Molecular oil palm breeding at Sampoerna Agro

Enrique Ritter1* and D Asmono2

1NEIKER-Tecnalia, Vitoria-Gasteiz, Spain
2Sampoerna Agro Tbk PT, Palembang, Indonesia

*E-mail: eritter@neiker.eus

Abstract. Sampoerna Agro and NEIKER perform jointly since many years collaborative R&D projects related to Molecular breeding and Marker assisted selection in Oil palm in order to accelerate the oil palm breeding program and to improve its efficiency. The detection of candidate genes for useful traits offers the possibility to apply them – after developing the corresponding markers – in marker assisted selection (MAS) within breeding programmes. The survey of the allelic diversity of such genes in available germplasm and the analyses of their particular effects permits to select the most efficient alleles or allele combinations for these purposes. Within this project, we have identified useful candidate genes for different productive and quality traits as well as for stress tolerance using different molecular tools. We have characterized the allelic variation in this germplasm in order to develop marker assisted breeding strategies through Model building. The developed Models which show large correlations between predicted and observed values explaining large portions of the total variance can be used to assign breeding values to parents and to predict progeny performances of their crosses. In this paper, we summarize the R&D activities which are being performed and practically implemented at Sampoerna.

1. Introduction

Sampoerna Agro and NEIKER perform jointly since many years collaborative R&D projects related to Molecular breeding and Marker assisted selection in Oil palm in order to accelerate the oil palm breeding program and to improve its efficiency. Molecular breeding can accelerate considerably breeding programmes, particularly in oil palm with long vegetative cycles and extended selection periods. The detection of candidate genes for useful traits offers the possibility to apply them – after developing the corresponding markers – in marker assisted selection (MAS) within breeding programmes [1]. The survey of the allelic diversity of such genes in available germplasm and the analyses of their particular effects permits to select the most efficient alleles or allele combinations for these purposes. Within this project we have identified useful candidate genes for different productive and quality traits as well as for stress tolerance using different molecular tools; we have characterized the allelic variation in this germplasm in order to develop marker assisted breeding strategies through Model building [2]. The developed Models which show large correlations between predicted and observed values explaining large portions of the total variance can be used to assign breeding values to parents and to predict progeny performances of their crosses. In this paper, we summarize the R&D activities which are being performed, as well as their concrete, practical implementation at Sampoerna.
Useful traits of interest (such as yield, oil production, oil quality and tolerances to biotic and abiotic stresses) are controlled by the effects of one, several or many different, so called Candidate Genes (CG) in each case. Their numbers depend on the particular trait. The particular phenotype of an individual depends on the genotype and is controlled by the effects of the specific alleles of the candidate genes of the corresponding trait and possible interactions [3]. In order to analyse the effects of the different, existing alleles of candidate genes it is necessary to link phenotypic and genotypic data, applying statistical methods such as Association Mapping.

2. Materials and Methods

2.1 Plant material
a) Parental lines for producing Tenera seeds through DxP crosses belonging to different Pisifera and Dura origins.

b) 440 DxP progenies descending from selected palms from above, which have been extensively characterized (over 15 years) with respect to production traits, bunch component traits and other traits of interest. 220 selected genotypes of these progenies with the positive and negative expression for traits of interest have been used for molecular characterization.

c) New germplasm accessions from recent prospections in Nigeria, Angola and Cameroon.

d) Accessions from Field Trials for evaluating drought resistance and nutrient use efficiency.

2.2 Candidate gene (CG) detection different tools
- Bulked Segregant Analyses coupled to differential cDNA-AFLP Analyses which allow the direct detection of potential CG.

- Co-location Analyses of sequenced markers, QTL and mapped genes of a high-density oil palm reference map [4].

- Known Candidate genes from oil palm or other species or key enzymes from relevant metabolic pathways (Palm Oil Biosynthesis, Transcription factors) using in silico mining or other Bioinformatics approaches.

2.3 Molecular analyses
For the molecular analyses, Primers were designed for each CG to amplify CG exon sequences using Primer 3 Software and validated first on a small set of genotypes.

Each CG was first amplified with a fusion primer pair containing a CG specific part and a universal part. Subsequently, each sample was re-amplified with a so called Key Primer composed of a part specific to the sequencing machine, a “Barcode” (=MIDS or Index) for each genotype and the universal part of the fusion primers. Barcodes and CG specific parts identify the origin of the sample. Samples were pooled and sent for sequencing using ION torrent, ION proton or MySeq Illumina platforms.

For Data analyses, in house developed Software was used for sequence processing, analyses of the allelic variability of CG in the collection, determination of allele composition of the evaluated genotypes and determination of effects of CG alleles and allele combinations through association mapping (AM). In addition, whole genome association studies (WGAS) and genomic selection (GS) methods are applied to the available data.

For special genes of interest, such as the Shell (Sh) gene or the Virescence (Vir) Gene, allele-specific primers were designed for routine selection of desired plant materials.

3. Results and Discussion
Hundreds of CG were identified, processed and analysed as described above. Numerous significant effects of alleles and allele combinations (AC) were detected for different CG for all traits of interest.

Beside the analyses of individual CG, also their cumulative effect was analysed by model building. Based on average allele and AC values, trait specific breeding values of the parental lines were
determined based on the allele composition of the parents for each CG. Moreover, progeny performance predictions were computed based on average parental breeding values or expected AC values for all CG. Highly significant correlations between the predicted and observed progeny performances were detected always.

The general applicability of the concept and the developed models was validated using subsets of parental data and even external data predicted with the allele or AC values of the original models. Based on these results, predicted progeny performance (PPP) matrixes for all possible crosses between available Dura and Pisifera parents were computed for each trait. These analyses were extended to combined predictions for multiple traits based on breeder preferences. Crosses were ranked based on these final scores and the resulting TOP crosses were extracted (Figure 1).

Based on these predictions, three comparative classical and molecular breeding field trials have been established targeting different combinations of traits of interest and involving a total of 48 crosses. They will be evaluated in the near future.

Figure 1. Visualization of TOP crosses for combined trait scores (A=12 points, B, C, D = 11,10 and 9 points, respectively).

Another example for the practical implementation of Molecular Breeding at Sampoerna is the routine screening of plant material from the breeding program with allele specific primers for the \( SH \) gene and Vir gene, respectively. For 2017 the analyses of a total of 1800 samples are targeted.

Pisifera parents are obtained from TxP or TxT crosses. The identification of the fruit type already in seedling stage allows to select and plant only the desired Pisifera genotypes and to save 50 or 75% of land for cultivation. A similar example represents the D\(x\)T crosses to evaluate the performance of Tenera genotypes.

The Vir allele screening for developing virescence cultivars allows detecting even the desired homozygous genotypes from crosses between heterozygous parents, which using only classical breeding would require one additional generation of selfings.

For preselecting of new germplasm from Nigeria, Angola or Cameroon, also simple molecular markers such as SSR are applied to select appropriate subsets of diverse materials based on the results of Cluster analyses and resulting dendrograms [2] (Figure 2).
Figure 2. Example of pre-selection of plant materials based on Cluster analyses derived from SSR markers.

Moreover, *Ganoderma* Tolerance Screenings are performed in the breeding material using promising candidate genes and the new traits of interest are targeted such as the ability for High Density Planting (HDP) and Sex Ratio.

4. Conclusions

In classical breeding, the breeding value of a parent is deduced from the progeny performances he is involved. Unfortunately, in oil palm, only a reduced number of progenies are available to evaluate a parent (2-3 crosses for Dura, 6-8 crosses for Pisifera parents).

However, many parents share the same allele of a particular CG. Parental alleles of CG are known and the information of all available progenies can be used to calculate the effects of CG alleles or allele combinations, providing more information in this way.

Molecular breeding allows to select already at the seedling stage the genotypes of interest and to plant only the desired plant material. The next Breeding Cycle can be already initiated at flowering time and it is also possible to plant only the most promising Crosses based on improved progeny performance predictions.

Finally, by inspecting carefully the allelic composition of genotypes, it is possible to pyramidalize favourable alleles of CG in the parents and to apply within selection in progenies.

5. References

[1] Crossa J, de los Campos G, Perez P, Gianola D, Burgueño J, Araus J L, Makumbi D, Singh R P, Dreisigacker S, Jianbing Yan, Arief V, Banziger M and H-J Braun 2010 *Genetics* **186** 713

[2] Ritter E, Pratiwi E, Lopez de Armentia E, Herrero J, Ningrum P Y, Santika B, Yulismawati, Sarimana U, Sembiring Z, Asmono D and Hernandez M *Euphytica* **2** 207

[3] Durstewitz G, Polley A, Plieske J, Luerssen H, Graner EM, Wieseke R and Ganal M W 2010 *Genome* **53** 948

[4] Wang N, Chen B, Xu K, Gao G, Li F, Qiao J, Yan G, Li J, Li H and Wu X 2016 *Front Plant Sci.* **24** 338