

The road to sorghum domestication

Concetta Burgarella^{1,2} (concetta.burgarella@gmail.com)

A. Berger^{1,2}, N. Terrier^{2,3}, A. Donkpegan^{1,2}, V. Ranwez^{2,4}, N. Chantret^{2,3}, S. Glémin⁵, M. Deu^{1,2}, D. Pot^{1,2}

¹ CIRAD, UMR AGAP, F-34398 Montpellier, France, ² AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France, ³ INRA, UMR AGAP, Montpellier, France, ⁴ Montpellier SupAgro, UMR AGAP, Montpellier, France, ⁵ CNRS, UMR ISEM (Institut des Sciences de l'Evolution Montpellier), Montpellier, France

Context

- Sorghum bicolor* is the fifth cereal worldwide for grain production and the staple food for 500 million people.
- It is a C4 species (as maize and sugarcane) native of North Eastern Africa. Due to its **adaptation to heat and drought** stress, its importance is expected to increase in the current context of global warming [1].
- The cultivated pool is characterized by a clear domestication syndrome (plant architecture, shattering, dormancy and seed weight; Fig. 1) and a **great phenotypic and genetic diversity** [2].
- In spite of its agronomic importance, the evolutionary dynamics of sorghum domestication and the genetic bases of adaptation to the different environmental contexts are not well understood.

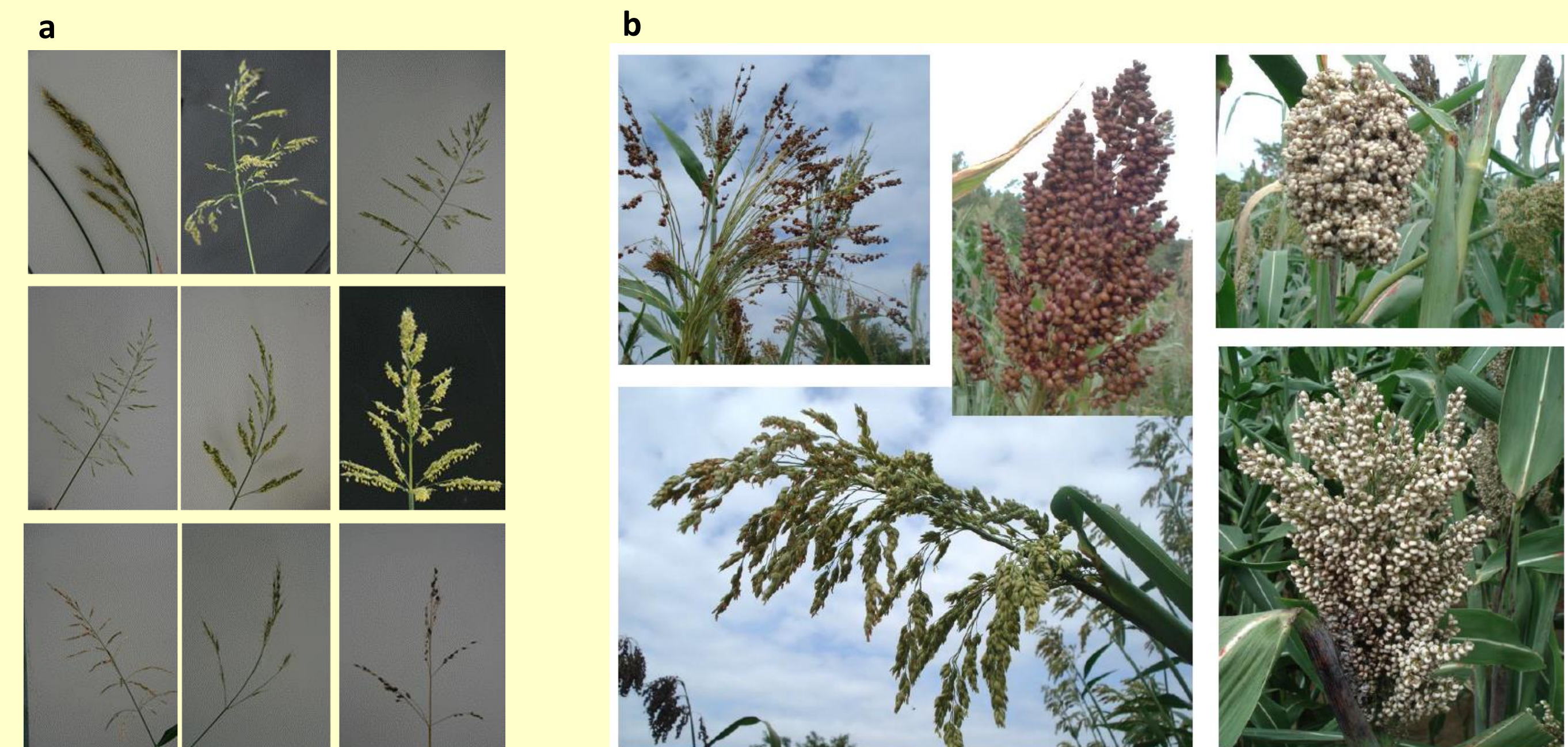


Fig. 1 - Wild (*S. b. ssp. verticilliflorum*, a) and cultivated (*S. b. ssp. bicolor*, b) sorghum panicles

Questions

- Is there an effect of domestication on gene expression?
- If yes, which genes and metabolisms are concerned?
- Are differentially expressed genes under selection?

Materials and Methods

- The transcriptome of leaf, flower and stem was sequenced in 9 wild and 11 landrace accessions.
- Reads were mapped on the Sorghum bicolor genome v.3.1. Differential expression analysis was performed with the new-tuxedo pipeline [3] and edgeR [4]. Genes were considered Differentially Expressed at a 1% False Discovery Rate.
- Genotype calling was done following the strategy proposed by [5] and polymorphism estimates were calculated with dNdSpiNpiS [6].

Table 1 - Expression and diversity of significantly differentially expressed (DE) versus non-DE genes

	Genes (DE analysis)	Ratio genes overexpressed in Crop / overexpressed in Wild	Genes (polymorphism analysis)	Crop			Wild		
				π_N	π_S	π_N/π_S	π_N	π_S	π_N/π_S
DE	949	0.23 ¹	469	0.49	3.76	0.13	0.60	4.68	0.13
non-DE	23698	1.1 ¹	9867	0.50	3.35	0.15	0.66	4.06	0.16

¹ Significantly different χ^2 test p-value < 2.2e-16

π_N : non-synonymous polymorphism; π_S : synonymous polymorphism. In 10³ units.

Table 2 – Gene Ontology terms most represented among DE genes

	Ontology	GO.ID	Term	P-value
Genes Overexpressed in Crop	Molecular.Function	GO:0005524	ATP binding	7.90E-08
	Molecular.Function	GO:0008017	microtubule binding	1.20E-07
	Molecular.Function	GO:0005086	ARF guanyl-nucleotide exchange factor activity	1.90E-05
Genes overexpressed in Wild	Biological.Process	GO:0015979	photosynthesis	9.80E-23
	Molecular.Function	GO:0016655	oxidoreductase activity, acting on NAD(P)H or quinone	5.70E-07
	Biological.Process	GO:0015986	ATP synthesis coupled proton transport	1.00E-07

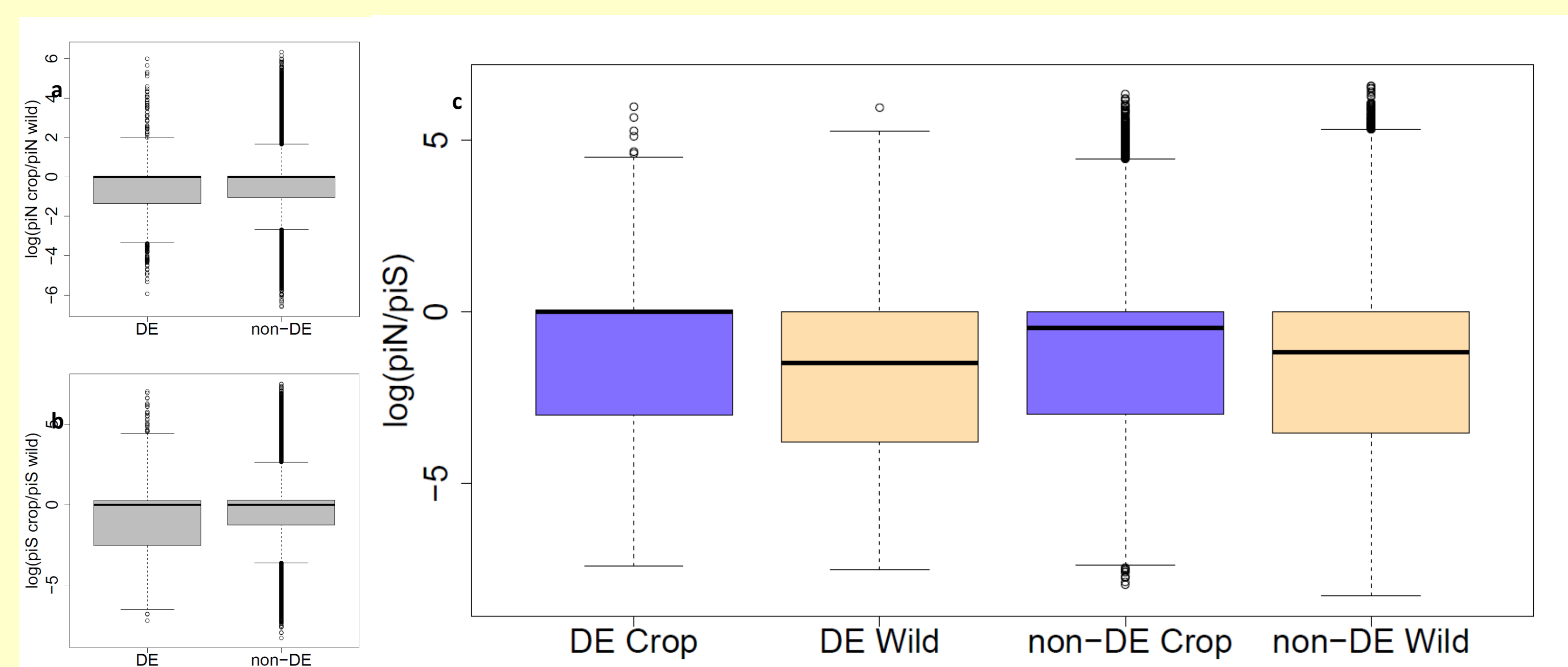


Fig. 2 – Comparison of polymorphism levels between DE and non-DE genes in crop and wild samples. Ratio crop/wild of π_N (a), π_S (b). π_N/π_S (c).

Results

- Changes in expression levels are observed in **949 genes**, which are mostly overexpressed in wild accessions (77 %, Table 1).
- Genes involved in **photosynthesis and in reduction-oxidation** processing of seed storage proteins are over-represented (Table 2).
- Among the DE, two genes for which functional evidences of impact on the phenotype are available have been identified: **dwarf2**, that regulates the stem internode length [8], and a **phosphoribulokinase**, involved in step 2 of photosynthesis [9].
- Selection against deleterious mutations** is stronger in DE genes than in non-DE genes for the wild compartment (lower π_N/π_S), but the contrary is observed for the crop pool (Fig. 2).
- Lower polymorphism diversity** (Table1) and **variability of the isoform expression balance** (expression ratio of the alternative proteins coded by a single gene) are observed in the crop pool (Fig. 3).

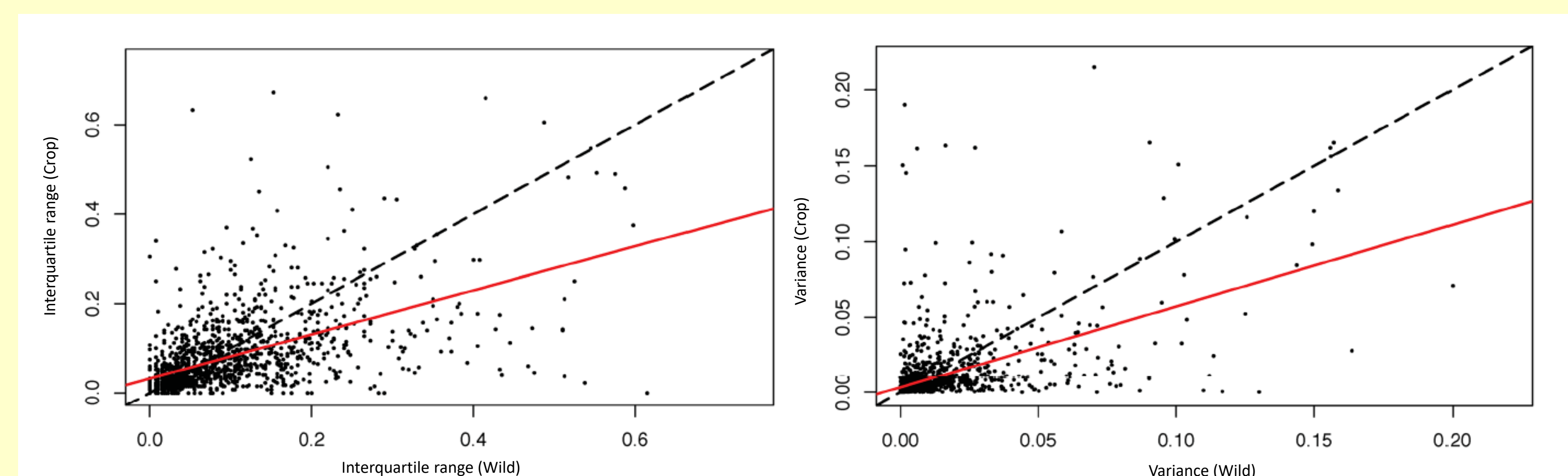


Fig. 3 – Variability of isoform expression balance in crop (y-axis) against wild (x-axis) samples [7]. Points are single genes, x=y black dashed line, linear interpolation (red line).