Genome-wide differentiation between male and female gene pools in sugar beet elite material

Authors: DESPREZ Augustin[1,2], HENRY Karine[3], DESPREZ Bruno[1], DEVAUX Pierre[1], CHARCOSSET Alain[2], TENAILLON Maud[2] and MOREAU Laurence[2]

1 Florimond Desprez Veuve & Fils SAS, BP 41, 59242, Cappelle-en-Pevele, France.
2 GQE - Le Moulon, INRAE, CNRS, AgroParisTech, Université Paris-Saclay, 91190, Gif-sur-Yvette, France
3 United Beet Seeds (UBS), Industriepark 15, 3300 Tienen, Belgium

Context and objectives

Oldest records of plants with expanded roots date around 3500 BC in Egypt and are thought to originate from the Near East (figure 1). Sugar beets diverged from fodder and wild beets in the late 17th century. In the 1940s, USDA researchers made **two key discoveries**: genes controlling **monogermy** of seeds (released in cultivar **SLC101** in 1950 by V.F. Savitsky), and **cytoplasmic male sterility** (F.V. Owen, 1945), enabling hybrid production without mechanical castration (figure 1). The challenges associated with the use of these traits were threefold: (i) developing high-performing monogerm varieties with female parents homozygous for the monogermy allele, (ii) establishing breeding methods to produce hybrids from male-sterile (female) and pollinator (male) pools and (iii) overcoming strong inbreeding depression.

To tackle these challenges, the conventional breeding of sugar beet relies on three-way hybrids that involve first crossing a **male-sterile (MS)** line with a **Type-O maintainer** (referencing F.V. Owen) to generate a female hybrid (MSF1), that is further crossed with a **male** (pollinator) (figure 2) (Hogaboam, 1957).

Using data from Florimond Desprez, the objective of our study is to investigate the genetic structuring into gene pools and the genomic impact of three-way hybrid selection through sugar beet breeding history.
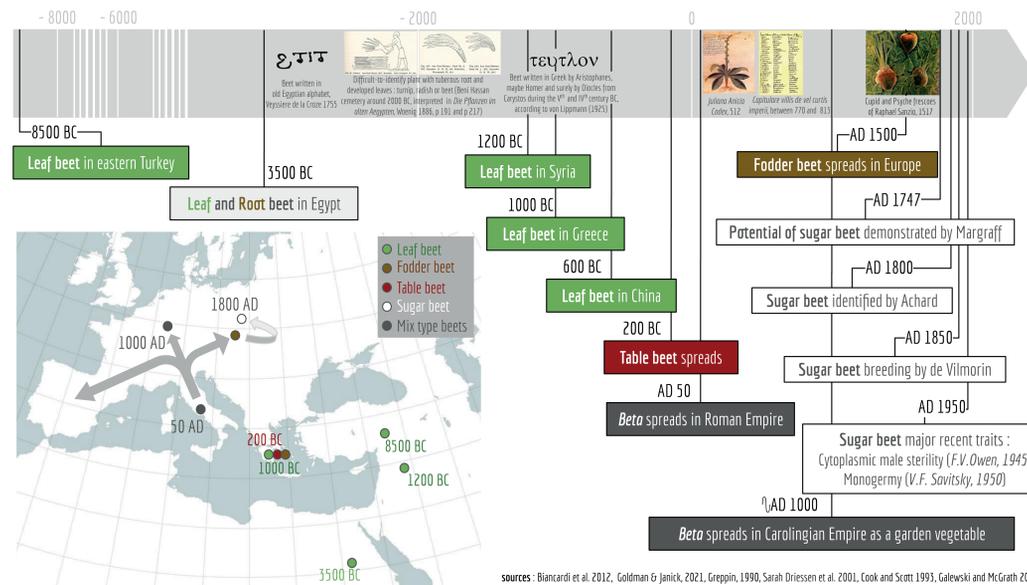

Figure 1: **From domestication to recent breeding, a glimpse through sugar beet breeding history** (BC: Before Christ, AD: Anno Domini = Common Era)

Material and methods

Material

Plant material includes over twenty years of germplasm and recent commercial hybrids. **30,771 plants** representing **3k accessions** were genotyped using 12,507 bi-allelic markers. Markers were mapped to the scaffolds of the 568.8 Mb reference genome EL10.2 (J.M. McGrath et al. 2020, 2023), of which **10,046** (80.3%) were retained (<25% missing data).

Methods

One consensus genotype per accession was obtained using maximum likelihood. **Genetic structuring and detection of outlier loci** were performed using Principal Component Analysis (PCA) computed with PCAdapt (Luu K. et al. 2017, Privé F. et al. 2020). **Genetic differentiation** among gene pools was measured by F_{ST} (Weir & Cockerham, 1983), using VCFtools. **Genetic diversity** within genetic pools was measured using the Modified Rogers' Distance (MRD) (J.S Rogers, 1991). We considered three periods based on the males' years of selection (1998-2004, 2005-2011 and 2012-2018), covering 21 years.
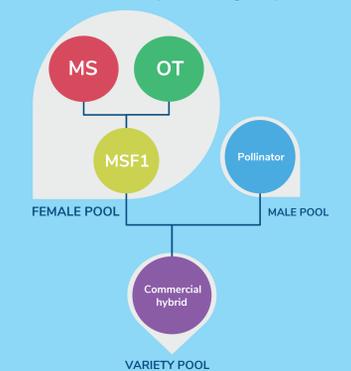

Figure 2: Three-way hybrid schemes used in sugar beet breeding. (MS: Male Sterile ; F1: first generation hybrid; OT: Owen-type nucleo-cytoplasmic genetic sterility maintainer)

Working hypothesis

Compared with genome-wide values (figure 3 A), we expect
- loci with additive effects for which the favorable allele is recessive, to be selected similarly in both hybrid parents and therefore to **display low differentiation and low genetic diversity** (figure 3 B).
- loci involved in hybrid vigor, to display a **high degree of differentiation that increases through time, between male and female pools** as well as low intra-pool genetic diversity (figure 3 C).

Because of the reduction of gene flow between pools, we expect an increase of divergence, even without intra-pool selection. We therefore rely on genome-wide levels of diversity and differentiation to detect outlier loci, using statistics such as Modified Roger's Distance within pools (MRD) and F_{ST}.

The introduction of new traits of interest, using more or less distant donors, might locally disrupt the structuring of male and female pools as can be revealed by tools such as PCAdapt.
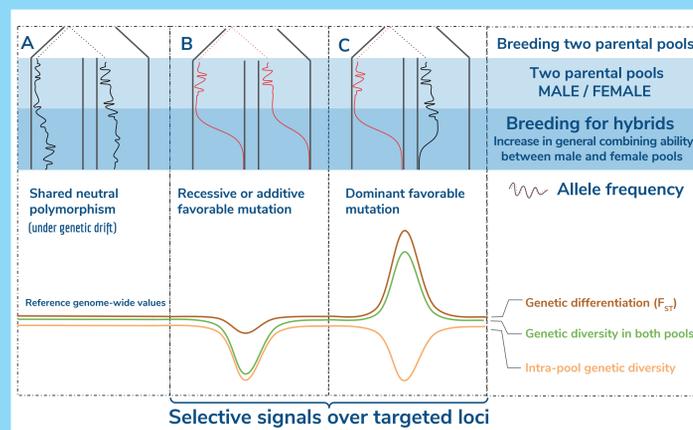

Figure 3: Detection of selection in a hybrid breeding context. Allele frequency evolution (upper panels) along with statistics used for detecting selection (lower panels) are shown. Expected changes in allele frequency and the resulting genomic patterns under selection (right panels) are compared to genome-wide reference values without selection (left panel).

Breeding structures genetic diversity into pools

The first PCA axis differentiates male and female gene pools, hybrids being intermediate (figure 4). The second axis is linked to the males' year of selection (not shown). Differentiation (F_{ST}) between males and females increases over 21 years (figure 5) while diversity slightly decreases (figure 6).
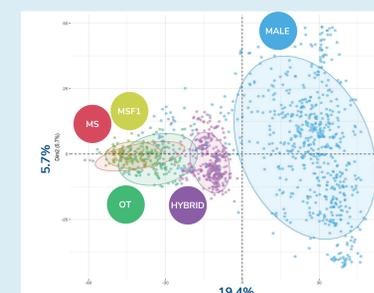

Figure 4: Principal Component Analysis (PCA) using 10k SNPs and colored by pool (ellipses cover 75% of individuals within each pool).
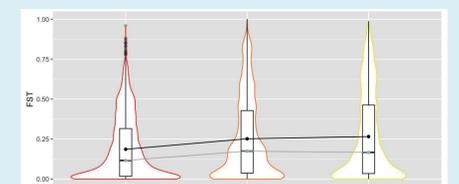

Figure 5: F_{ST} computed on 10k SNPs over 7-years periods (3) covering 21 years total over 7-years periods (3) covering 21 years total
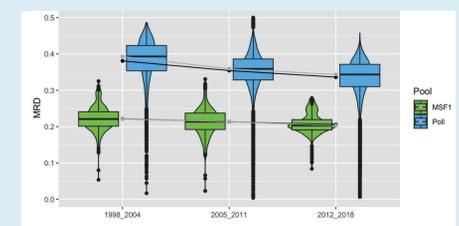

Figure 6: MRD computed on 10k SNPs within males (Poll) and females (MSF1) over 7-years periods (3) covering 21 years total

Large genomic regions under selection

Structure is highly **heterogeneous along chromosomes** (figure 7). Past selection and trait introductions shape genome-wide differentiation (figure 8). Some regions display patterns of consistent evolution across breeding periods (continuous increase).

Monogermy loci on chromosome 4 are associated with a very narrow but distinct pattern of differentiation between males and females (result not shown), which matches with an old introduction (high recombination on each side of the selected haplotype).
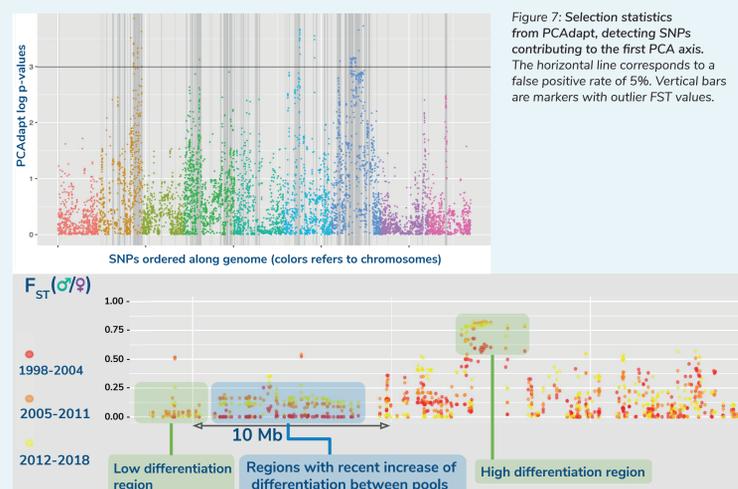

Figure 7: Selection statistics from PCAdapt, detecting SNPs contributing to the first PCA axis. The horizontal line corresponds to a false positive rate of 5%. Vertical bars are markers with outlier FST values.

Figure 8: F_{ST} (Weir & Cockerham, 1983) between males and females along a 30 Mb region of chromosome 2

Perspectives

Identifying the key aspects of the resulting evolution of genetic structuring, and the **impact of successive introductions is a first step in developing more efficient breeding and pre-breeding methods.** Focusing on signatures of selection will allow associating patterns with major genes and known Quantitative Trait Loci (QTLs). Narrow genetic diversity in some regions may compromise long-term genetic gain, and increases the genetic vulnerability to unpredictable environmental conditions. It has been observed with recurrent introductions of improved donors that it is possible to **maintain the genetic diversity and increase mid- and long-term performances with only limited short-term penalty** in a hybrid context (Allier et al., 2020). We aim at comparing simulated breeding programs to (i)consider genomic selection and optimal cross selection to recurrently improve genetic resources (i.e. pre-breeding), (i) bridge the improved genetic resources with elites (i.e. bridging) and (i) directly manage introductions into the elite breeding population.

REFERENCES

Thanks to Kevin Dorn (USDA-ARS), for his valuable advice and constructive external perspective.

Savitsky, V.F. (1950). Monogerm sugar beets in the United States. Proceedings of ASSBT 6: 156–159.
Owen, F.V. (1945) Cytoplasmically inherited male sterility in sugar beet. Journal of Agricultural Research, 71, 423–440.
Hogaboam, G.J. (1957) Factors influencing phenotypic expression of cytoplasmic male sterility in the sugar beet (Beta vulgaris L.) J.Am. Soc. Sugar Beet Technologies. 9:457-465.
Alachiotis, N., Pavlidis, P. (2018) RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Commun Biol 1, 79.
McGrath, J. M. et al. (2020) A contiguous de novo genome assembly of sugar beet EL10 (Beta vulgaris L.). 2020.09.15.298315.
Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution, 38(6), 1358–1370.
Rogers, J.S (1991). A Comparison of the Suitability of the Rogers, Modified Rogers, Manhattan, and Cavalli-sforza and Edwards Distances, Systematic Biology, Volume 40, Issue 1, Pages 63–73
Privé, F., Luu, K. et al. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. Molecular Biology and Evolution.
Luu, K., et al. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources, 17(1), 67–77.
Wright, S. (1949) The Genetical Structure of Populations. Annals of Eugenics 15, 323–354.
Malécot, G. (1948) Les mathématiques de l'hérédité. Masson & Cie. Paris.
Allier, A., Teyssèdre, S., Lehermeier, C. et al. (2020) Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics 21, 349.