GROWING STRAWBERRY FROM THE EQUATOR TO THE ARCTIC

GENOFRA: High-density genotyping of *Fragaria ×ananassa* and wild relatives

Anja Kögler[a], Torsten Wenke[b], Stefan Wanke[a], Falk Zakrzewski[b], Konrad Grützmann[b], Henning Wagner[c], Yannick Wolfsteiner[c], Christoph Neinhuis[a], and Klaus Olbricht[c]

[a] Dresden University of Technology, Institute of Botany, Zellescher Weg 20b, 01217 Dresden

[b] Asgen GmbH & Co. KG, Egon-Erwin-Kisch-Str. 6, 01069 Dresden

[c] Hansabred GmbH & Co. KG, Radeberger Landstraße 12, 01108 Dresden

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

asgen GmbH & Co. KG

TECHNISCHE UNIVERSITÄT DRESDEN

hansa bred

GENOFRA: High-density genotyping of *Fragaria* ×*ananassa* and wild relatives

Abstract

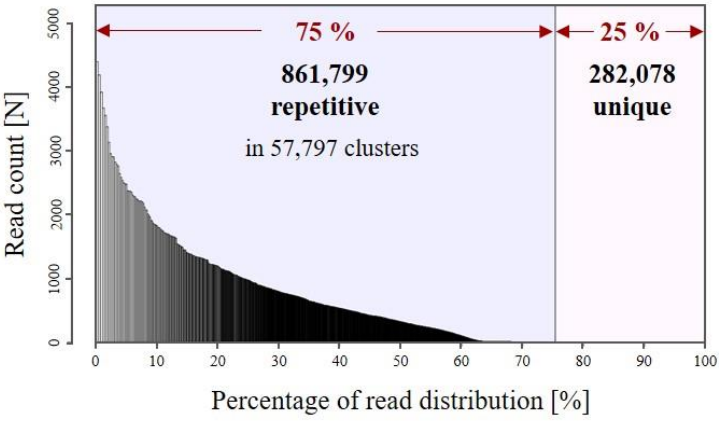
The strawberry *Fragaria* ×*ananassa* has been cultivated for about 260 years since their ancestor species were introduced to Europe in the 17th and 18th century. In the past selective breeding strategies predominantly focused on performance parameters such as yield, fruit size, shelf life or ripening times resulting in an increasing impoverishment of other important properties such as aroma, taste and resistance to pests and abiotic stress. These negative domestication effects in cultivated plants are not only determined by the intensity of the breeding process, but also by low genetic diversity of the original breeding material.

The collaborative project GENOFRA aims to compare the genetic diversity among a *Fragaria* ×*ananassa* sample pool by high-density genotyping of the non-repetitive genome fraction. The sampling includes 185 cultivated varieties (cultivars and clone selections) and an established segregating F2 population (crossbred of *F.* ×*ananassa* 'Senga Sengana' with *F. chiloensis* followed by F1 selfing). The identification of the contributing genome portions as well as the original genetic resources of the *F.* ×*ananassa* crossing parents (*F. chiloensis* and *F. virginiana*) will enable the representation of allelic variations (recombination effects) and the depletion of genetic diversity during breeding history. Additionally, we will complement the aforementioned genomic resources by a collection of wild strawberry species aiming at the determination of alleles or allele combinations associated with important breeding traits to develop molecular markers for the selection of domestication-relevant performance parameters.

GROWING STRAWBERRY FROM THE EQUATOR TO THE ARCTIC

GENOFRA: High-density genotyping of *Fragaria ×ananassa* and wild relatives

Determination of the genomic repeat content



The *RepeatExplorer* software (Novák et al., 2010, 2013) was used to determine the genomic repeat content.

Approximately 75% of the strawberry genome consists of repetitive DNA.

Of the total ~1.1 million NGS reads analyzed, the repeat fraction was grouped into 57,797 clusters representing different groups of repetitive sequences (e.g. transposable elements, satellite DNA, rDNA loci).
Only 282,078 reads (25%) correspond to the unique genome fraction, including genes that will serve as the basis for genotyping-by-sequencing (GBS).

Novák, P., Neumann, P. and Macas, J. (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics, 11, 378.

Novák, P., Neumann, P., Pech, J., Steinhaisl, J. and Macas, J. (2013) Genome analysis RepeatExplorer : a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics, 29, 792–793.

GENOFRA:
High-density genotyping of *Fragaria* ×*ananassa* and wild relatives

In silico restriction digestion of the genome

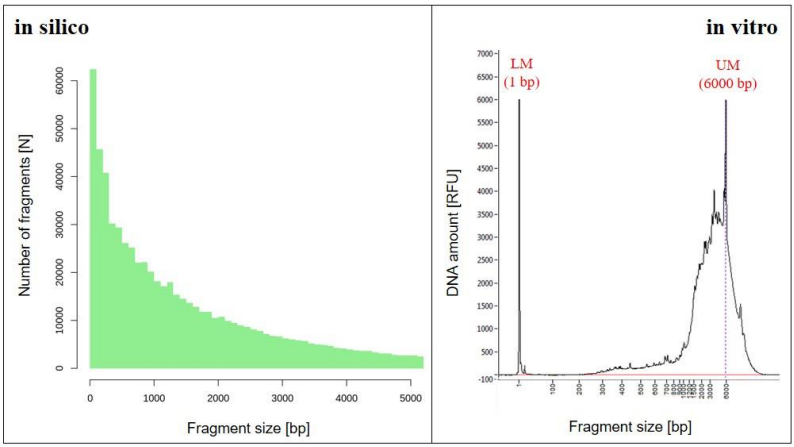
Prior to GBS, the number of expected DNA fragments per genome was determined for several restriction enzymes based on the reference genome ('Camarosa'[1]).

Restriction enzyme	Number of fragments
*Mse*I	5.164.886
*Psi*I	424.881
*Eco*RI	240.671
*Sap*I	55.479
*Srf*I	1.606

Of the 424.881 *Psi*I restriction fragments predicted by in silico digestion, 33,161 are between 400 bp and 600 bp (relevant size range for sequencing) which is a suitable amount of loci for GBS.

However, the size distribution of in silico analyses are always biased as the reference assembly mainly represents the smaller non-repetitive genome fraction.

Distribution of *Psi*I restriction fragment lengths



[1] www.rosaceae.org/analysis/322

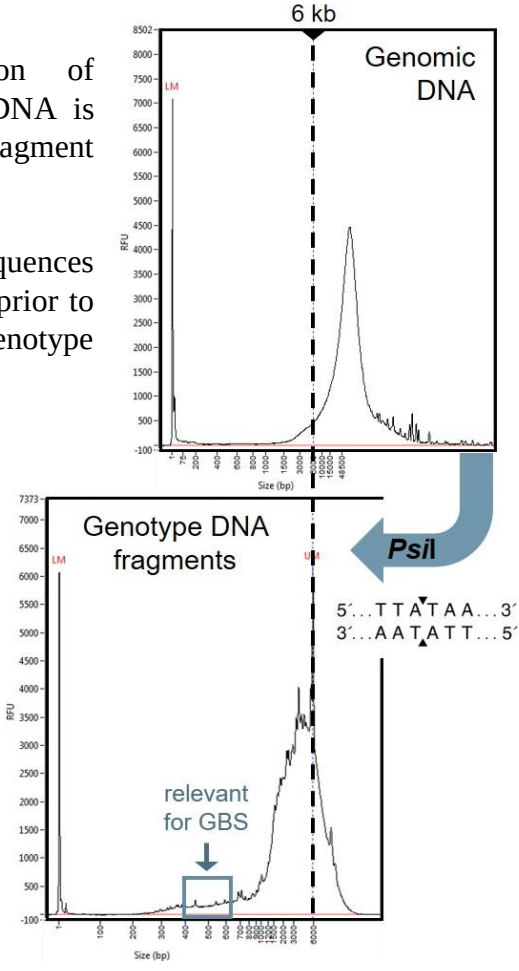
GROWING STRAWBERRY FROM THE EQUATOR TO THE ARCTIC

GENOFRA: High-density genotyping of *Fragaria* ×*ananassa* and wild relatives

*Psi*I digestion and depletion of repetitive DNA sequences from genotype pools

The *Psi*I digestion of genomic genotype DNA is evaluated by fragment length analysis.

Repetitive DNA sequences are largely removed prior to sequencing from genotype pools.



GROWING STRAWBERRY FROM THE EQUATOR TO THE ARCTIC

GENOFRA: High-density genotyping of *Fragaria* ×*ananassa* and wild relatives

Genotyping-by-sequencing

***Fragaria* genotypes**	**Count**
segregating F2 population	94
cultivated varieties	116
wild species	171
Total	**381**

DNA extraction

Genomic DNA

*Psi*I cleavage

Genotype DNA fragments

Transfer to 96-well plates

Genotyping-by-sequencing (GBS)

GROWING STRAWBERRY FROM THE EQUATOR TO THE ARCTIC

GENOFRA: High-density genotyping of *Fragaria* ×*ananassa* and wild relatives

Determination of GBS loci

Quality trimming/length trimming 100 nt

Filtering for restriction site (*PsiI* – TAA)

Clustering in vSEARCH

Filtering for global identity

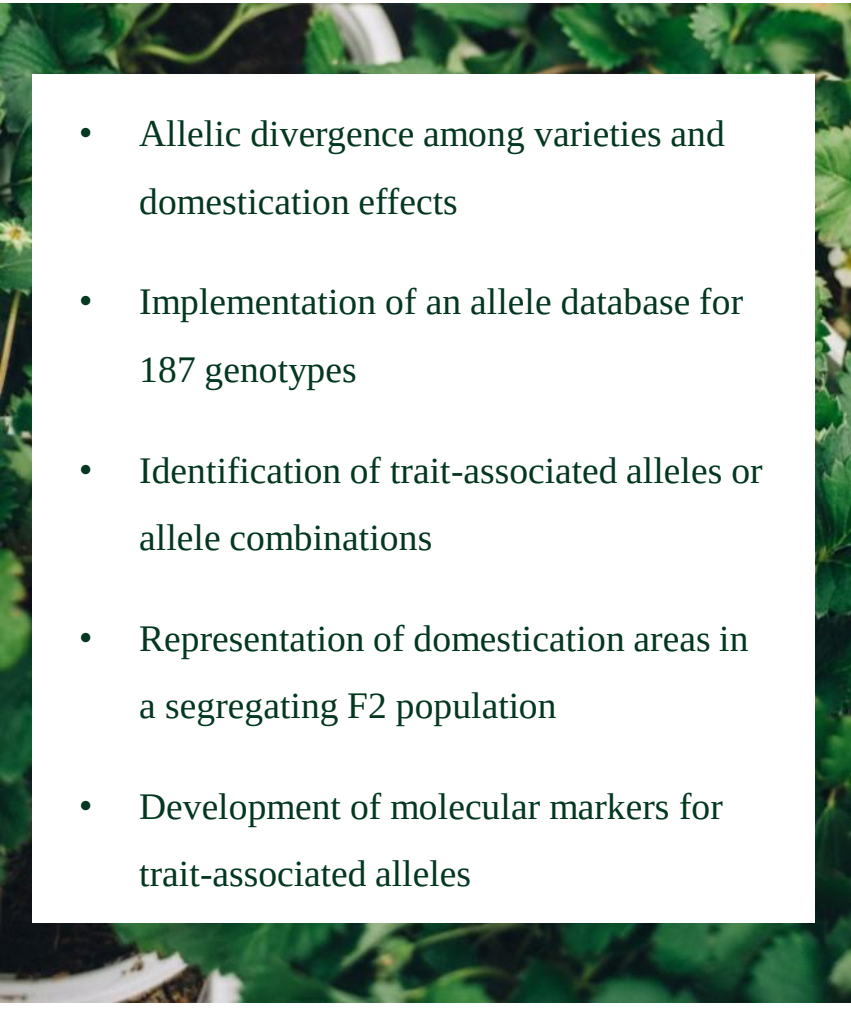
Determination number of alleles per locus

Extraction of alleles per locus

GROWING STRAWBERRY FROM THE EQUATOR TO THE ARCTIC

GENOFRA: High-density genotyping of *Fragaria ×ananassa* and wild relatives

Our high-density genotyping results will help to answer crucial issues:



- Allelic divergence among varieties and domestication effects
- Implementation of an allele database for 187 genotypes
- Identification of trait-associated alleles or allele combinations
- Representation of domestication areas in a segregating F2 population
- Development of molecular markers for trait-associated alleles