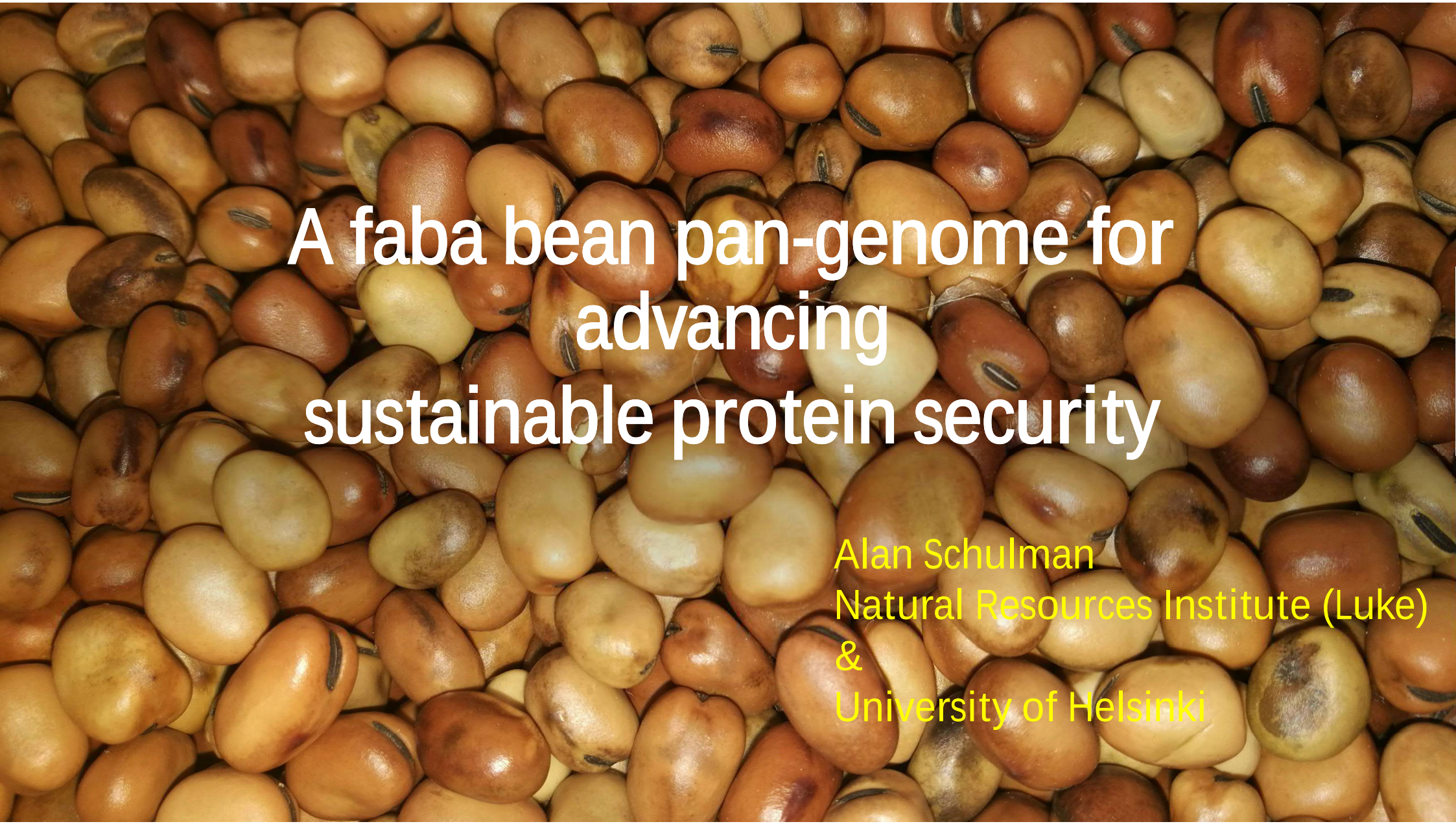
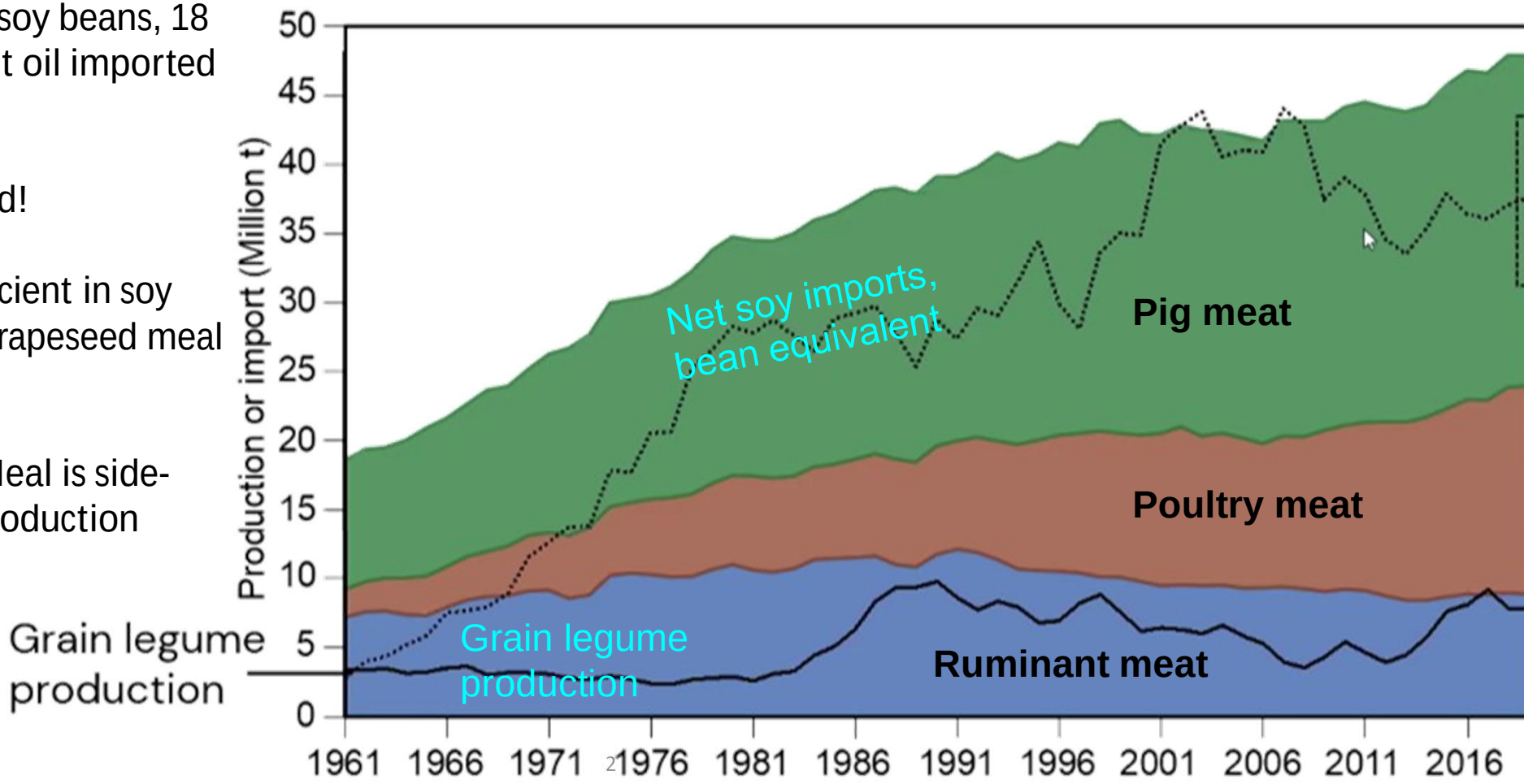


A faba bean pan-genome for advancing sustainable protein security

Alan Schulman
Natural Resources Institute (Luke)
&
University of Helsinki

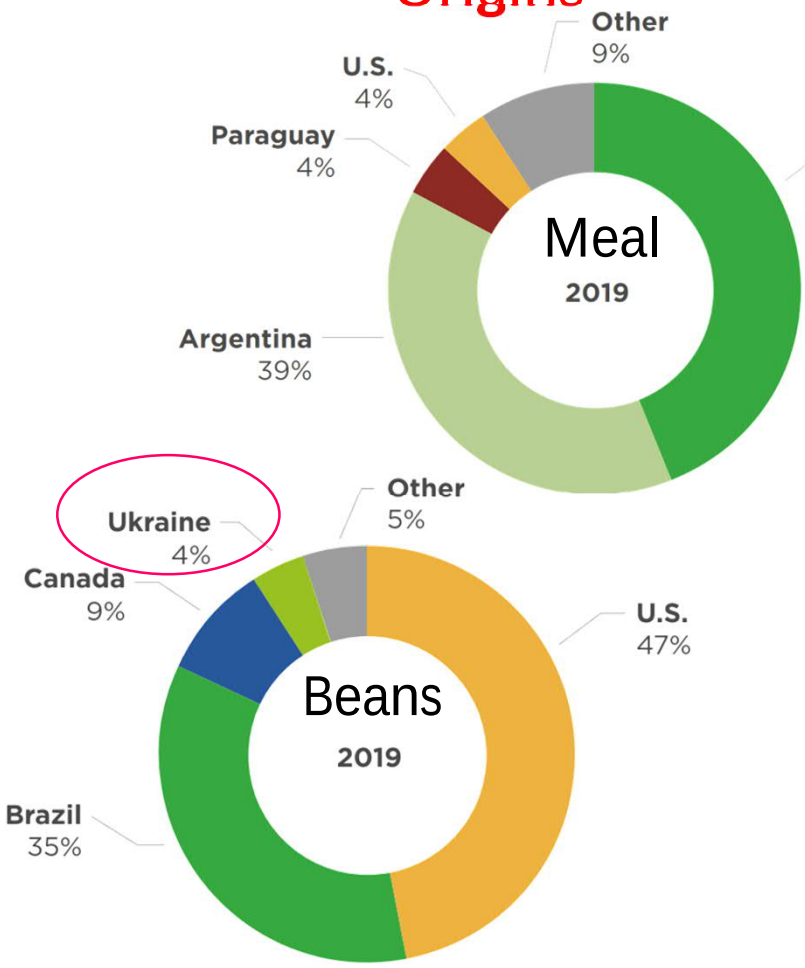
EU's plant protein imports driven by increased poultry and pig meat production

- Average 16 Mt soy beans, 18 Mt meal, 0.2 Mt oil imported annually
- = 13.5 M ha land!
- EU 5% self-sufficient in soy meal vs 79% in rapeseed meal (FEFAC, 2017)
- Soy is cheap! Meal is side-stream of oil production

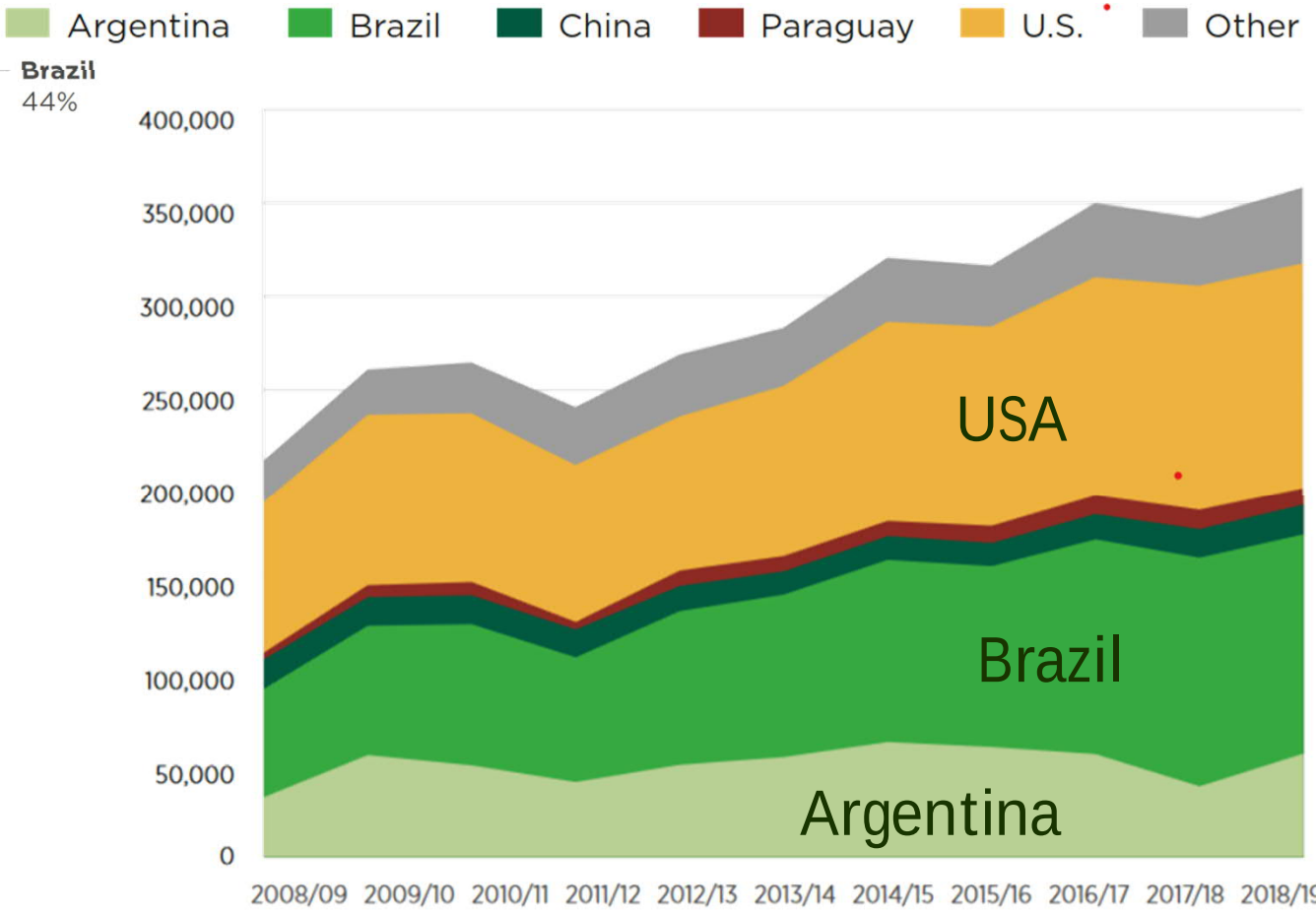


Where EU soy comes from

Origins

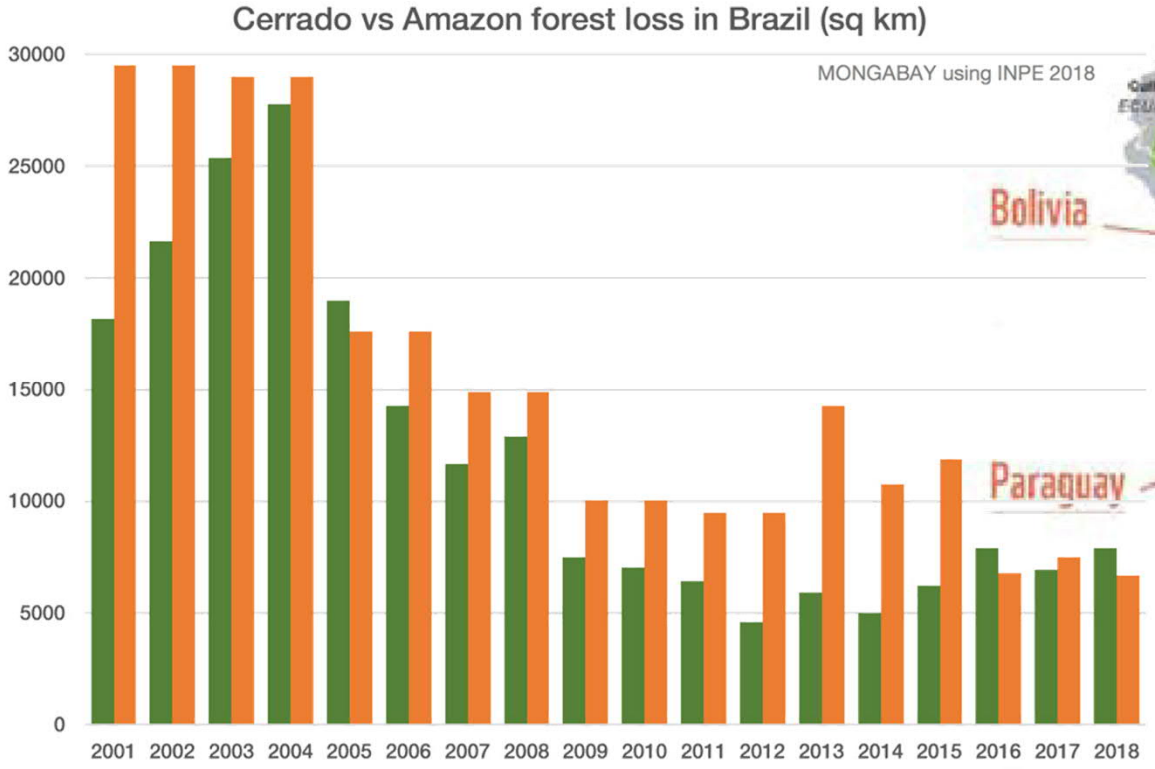


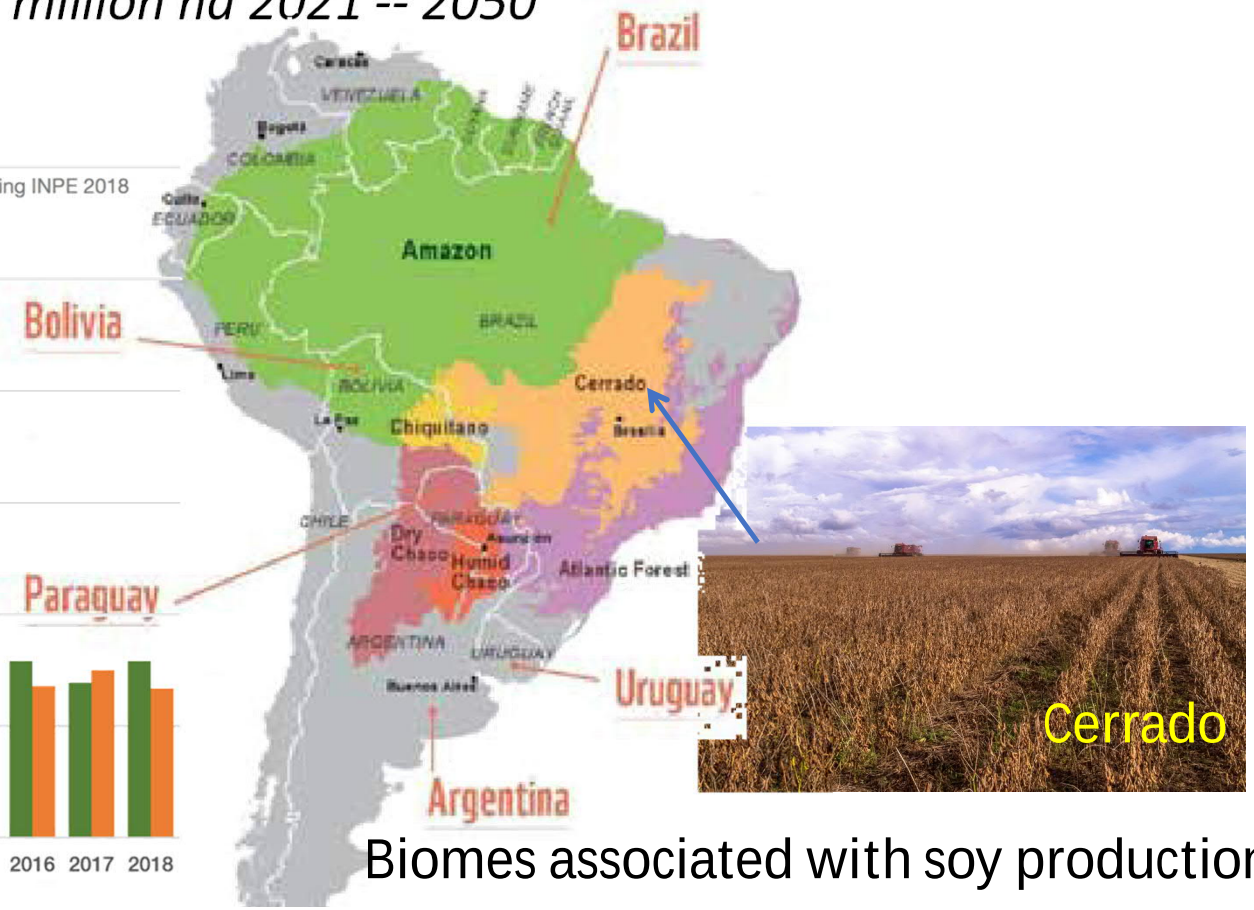
Soybean production, key countries 2008 – 2019 (in kt)



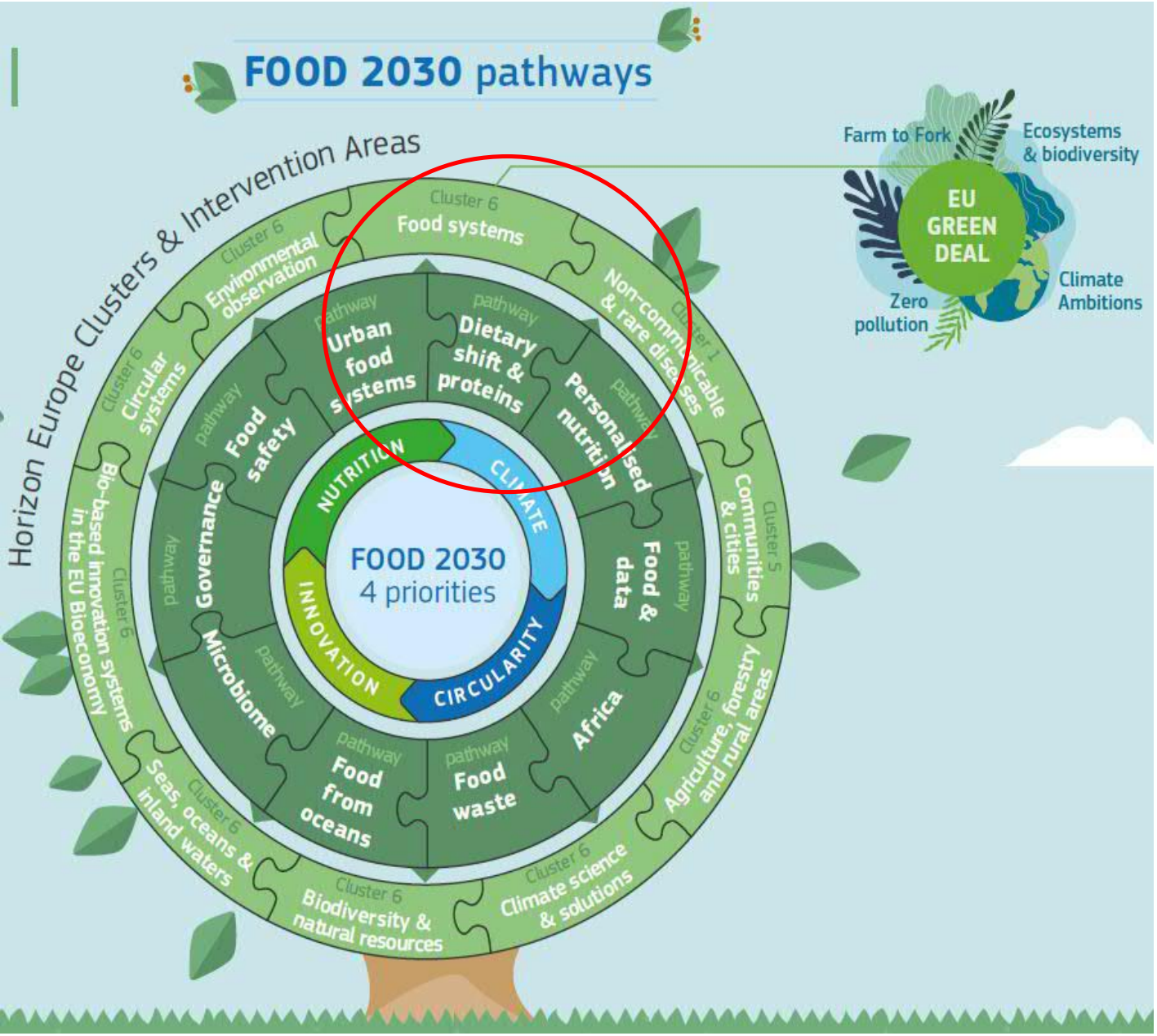
Bye, bye Brazil (rainforest & Cerrado)

- Brazil supplies 35% of world, 37% of EU, soy imports
- 2002 EU pigswill ban after 2001 F&M disease led to expansion into rainforest
- Recent expansion into Cerrado (savannah) may lead to its complete destruction:
- *Brazilian soy cultivation to expand by 12 million ha 2021 -- 2050*





Biomes associated with soy productior



5 RESEARCH AND INNOVATION NEEDS
1. ON THE IMPACT OF ALTERNATIVE PROTEINS AND DIETARY SHIFTS ON THE ENVIRONMENT AND HEALTH

•...The alternative proteins to be considered are (both old and new sources): plant-based proteins, microbial-based proteins, marine-based proteins, insect-based proteins, meat and fish meat alternatives (animal stem cells from living animals for cultured meat and fish meat), synthetic proteins from CO2 or other chemical sources.

European Commission - Press release

Commission acts for global food security and for supporting EU farmers and consumers

Brussels, 23 March 2022

Today, the European Commission has presented a range of **short-term and medium-term actions to enhance global food security** and to **support farmers and consumers in the EU** in light of rising food prices and input costs, such as energy and fertilisers. The surge in global commodity prices, further accelerated by Russia's invasion of Ukraine, highlights again the need for EU agriculture and food supply chains to become more resilient and sustainable, in line with the Farm to Fork strategy.

among the proposals presented by the Commission. The Commission calls on Member States to use all the available instruments in their **CAP strategic plans** for the period 2023-2027 in that regard. This concerns for example the use of risk management tools, the development of precision farming or coupled support to boost protein crops.

Food availability is currently not at stake in the EU, since the continent is largely self-sufficient for many agricultural products. However, our agricultural sector is a net importer of specific products, for example feed protein. This vulnerability, together with high input costs, such as fertilisers and fossil energy, is causing production challenges for farmers and risks driving up food prices.

The EU has a protein production deficit.

This is a food and nutritional security risk.

What to do?

Answer: Grow protein crops (i.e., legumes)

- EU defines “protein crops” as
  - pea (*Pisum sativum*)
  - faba bean (*Vicia faba*)
  - lupins (*Lupinus albus, L. angustifolius, L. luteus*)
  - dried alfalfa (*Medicago sativa*)
  - Excluding soybean: it is an oil crop!
- Imports of cheap soy meal injured European legume breeding
- Renewed interest over last 15 years:
  - food and feed security
  - sustainability
  - environmental impacts

Thanks to Fred Stoddard

Why grain legumes?

- Protein yield per hectare
- Europe imports 70% of its plant protein *supplement* requirements
  - = 16% of total protein feed requirements
- Legume protein achievable without N fertilizer inputs
- Break-crop effects > other broadleaved crops

- Problem: low prices for farmers in Europe
- Solution: develop more food uses

Why faba bean (*Vicia faba* L.)?

- Highest protein content (29%) and global average yield (1.7 t/ha) of starchy legumes
- Per ton of grain, fixes ~60 kg of N, takes up 20 kg from soil, leaves 40 behind, 40 goes into grain
- Important crop on all inhabited continents, 62°N to 45°S

BUT

- Mixed breeding system, partly reliant on bees
  - Difficulties for breeder but not farmer
- Does not like drought, heat, soil acidity
- Diseases, pests
- Yield instability
- Genetics, genomics, breeding has received less attention than those of soy or major cereals

Thanks to Fred Stoddard

Problems: Disease and pest resistance

- Chocolate spot (*Botrytis fabae*):
  - occasionally catastrophic (1 year in 10)
  - World-wide, most important faba disease
- 22°C, 95% RH, damp leaf surfaces, 2 days→ crop dead
- Sources of resistance available, selection difficult

Thanks to Fred Stoddard

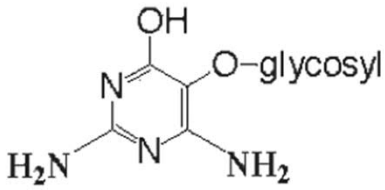
Other diseases:

- Ascochyta blight, rust, downy mildew
- Aphids
- *Sitona* leaf weevils
  - adults eat leaves, larvae eat root nodules
- Bruchid seed weevils

Problems: Antinutrients

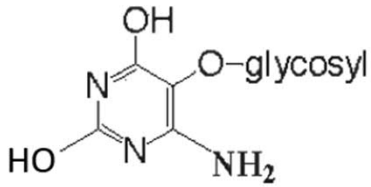
- Vicine + convicine
- Tannins in seed coats
- Protease inhibitors
- Oligosaccharides (→ flatulence, in pigs as well as people)
  - alpha-galactosides of sucrose: raffinose, stachyose and verbascose

Vicine



(aglycone: divicine)

Convicine



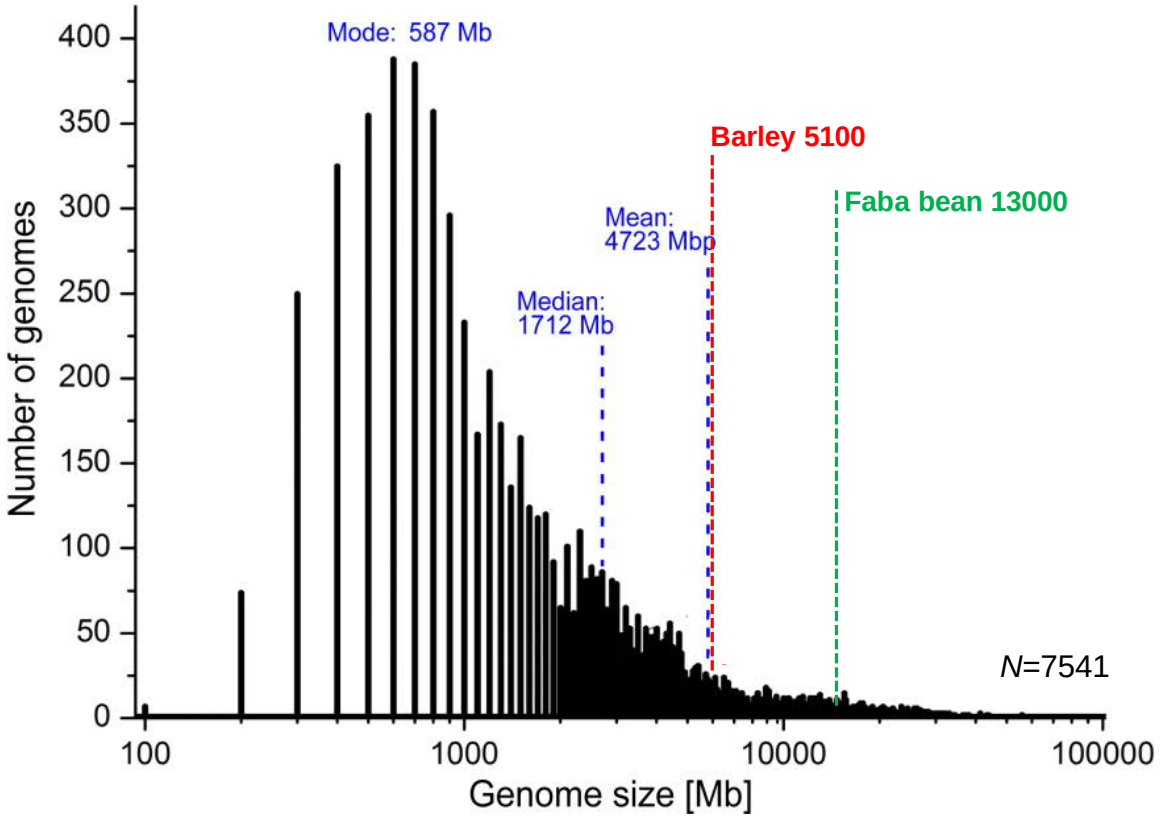
(aglycone: isouramil)

Problems: Quality

- Seed lipases→ “beany” flavor in processing
- Protein quality: legumin vs vicilin content

Problems: Huge genome (13 Gbp, diploid), small investment

Angiosperm genome sizes

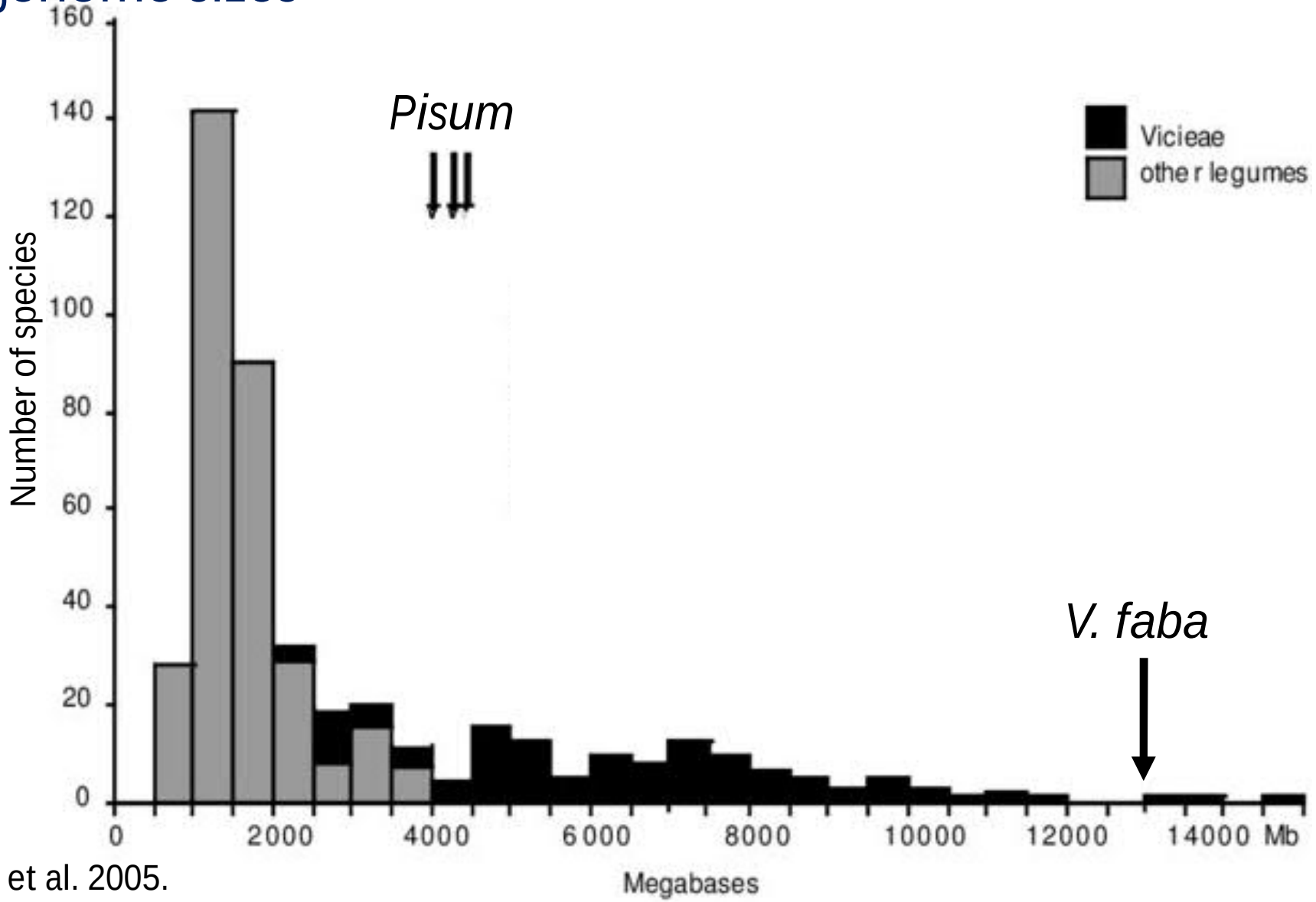


Mean:	4723
Median:	1712
Mode:	587
Max:	101370
Minimum:	63

Distribution of determined angiosperm monoploid genome sizes. Bins of 100 Mb vs number of genomes in each bin (total 7541).

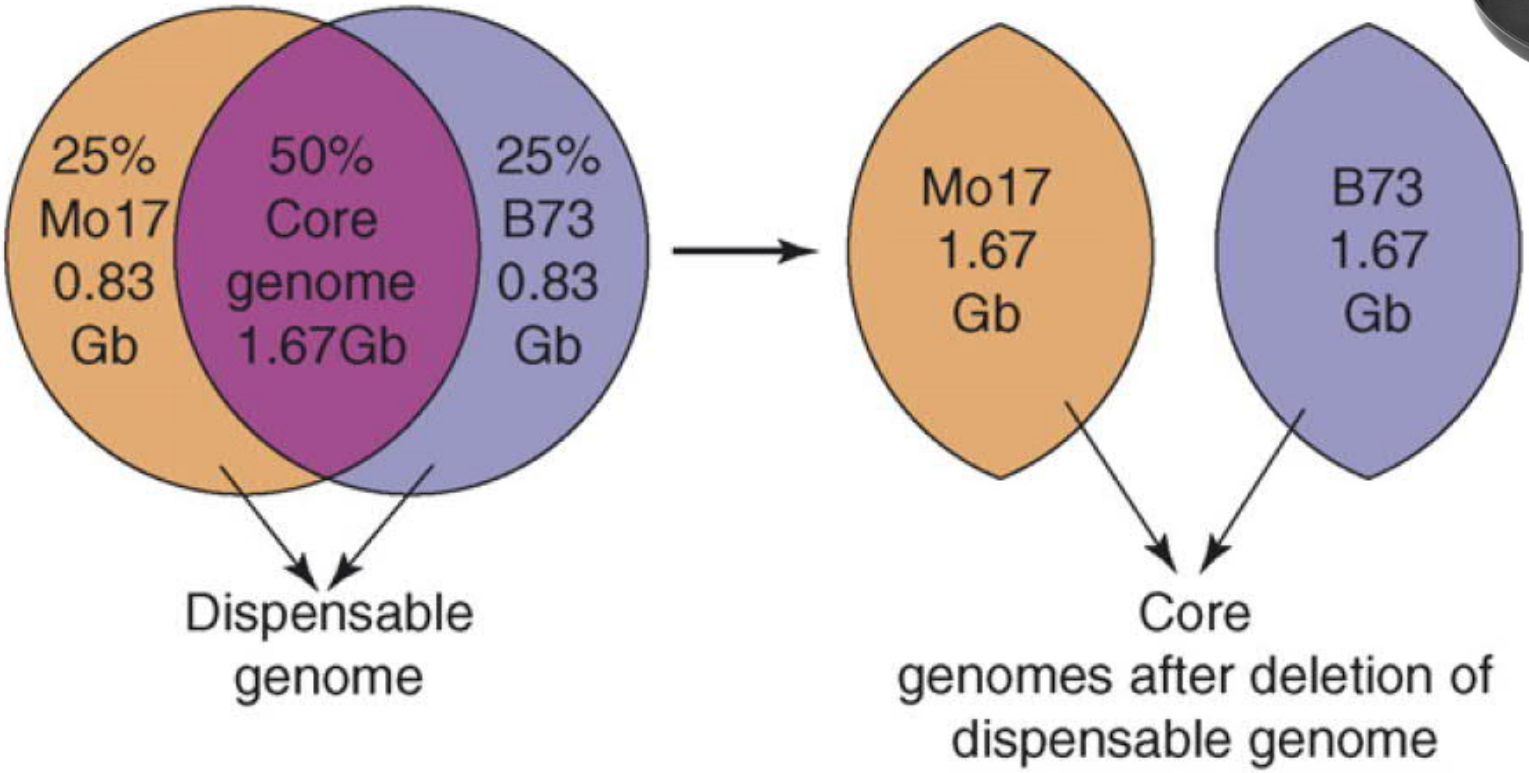
Data source: http://data.kew.org/cvalues/ , accessed 07.02.2017

Legume genome sizes



Modified from Ellis et al. 2005.

The Core and Pan-Genome Concept



Pan genome: full set of genes in a species
Core genome: genes shared by all individuals in a species

Core- Pan = “Dispensible Genome”

Morgante 2007

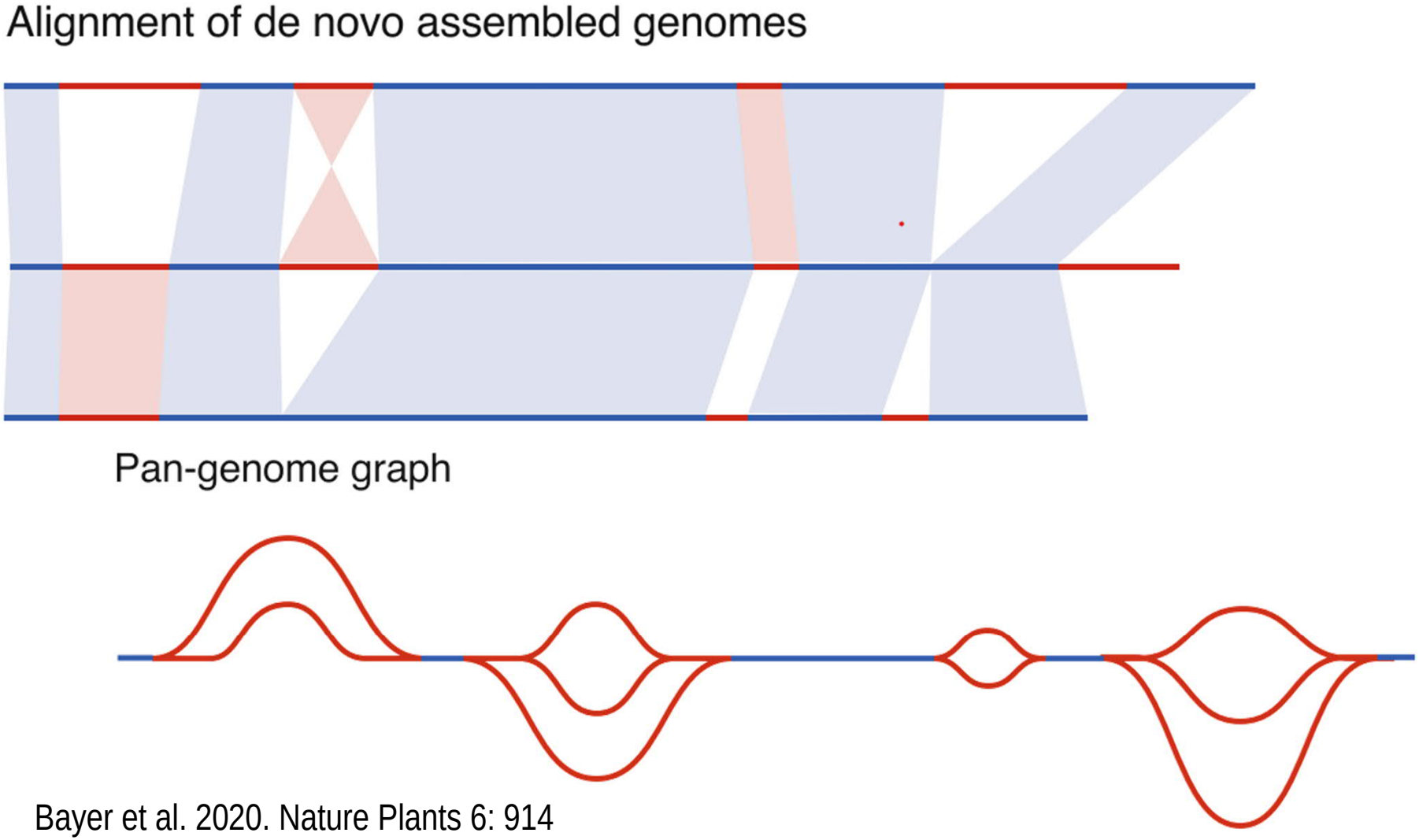
Rice Core and Pan genome

3010 rice genomes sequenced to 14.3× depth. >12 000 novel genes absent in the reference genome were found.

Gene category	Count
Total genes	50 995
Core genes	23 914
Candidate core genes	4986
Distributed genes	22 095
Subspecies-unbalanced genes	13 617
Indica-dominant genes	5579
Japonica-dominant genes	6038
Subspecies-specific genes	853
Indica-specific genes	587
Japonica-specific genes	147
AUS-specific genes	67
ARO-specific genes	52
Subgroup-unbalanced genes	11 581
Indica-subgroup-unbalanced genes	9816
Japonica-subgroup-unbalanced genes	3418
Random genes	5316

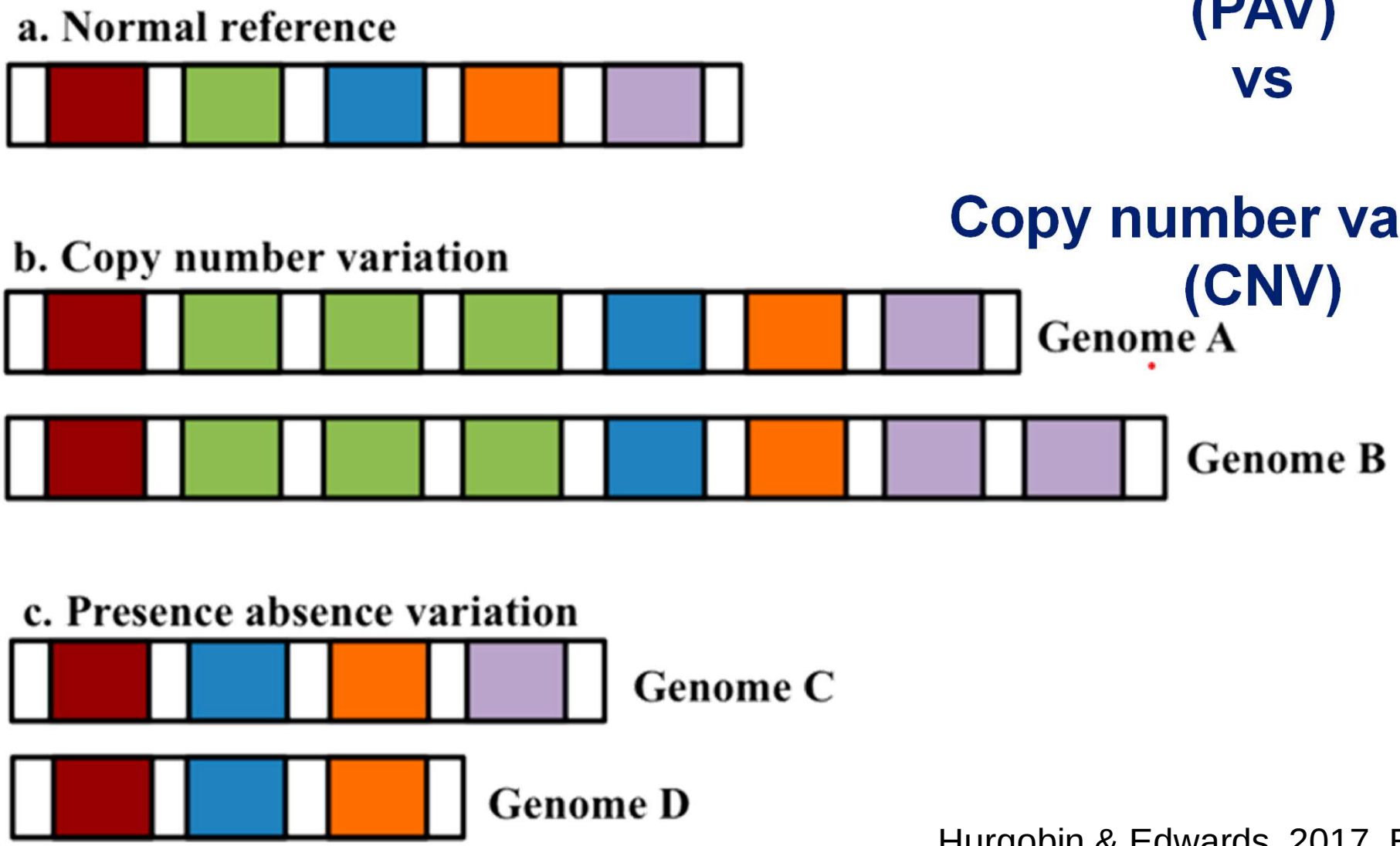
Sun C et al. 2017. Nuc Acids Res. 45: 597-605.

The Core and Pan-Genome Concept



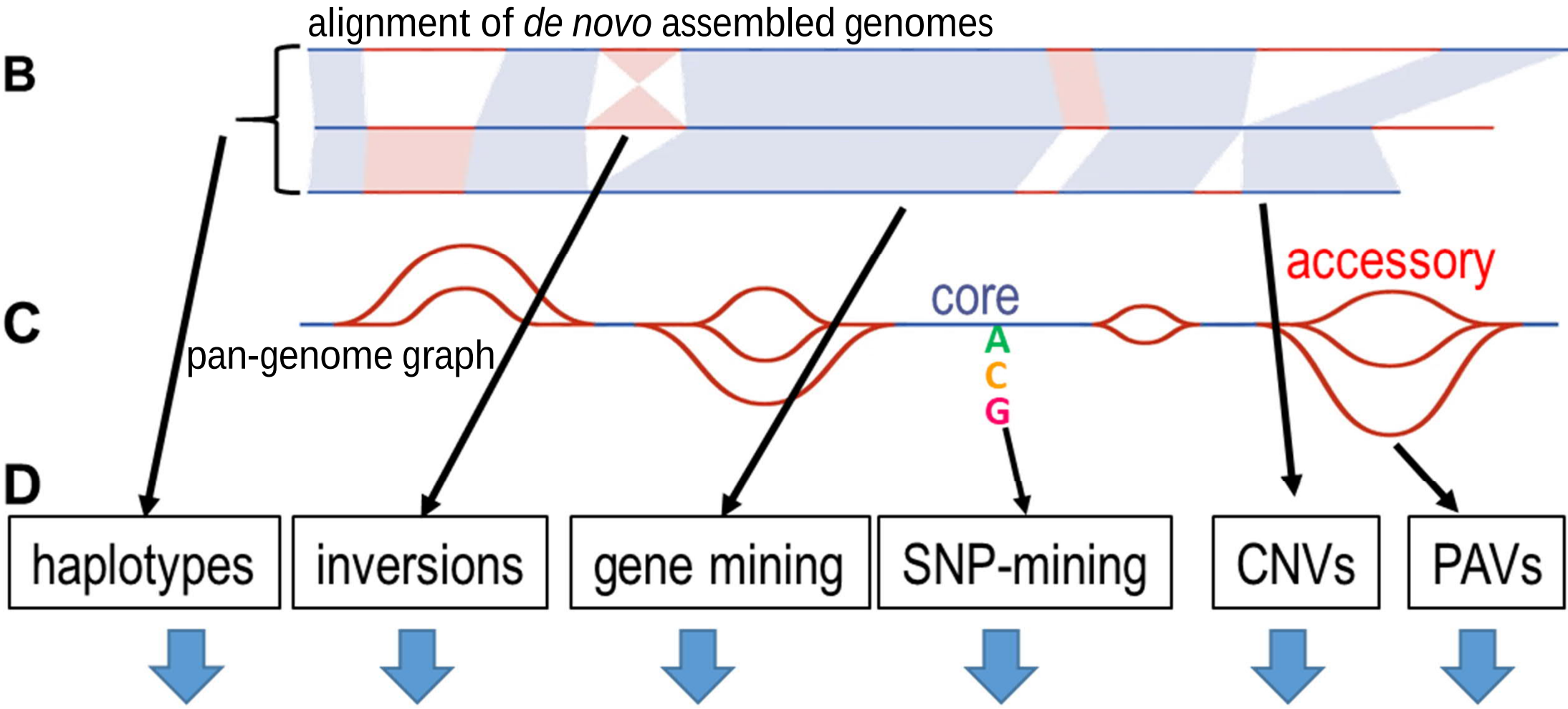
Bayer et al. 2020. Nature Plants 6: 914

Presence-Absence Variation (PAV) vs Copy number variation (CNV)

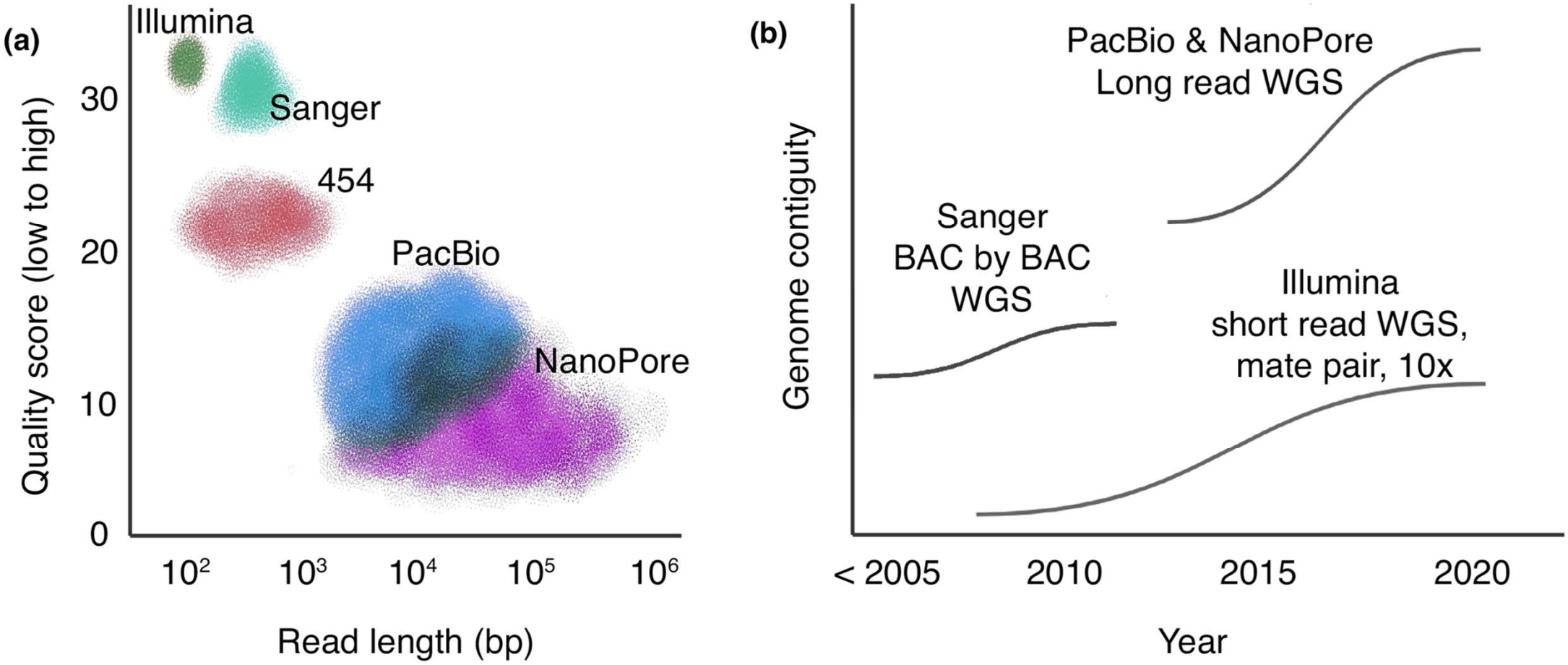


Hurgobin & Edwards. 2017. Biology 6: 21

Pan-genome analysis



Advances in sequencing technology drive increasing contiguity



Michael & VanBuren. 2020. Curr Opin Plant Biol 54:26

Faba sequencing strategy:

PacBio Sequel II CCS HiFi reads

= circular consensus sequencing

Faba reference genome consortium

Nadim Tayeh, Gregoire Aubert, Jonathan Kreplak
PacBio sequencing, data analysis

Agnieszka Golicz, Rod Snowdon
comparative genomics

Francis Ogbonnaya
PacBio sequencing

Institute of Experimental Botany of the Czech Academy of Sciences
Dolezel Jaroslav
Genome size

Olaf Sass, Felix Dreyer
PacBio sequencing

Coordinator: Stig U. Andersen
PacBio sequencing
Hi-C scaffolding
Assembly

Murukarthick Jayakodi
Assembly, methylome

Kirstin Bett
Assembly

Alan Schulman
Repeat analysis

Ana M. Torres
Genetic maps

Donal O'Sullivan
Genetic maps

Jiří Macas
Repeat analysis

Shusei Sato
Assembly

https://fabagenome.dk

PanFaba Goal

Build a *de novo* pan-genome from five accessions that sample the genetic diversity for *V. faba*

Criteria for plant lines:

1) sufficient inbreeding to confer a high level of homozygosity (>95.5%) for sequencing

2) sampling of an important diversity pool, with useful traits and alleles, for later functional analyses;

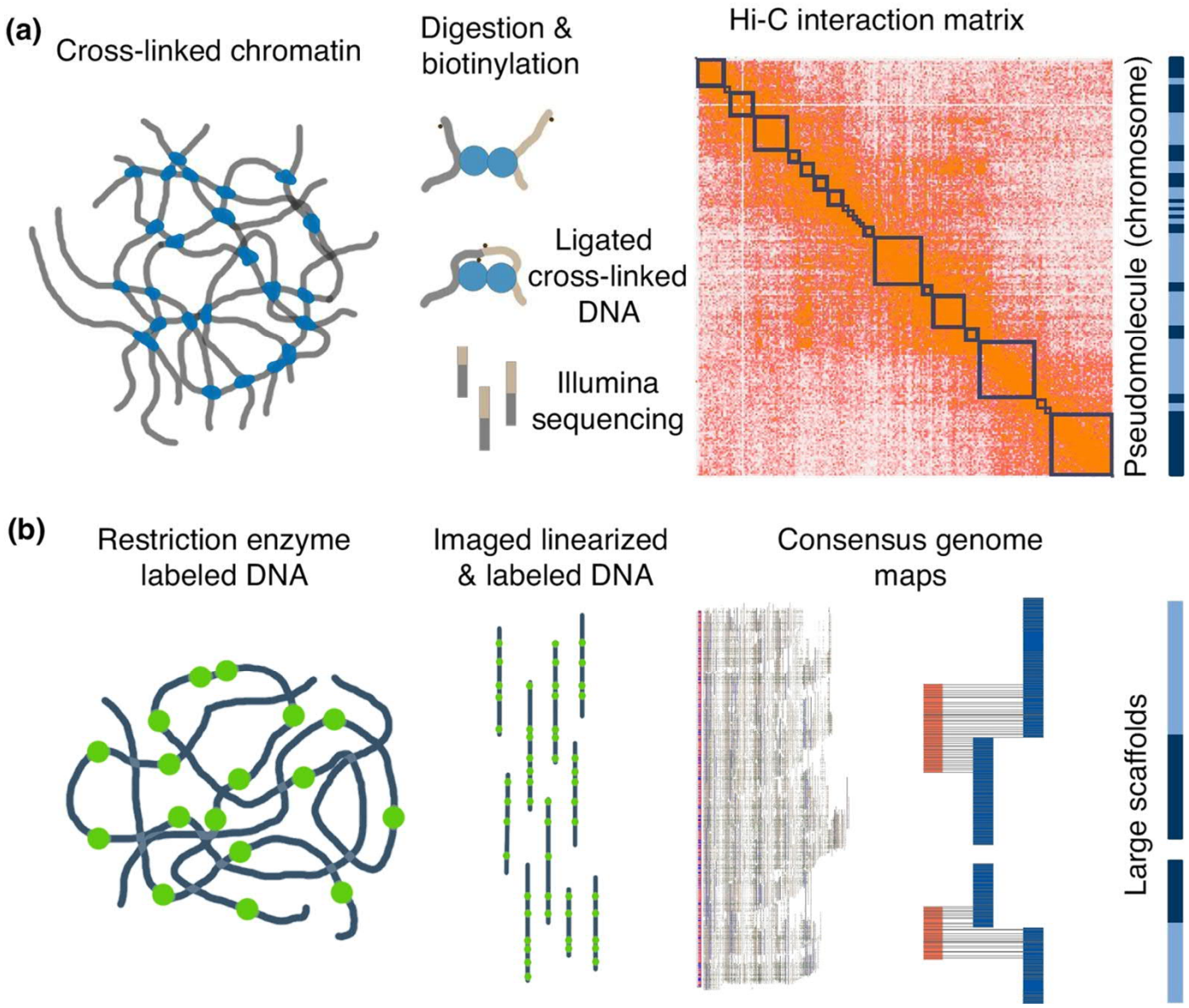
3) being a parent of a mapping population to aid assembly and orientation of scaffolds as well as for genotype—trait association;

4) relevance to collaborative efforts for faba bean improvement

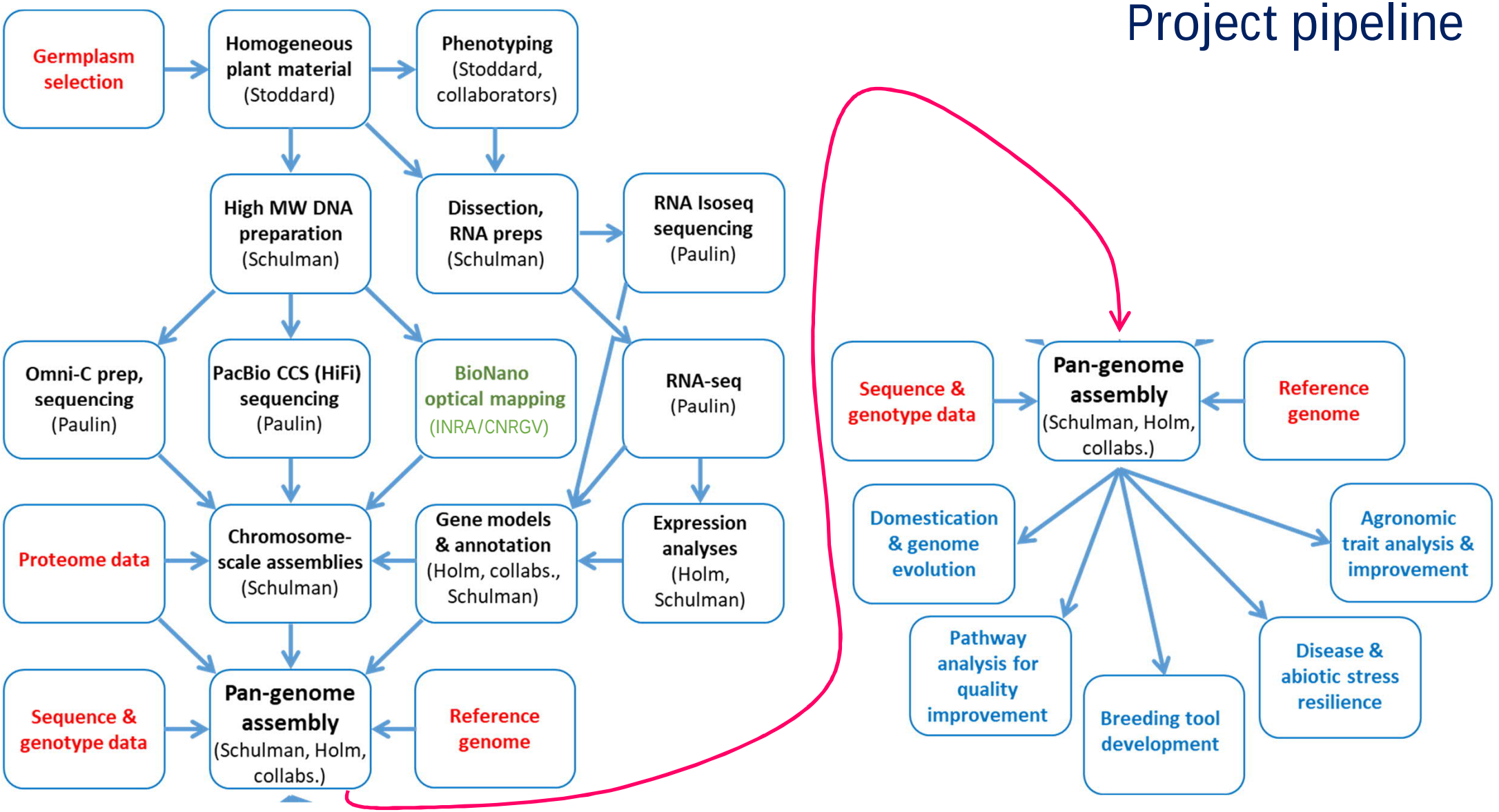
PanFaba: technical

- In-house PacBio Sequel II;
  - budgeted for 26X coverage, 25 Gbp HiFi / cell, 400 Gbp / line.
- In-house IsoSeq for gene models, 5 libraries for each line
  - 1) developing seeds, three stages; 2) etiolated seedling shoots; 3) sterile-grown seedling roots; 4) pre-fertilization flowers, 5) developing pods, expanding and filling stages.
- RNA-seq data for genotype/phenotype association: three tissues, four lines
  - developing seeds for quality; roots and leaves for drought, aluminum, acid response
- BioNano Saphyr II by INRAE/CNRGV Toulouse, France
- Hi-C as Omni-C in-house
- Computer-infra from UH, CSC, LUKE ; assembly Hifiasm
  - CSC ePouta: 80-core/1.4 TB RAM

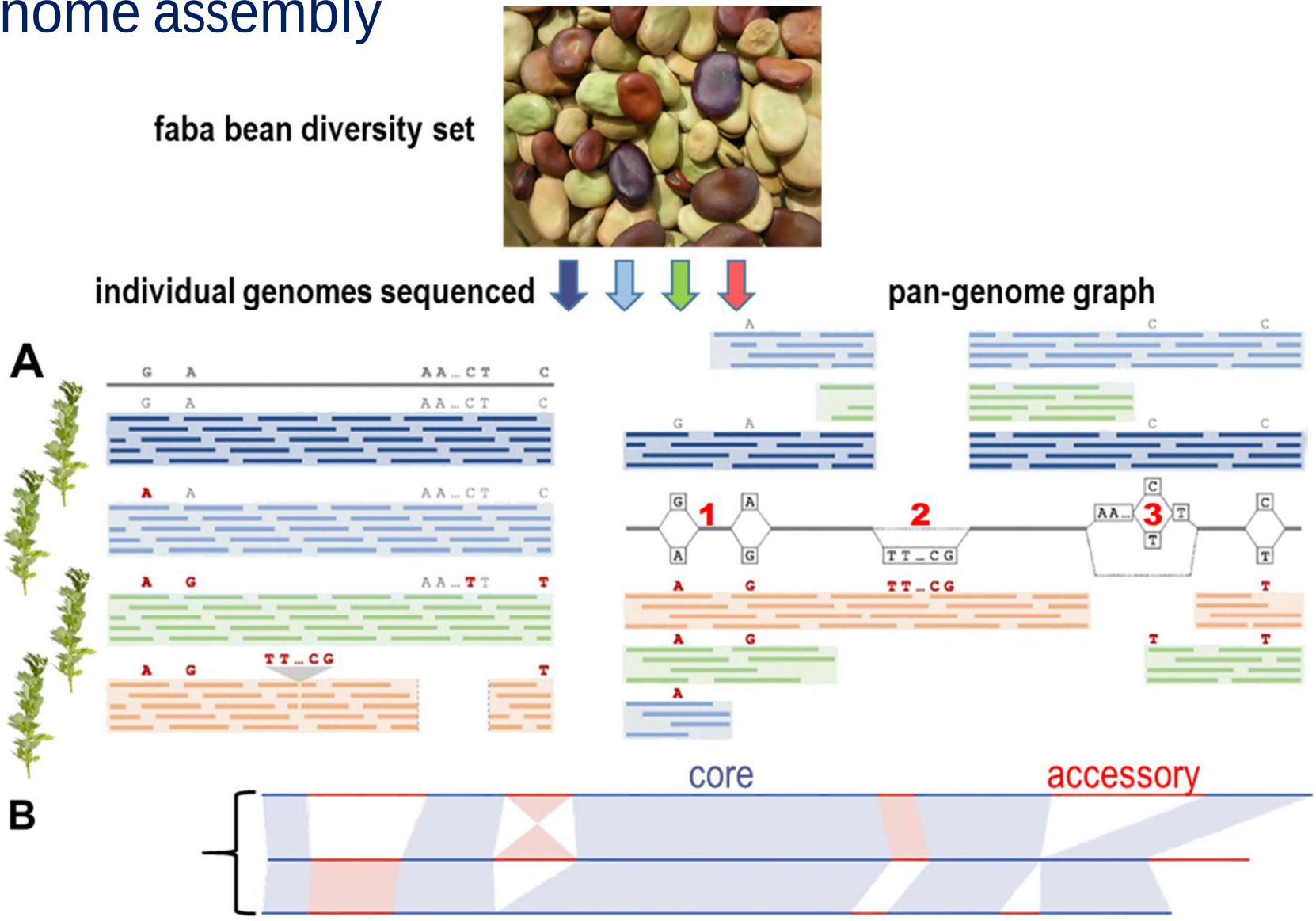
Scaffolding strategies



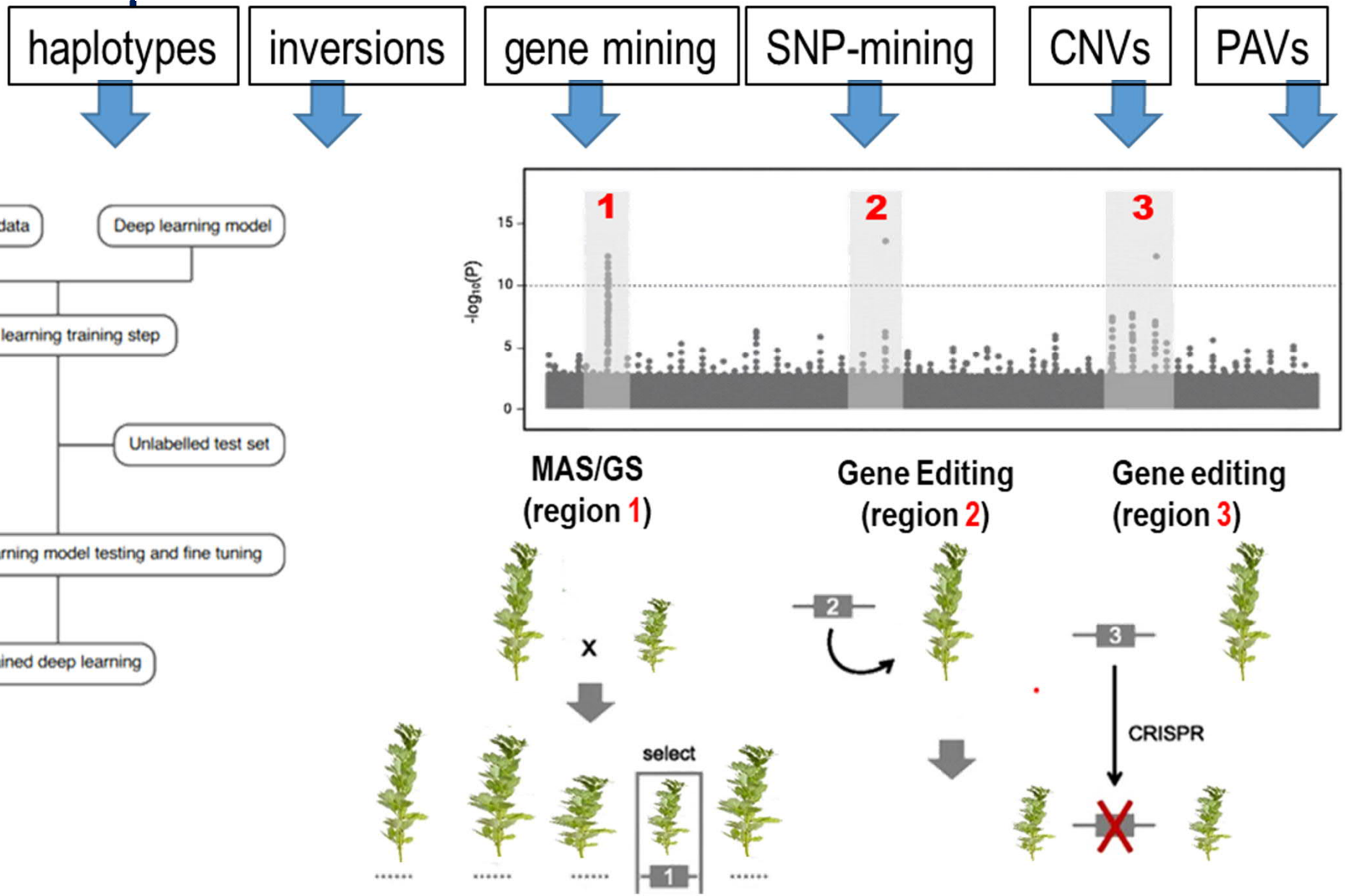
Project pipeline



Pan-genome assembly



Pan-genome exploitation



The Pan-genome expands!

- Rod Snowdon, Agata Dasvkowska , Björn Usadel (Univ. Hohenheim, HHU Düsseldorf, DE)
  - 6 *de novo* genomes + Tiffany (already complete)

- Murukarthik Jayakodi (IPK, Gatersleben, DE)
  - 5 *de novo* genomes

Total: →17 *de novo* á 13 Gbp!

Multiple other tools in hand, or being developed

- 60 k chip (Donal O'Sullivan, Univ Reading)
- SPET genotyping, 90k probes, 18 Mb gene space (Stig U. Andersen, Univ Aarhus)
- 100s of genotyped inbred lines (NORFAB, ProFaba projects)

- Gene editing under development … (Schulman, other groups).

PanFaba thanks!

PI partners:

- Alan Schulman (Coordinator)
- Fred Stoddard (germplasm, plant material, phenotyping)
- Lars Paulin + Petri Auvinen (libraries, sequencing, pre-assembly)
- Liisa Holm + Petri Törönen (annotation)

Do-ers:

- Wei Chang (RNA wet-lab)
- Pia K Laine (PacBio HiFi data)
- Anne- Mari Narvanto et al. (technical)
- Marco Salgado (gene models, annotation)
- Jaakko Tanskanen (assembly, bioinformatics)
- Petri Törönen (gene family bioinformatics)

Collaboration

- Reference genome consortium ☺
- Nathalie Rodde et al. (INRAE; BioNano)