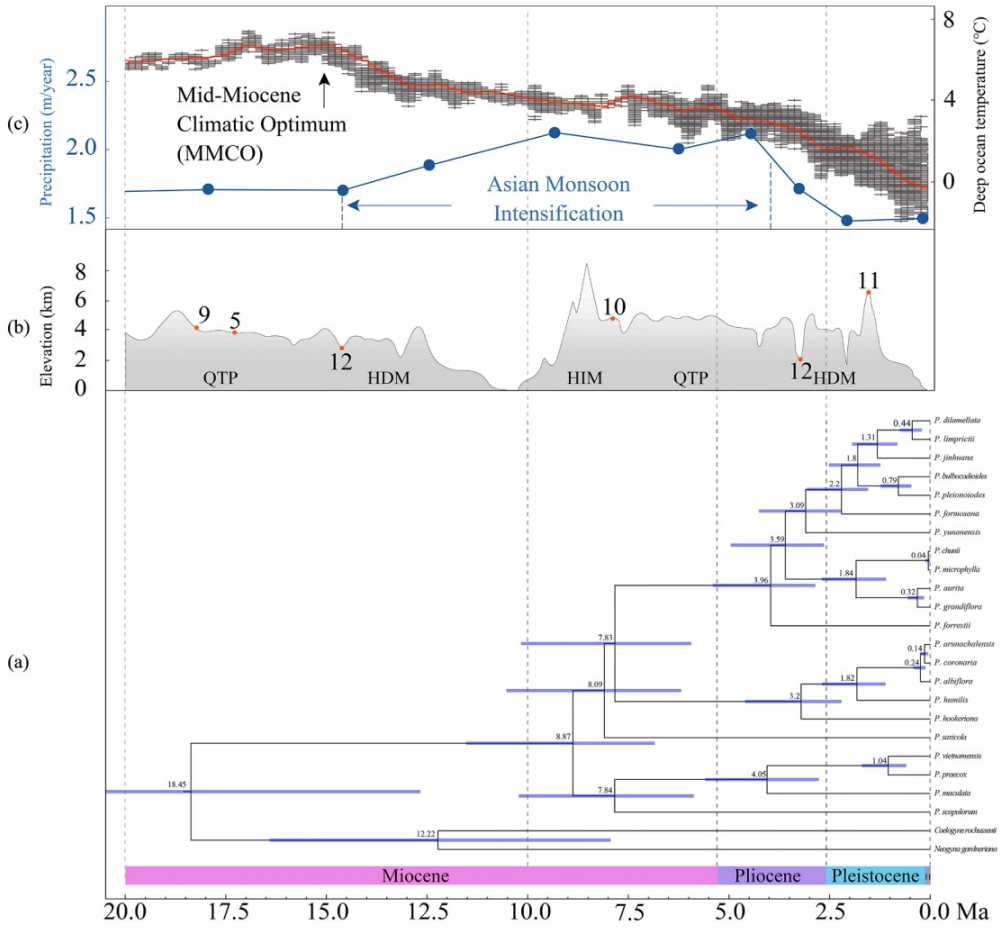
calibration | event

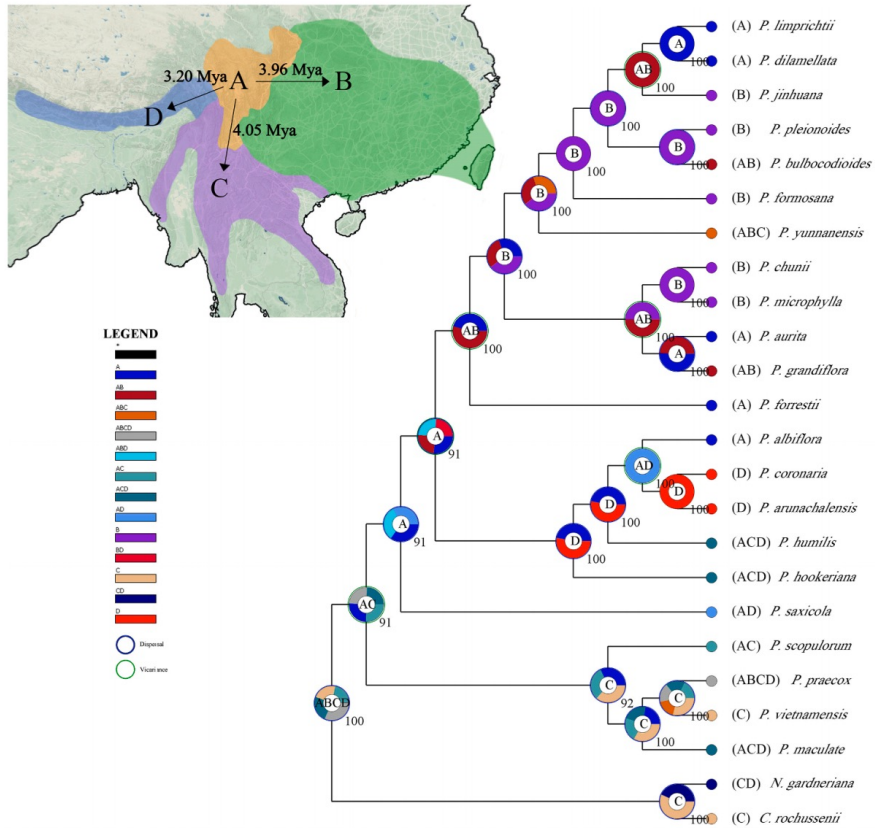
TIMETREE | Papers

BEAST | MCMCTREE

PSMC | BAMM

RSAP





Wu et al., 2023

what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

whole genome duplication

calibration

event

biogeography

gene family evolution

function evolution

adaptation

diversification

collinearity and syntenic analysis

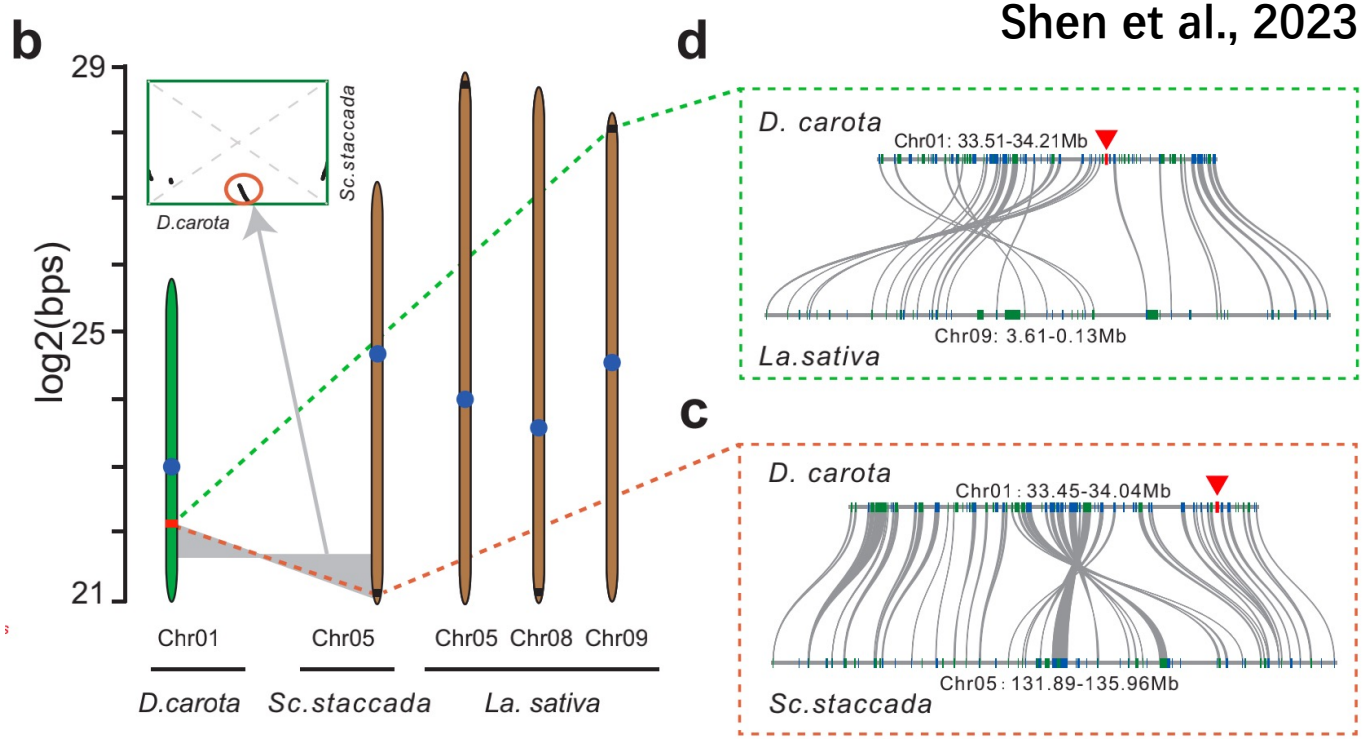
mcscanx | mummer | lastZ | blast | circos

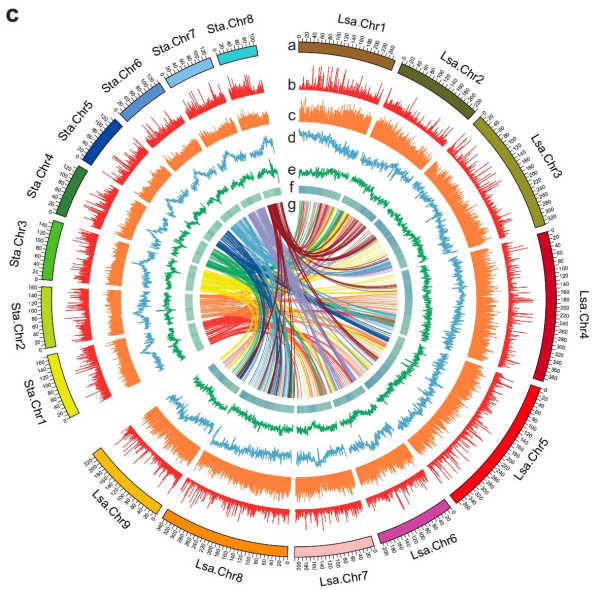
rearrangement

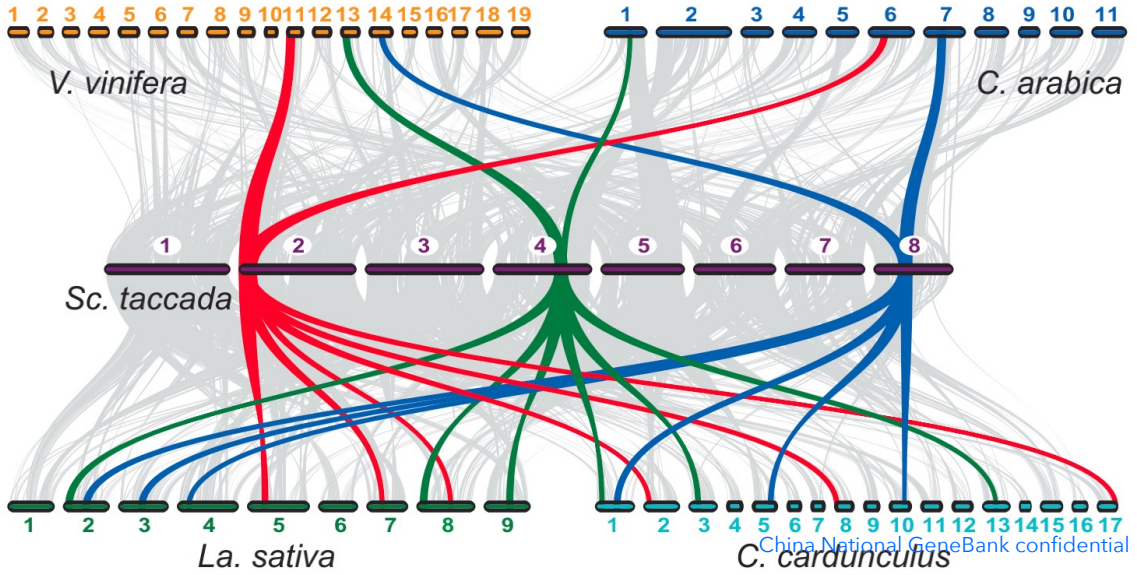
present or not

WGD

Shen et al., 2023







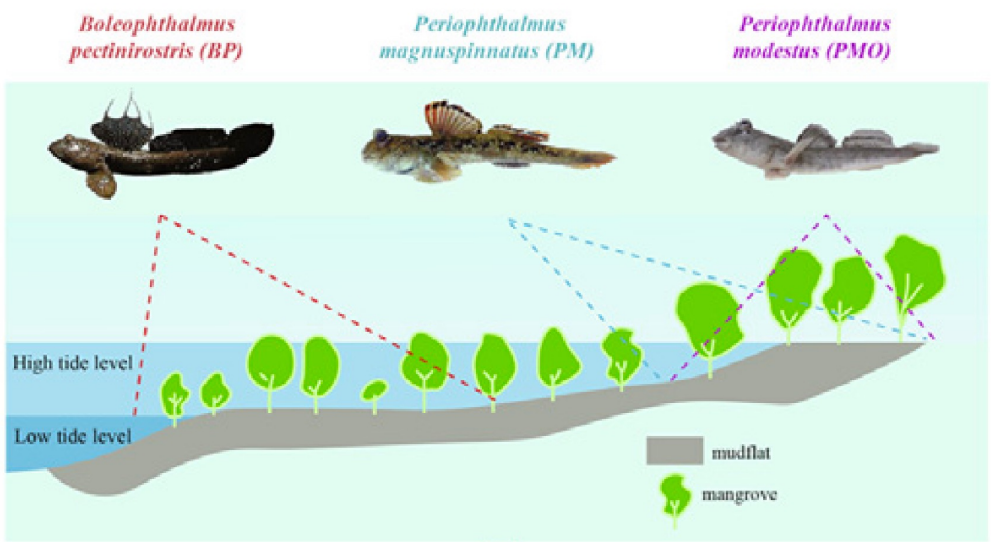
collinearity and syntenic analysis

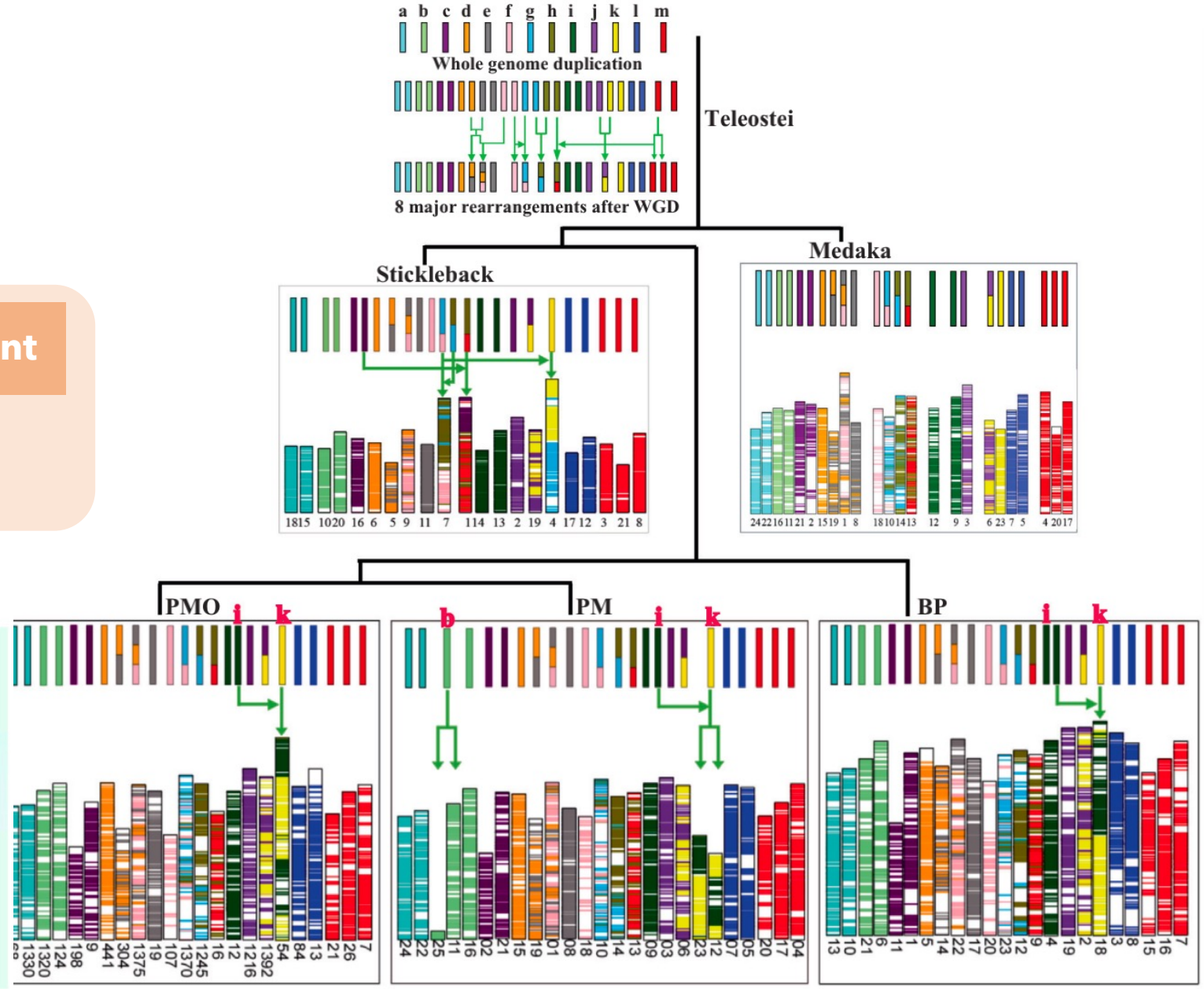
mcscanx | mummer | lastZ | blast | circos

chromosome evolution

rearrangement

WGD





Bian et al., 2023

what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

whole genome duplication

calibration

event

biogeography

gene family evolution

function evolution

adaptation

diversification

whole genome duplication

wgd | paml | i-ADHoRe

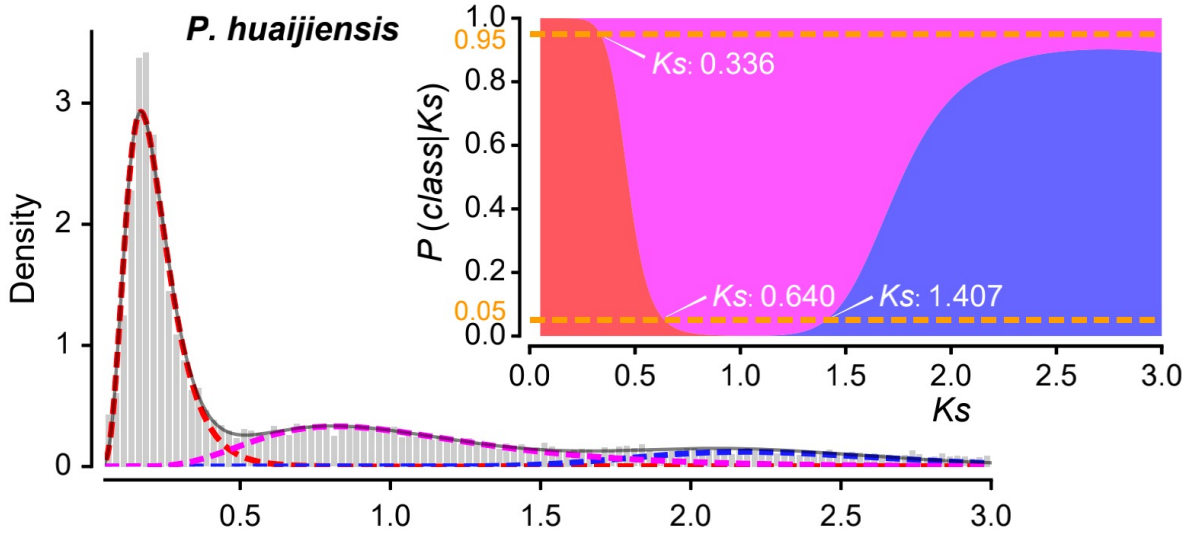
MAPS | DupGen_finder

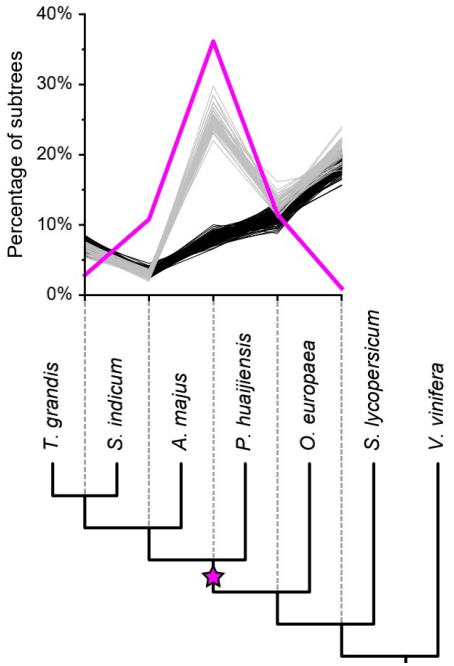
Ks distribution | phylogeny

Wgd or not

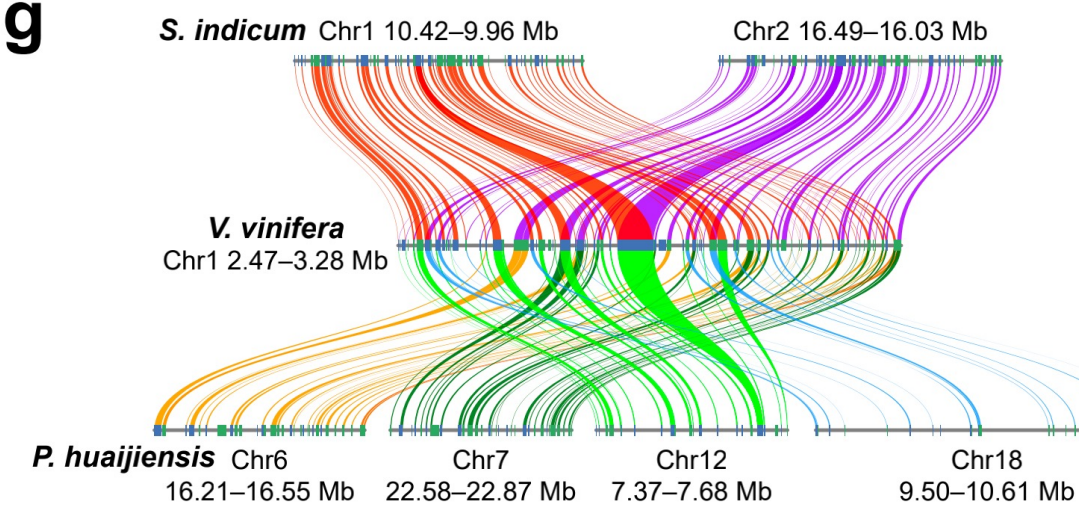
Share or not

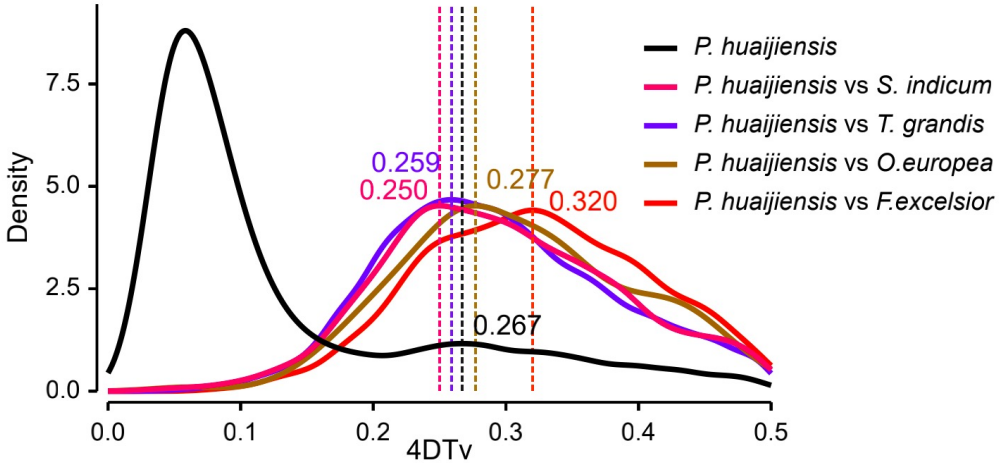
Gene retain





g





what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

calibration

event

biogeography

whole genome duplication

gene family evolution

function evolution

adaptation

diversification

gene family evolution

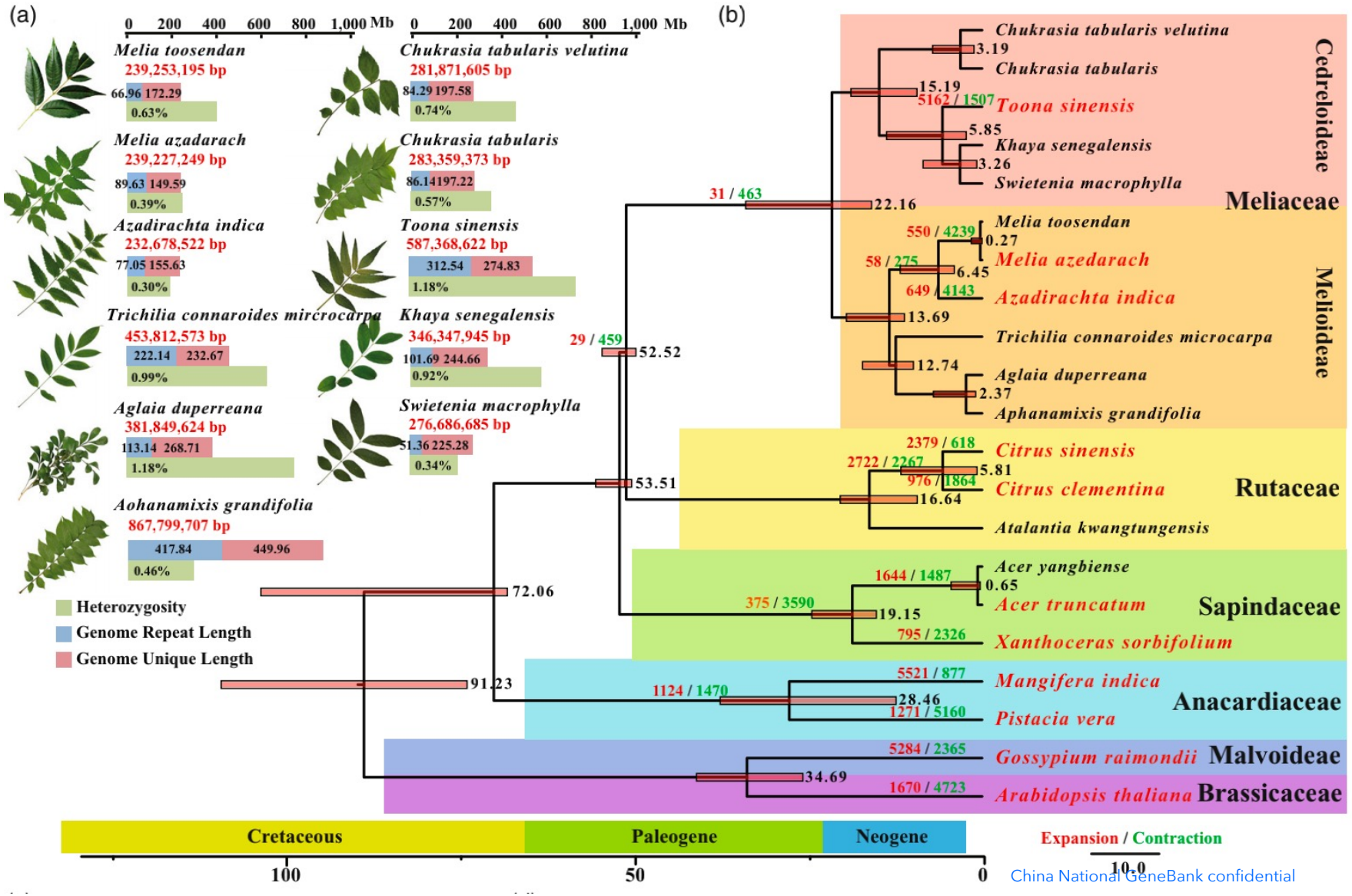
orthoFINDER

orthoMCL | cafe

expansion

contraction

specific gene



gene family evolution

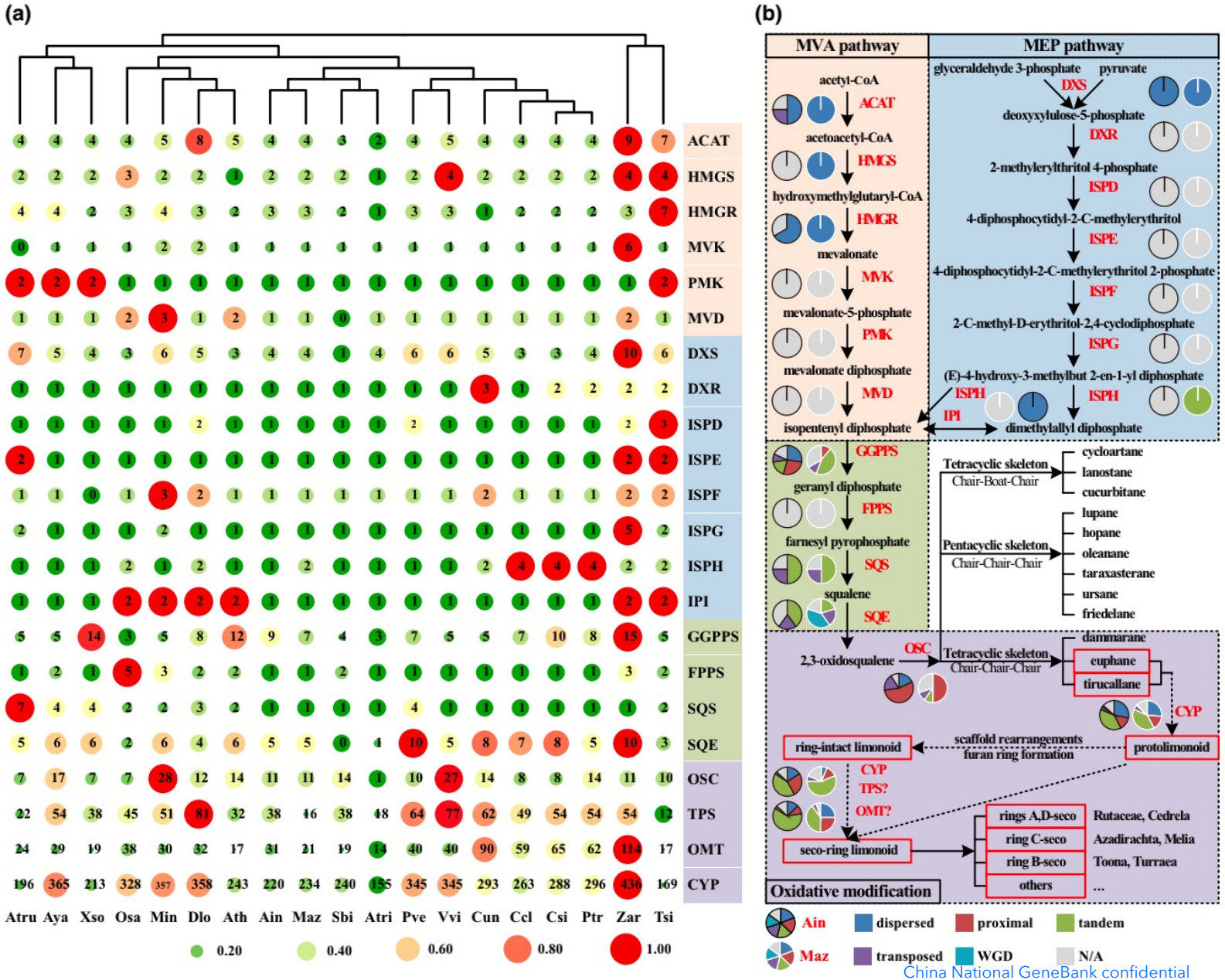
orthoFINDER

orthoMCL | cafe

expansion

contraction

specific gene



what we can do

collinearity and syntenic analysis

chromosome evolution

function analysis

phylogenetic reconstruction

whole genome duplication

calibration

event

biogeography

gene family evolution

function evolution

adaptation

diversification

downstream analysis

phylogenetic reconstruction

collinearity and syntenic analysis

biogeography

whole genome duplication

gene family evolution

function analysis

traits

Gene expression pattern

tissue

stage

period

function evolution

history

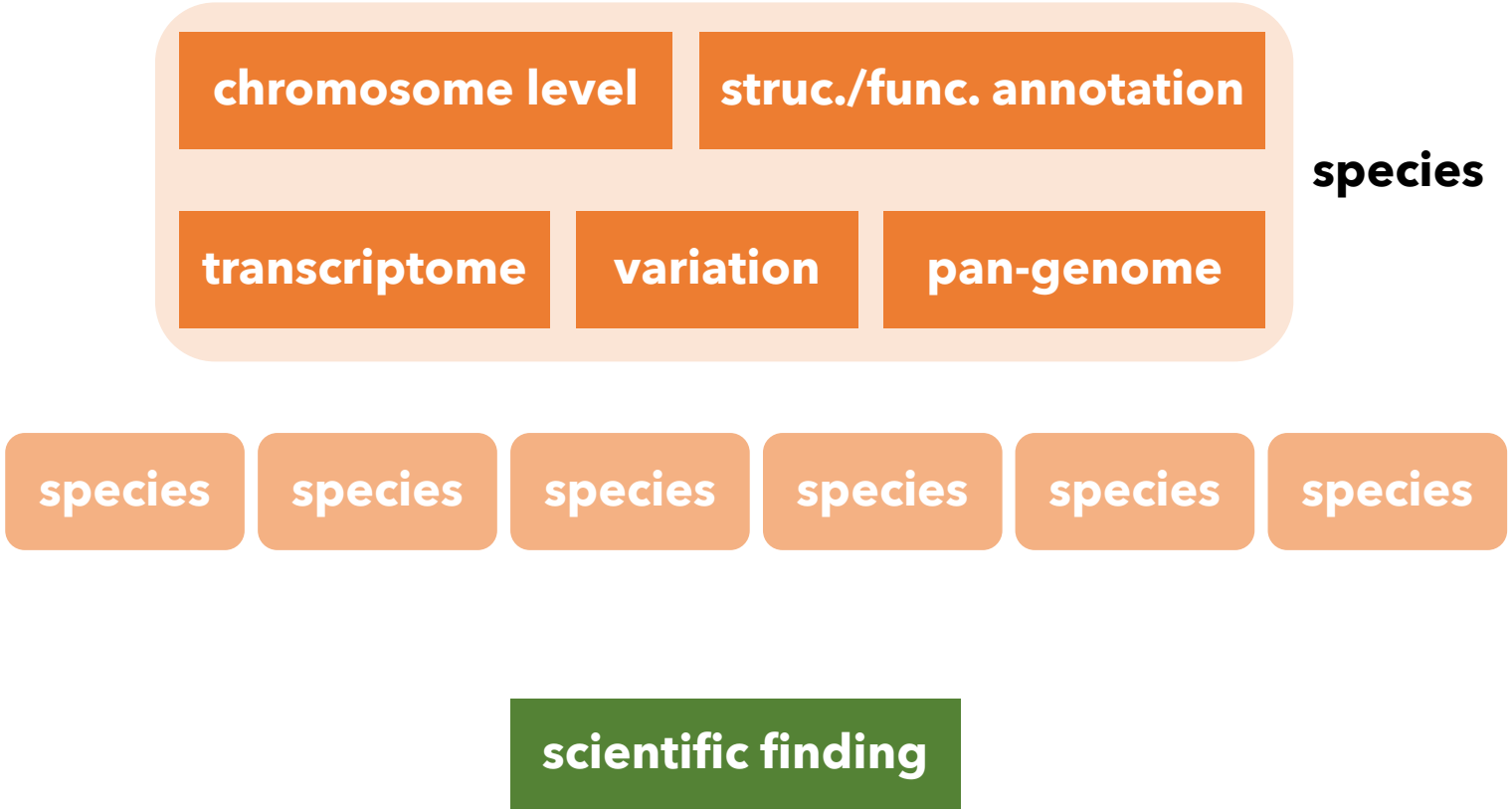
genes

SNPs

introgression

selection

what is CNGBdb doing now



what is CNGBdb doing now

Genome portal

species | species | species | species | species | species

scientific finding

what is CNGBdb doing now

Genome portal

species

Dataset

species

species

Dataset

species

species

Dataset

species

scientific finding

what is CNGBdb doing now

Genome portal

Dataset
species
species

species
Dataset
species

specific gene

single copy

scientific finding

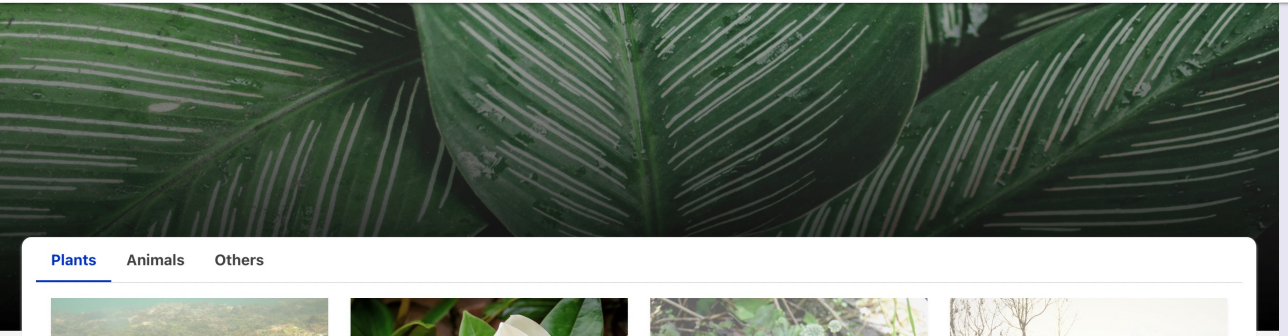
species
Dataset
species

Ka/Ks matrix

syntenic block

what is CNGBdb doing now

Home Search species Genome browser



Plants Animals Others

Algae

Angiosperm

Bryophyte

Fern and horsetail

Home Search species Genome browser

Angiosperm

Angiosperms, also called flowering plants, includes all forbs (flowering plants without a woody stem), grasses and grass-like plants, a vast majority of broad-leaved trees, shrubs & vines, and most aquatic plants. The term "angiosperm" is derived from the Greek words ἀγγεῖον /angeion ('container, vessel') and σπέρμα / sperma ('seed'), meaning that the seeds are enclosed within a fruit. They are by far the most diverse group of land plants with 64 orders, 416 families, approximately 13,000 known genera and 300,000 known species. Angiosperms were formerly called Magnoliophyta.

Source: Wikipedia

/ Search species / Angiosperm

Tree view

- Acorales (5)
- Alismatales (53)
- Apiales (71)
- Aquifoliales (24)

List

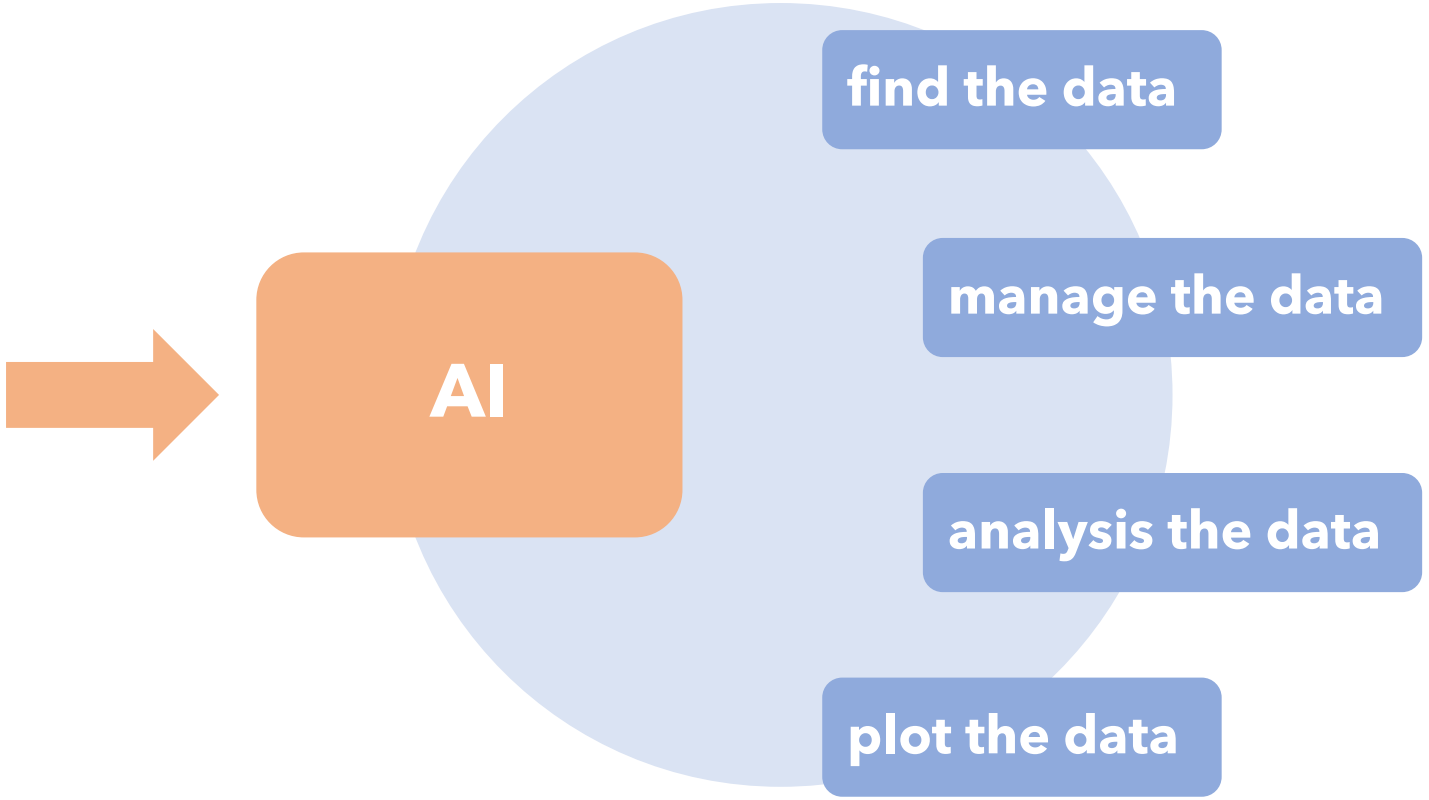
Show 10 results

View all selected information | Download selected information

Species	Taxon_id	Order	Family	Genus	Ploidy	Chromc
Abelia macrotera	1630337	Dipsacales	Caprifoliaceae	Abelia	--	--

what is CNGBdb doing now

Genome data framework



conclusion

- Comparing occurrs on stuffs with similarity but different in some respects.
- Comparative genome analysis connects the difference between molecular traits and morphological traits.
  - phylogenetic reconstruction
  - biogeography
  - collinearity and syntenic analysis
  - whole genome duplication
  - gene family evolution
  - function analysis and evolution
- CNGBdb are preparing the genome portal in the species diversification level.
- CNGBdb are preparing the AI4S (AI for science) tools to help manage and analysis the data.

China National GeneBank 国家基因库

华大生命科学研究院 BGI·Research

THANKS!

Jieyu (Jerry) WANG

wangjieyu@genomics.cn