UNIVERSITY OF
HOHENHEIM

Genetic diversity

3502-470 Plant Genetic Resources

Prof. Karl Schmid
SS 2025

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

Plant genetic resources

The paradox of genetic diversity

Loss of genetic variation by breeding

Quantification of genetic variation

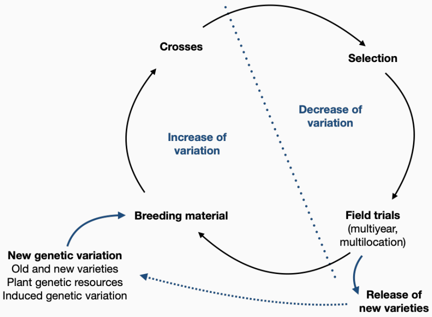
Plant genetic resources

The paradox of genetic diversity

Loss of genetic variation by breeding

Quantification of genetic variation

Innovation cycle in plant breeding



Crosses

Selection

Decrease of variation

Increase of variation

Breeding material

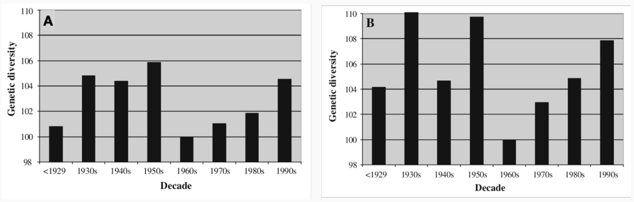
Field trials (multiyear, multilocation)

New genetic variation
Old and new varieties
Plant genetic resources
Induced genetic variation

Release of new varieties

After Jorasch (2019)

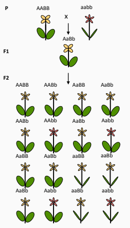
Genetic diversity of varieties in the 20th century



Wan de Wouw TAG (2010)

The paradox of genetic diversity

Different alleles at 2 loci



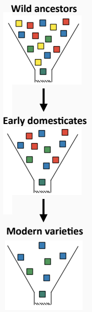
$3^2 = 9$ genotypes

Different Alleles at 21 loci

P `AABBCCDDEEFFGGHHIIJJKK`
`LLMMNNOOPPQQRRSSTTUU`
x
`aabbccddeeffgghhiijjkk`
`llmmnnooppqqrrssttuu`
$\Downarrow$
F1 `AaBbCcDdEeFfGgHhIiJjKk`
`LlMmNnOoPpQqRrSsTtUu`
$\Downarrow$
F2 $3^{21} = 10,460,353,203$ genotypes

This is more variation than a plant breeder can handle!

Loss of genetic variation in crop history



Modified after McCouch and Tanskley, Science (1997)

Reduction of genetic variation by

- Bottlenecks: Only a small proportion of the total genetic variation is included in breeding population
- Genetic drift: Random fixation of alleles
- Selection: Fixation of advantagous alleles / Loss of disadvantagoues alleles

---

Types of genetic variation

Useful / Beneficial / Advantageous:

- yield increase
- better adaptation
- higher resistance
- etc.

Neutral:

Genetic variation without phenotypic effects

Deleterious / Disadvantageous:

Any genetic variation that reduces the yield and quality of the crop or that simply reduces the fitness

---

Genetic vulnerability due to a narrow genetic basis

Famine memorial in Dublin

Wikipedia

- **Potato leaf blight in Ireland 1845-1849**, which was caused by the pathogen *Phytophtora infestans*. As a result one million people starved to death and one million people emigrated to the US.
- **Southern leaf blight** in 1970, where corn with the so-called texas-T cytoplasm was wiped out by Southern leaf blight.

Genetic vulnerability and genetic erosion

- **Genetic vulnerability:** Condition that results when a widely planted crop is uniformly susceptible to a pest, pathogen or environmental hazard as a result of its genetic constitution, thereby creating a potential for widespread crop losses
- **Genetic erosion:** The loss of individual genes and the loss of particular combinations of genes (i.e. of gene complexes) such as those maintained in locally adapted landraces.
- Alternative uses of the term 'genetic erosion':
  - Loss of genes or alleles
  - Loss of varieties

Source: Second FAO Report 'State of the worlds plant genetic resources for food and agriculture', 2010

Key questions

- What is the level of genetic diversity in modern elite breeding populations?
- Does 'exotic' breeding material have a higher level of genetic variation that can be introgressed into modern varieties?

Future questions

- What part of genetic diversity is useful or deleterious?
- How can useful genetic variation be identified?

What are plant genetic resources (PGR)?

- 4th international technical conference of FAO on PGR, Leipzig 1996:
  ... *generatively or vegetatively reproducible material of plant of current or potential value, including landraces, related wild sprecies and wild forms and special genetic material of crop plants*

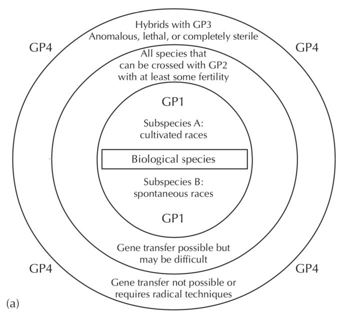
- A simpler definition by Becker (1993):
  ... *the complete genetic material that is available for breeding of a crop plant.*

Sources of new genetic variation

Plant Genetic Resources for Food and Agriculture (PGRFA)

- New and minor crops (“Neodomestication”)
- Close relatives
- Wild ancestors
- Land races
- Modern elite varieties of different geographic origin
- Induced mutations

Sources of genetic diversity: Crop gene pools



Sources of genetic diversity for canola (*Brassica napus*)

Primary gene pool:

- Modern elite varieties of canola (*Brassica napus*)
- Landraces
- German Brassica napus

Secondary gene pool:

- *Brassica rapa*
- *Brassica oleracea*
- *Brassica nigra*

Tertiary gene pool:

- *Raphanus*
- *Crambe*
- *Arabidopsis*

Wikipedia

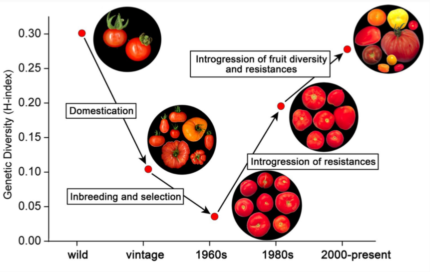
Brassica napus

Keye questions in the context of PGR

1. Level of genetic diversity in wile and modern elite breeding populations?
2. Higer level of variation in exotic material for introgression?
3. What type of genetic diversity is useful or deleterious?
4. How can useful genetic variation be identified?

Successful integration of new genetic variation

Genetic diversity in Dutch tomato varieties



Schouten et al., Front. Plant Sci (2019)

Plant genetic resources

The paradox of genetic diversity

Loss of genetic variation by breeding

Quantification of genetic variation

Genotype and allele frequencies

- Assumption: Diploid species
- Locus A, B, C, ...
- Alleles at locus A: A_1,A_2,...
- Frequency of alleles: $A_1 : p_1$, $A_2 : p_2$

Assuming one locus and two alleles:

Genotype:	A_1A_1	A_1A_2	A_2A_2
Relative genotype frequency:	x_{11}	x_{12}	x_{22}

sum of relative genotype frequencies: $x_{11} + x_{12} + x_{22} = 1$

Methods to quantify genetic variation

Single Markers:

- Number of alleles, a or allele richness, A
- Gene diversity or Hardy-Weinberg heterozygosity, H_E
- Biallelic SNPs: Minor allele frequency (MAF)

Multiple markers:

- Average gene diversity or mean heterozygosity
- Number of haplotypes
- Haplotype diversity

Complete DNA sequences:

- Number of segregating sites, S
- Nucleotide polymorphism, P
- Nucleotide diversity, π

Gene diversity or Hardy-Weinberg heterozygosity

Definition: Calculate the expected frequency of heterozygous genotypes under the Hardy-Weinberg-Equilibrium given the frequencies of the different alleles at a locus

- Theoretical measure:
$$H_E = 1 - \sum_{i=1}^{a} p_i^2$$
where p_i is the frequency of allele i and a is the number of different alleles.
- For small sample sizes ($n < 50$), the following correction is recommended :
$$\hat{H}_E = \frac{n}{n-1}(1 - \sum_{i=1}^{a} \hat{p}_i^2)$$
($\hat{p}_i$= relative frequency of allele i in the sample)

Analysis of multiple markers

- Average gene diversity or mean heterozygosity:

$$\hat{H} = \frac{1}{L}\sum_{i=1}^{L} \hat{H}_{E_i}$$

where L is the number of markers (or loci) analyzed

- Haplotype: The combination of alleles on a chromosome in a given individual
- Haplotype count: Number of different haplotypes formed by the markers
- Haplotype diversity:

$$HD = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2)$$

with n as the number of individuals (sequences), k as the number of different haplotypes, and p_i as the relative frequency of haplotype i

---

A DNA sequence alignment

Alignment 1:

```
seq1  GATCTATATA
seq2  GAACTATATA
seq3  CATCATCATA
seq4  GACCTATATC
```

Parameters:

- $n = 4$ sequences
- $L = 10$ sequenced nucleotide positions
- $S = 6$ polymorphic nucleotide positions

Quantification of DNA sequence variation

Nucleotide polymorphism:

$$P = \frac{S}{L}$$

- S: number of polymorphic nucleotide positions (segregating sites)
- L: total number of sequenced nucleotide positions

From the previous slide:

$$P = \frac{6}{10} = 0.6$$

Quantification of DNA sequence variation

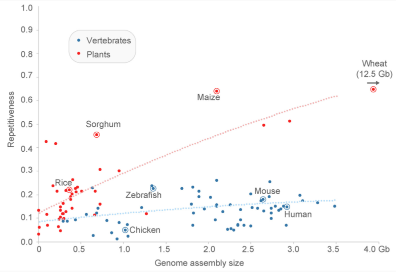
Nucleotide diversity:

$$\pi = \frac{n}{n-1} \sum_{i=1}^{k} \sum_{j=1}^{k} p_i p_j \pi_{ij}$$

- n: number of individuals
- k: number of different haplotypes
- p_i: proportion of haplotype i
- p_j: proportion of haplotype j
- π_{ij} as pairwise differences of haplotypes i and j. The difference is measured as the proportion of different SNPs between two haplotypes. If 4 out of 10 sites are different between two haplotypes, $\pi_{ij} = 0.4$

Limitations of current estimates

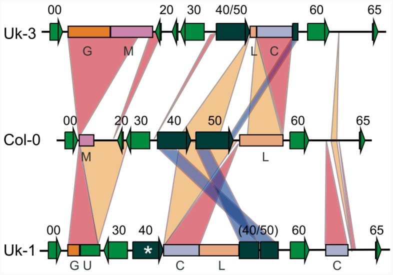
Plant Genomes are highly repetitive



Jiao and Schneeberger (2017)

Limitations of current estimates

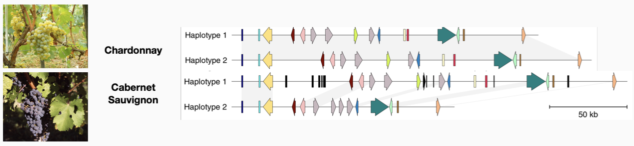
Rearrangments and gene duplications



Bomblies et al. 2007

Limitations of current estimates

Diversity in grapevine genomes



Zhou et al. (2019)

References i

Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, Cantu D, Gaut BS (2019) The population genetics of structural variants in grapevine domestication. Nature Plants 5(9):965–979, DOI 10.1038/s41477-019-0507-8, URL `http://www.nature.com/articles/s41477-019-0507-8`