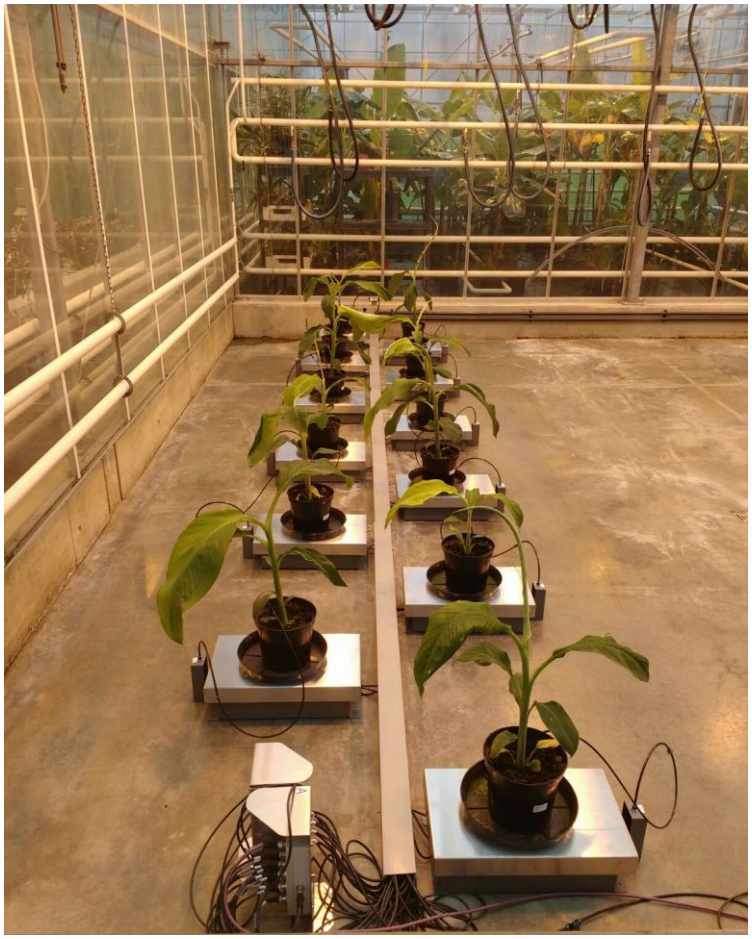
More fruit for food security: developing climate-smart bananas for the African Great Lakes region

TECHNICAL REPORT: Jan 2018 – July 2019



Simulating drought in banana plants, S. Carpentier, Bioversity

Executive summary

Banana (*Musa* spp.) is an extremely important crop in the Least Developed Countries, providing a staple food for more than 400 million poor people. Drought stress globally causes estimated yield losses of 37-61 million tonnes annually, worth at least US$13-22 billion, with a third of these losses incurred in Africa.

The overarching goal of the project is to identify drought-tolerant banana cultivars to help close the yield gap in drought-prone areas. This project focus on the African Great Lakes region, which is one of the poorest areas in the world, encompassing Burundi, north-eastern Democratic Republic of Congo (DR Congo), western Kenya, Rwanda, north-western Tanzania, and Uganda. The region is also one of the most important banana-growing areas in the world and is home to a unique group of banana cultivars that are well adapted to this unique colder highland environment.

To support this research, we rely on the Bioversity International's Musa Germplasm Transit Centre (ITC), hosted at the Katholieke Universiteit Leuven (KU Leuven) in Leuven that houses germplasm samples representing more than 1000 banana cultivars and 29 *Musa* wild relatives, which together provide a substantial potential genepool to combat drought stress.

As described in the original proposal, the first year of the project was focused mostly on two main Work packages (1 and 4).

- Deploying our high-throughput screening methods to assess the growth potential of the banana cultivars in the ITC collection for the agro-ecological zone of the African Great Lakes region under drought stress (WP1).
- Complementing the phenotyping by molecular analyses to unravel the molecular basis of mechanisms behind banana's drought tolerance, to support future high-throughput discovery of drought-tolerant cultivars and breeding (WP4).

Significant progresses were achieved in the first year that will lead to at least two peer review open access publications next year. Therefore, the technical report follows the structure of the manuscripts that are currently in preparation.

Table of content

Evaluation of the 'bananaTainer', as a proxy for the African highland climate

Report Jan 2018-Dec 2018

Abstract

Future agriculture is challenged to secure food, feed, fuel and fiber production through sustainable intensification. This requires more effective yield per area also taking into account environmental constraints. Germplasm collections provide valuable genetic resources with potential to augment the resilience of agro-ecosystems towards future (unpredictable) climate conditions. An example of such a collection is the world's biggest Musa collection (ITC, Bioversity International) holding over 1500 *in vitro* accessions, hosted at the KU Leuven. The climate-controlled growth container, so called "bananatainer", introduced here, which is designed to simulate the climate of the East-African highlands and perform high-throughput evaluation of growth performance. The challenges in the highlands are the lower temperatures, a dry season of 2-3 months and lack of nutrient such as potassium and nitrogen. We established a pipeline, growing 504 banana plantlets of 21 cultivars for 6 weeks in a continuously recirculated and aerated nutrient solution in the bananaTainer. This allowed so far to rank the growth of 57 cultivars, representing 631 ITC accessions. The best performing cultivars are 3 AAB cultivars (Uzakan, Silk, Raja), 1 AAA (Lacatan), 1 local Mshare cultivar (Huti Green Bell), 1 ABB cultivar (Simili Radjah), and 1 malaccensis wild accession (Pahang).

Water scarcity was mimicked by addition of 5% Poly ethylene Glycol (PEG), lowering the plant water availability to osmotic water permeability coefficient (pF) 2.7. This system has been proven to be successful in small scale. However, when upscaling to high throughput conditions, the system was vulnerable towards micro-organisms. In the bananaTainer, the added PEG was metabolized by Cupriavidus sp., Pseudomonas putida, and Acinetobacter sp. This depleted the nutrients, rendering them unavailable for the plants. The bananaTainer is standard equipped with a UV-treatment. The UV-treatment was successful to keep micro-organism growth under control in the bananaTainer but unsuitable in combination with PEG. The addition of an alternative extra disinfection was not appropriate as this caused root damage and decreased plant transpiration. We conclude that the bananaTainer and the plant growth model is excellent to simulate the impact of lower temperatures and specific nutrient loss on plant growth. However, Polyethylene glycol is not suitable to simulate drought stress due to microbial degradation. An alternative where drought is simulated in an inert matrix is being evaluated currently.

Introduction

By 2050 the demand of food, feed, fiber, and fuel from agricultural sources is expected to double. Meeting these rising demands requires sustainable intensification of the production: more effective yield on the same area (Foley et al., 2011). However, the current yield increases of the four major crops (maize, rice, wheat, and soybean) are not sufficient to meet the required crop yield increase (~ + 2.4% yearly) (Ray, Mueller, West, & Foley, 2013). Also the further increasing impact of climate variability is estimated to impact worldwide agricultural productivity. Taking this into account, an example pathway towards sustainable intensification is creating augmented resilience in agricultural production systems in order to reduce the risks (Lipper et al., 2014). The impact of abiotic stress on agricultural productivity is more determined by environmental variability rather than the (a)biotic stress factor (Langridge & Reynolds, 2015). Elite, high yielding, cultivar breeding has traditionally aimed at increasing productivity, successfully augmenting production, but narrowing genetic diversity, and thus limiting stress plasticity (Dalal, Attia, & Moshelion, 2017; Dwivedi et al., 2016). Climate smart cultivars play a central role in this regard: agriculture should focus on the right cultivar in the appropriate environment, more than applying non-sustainable management. Plant landraces are adapted to a certain agro-ecological environment, and are a major source of genetic gain (alleles) towards environmental plasticity. It is of utmost importance that germplasm pools are taken up into climate smart plant breeding and cultivar selection protocols (Dwivedi et al., 2016; Mickelbart, Hasegawa, & Bailey-Serres, 2015). This requires that available (stored, or wild) germplasm collections are explored and characterized by full integration of genomic and plant physiological phenotyping.

Characterizing germplasm in the field risks to be masked by environmental and spatial heterogeneities combined with impracticalities, leaving the scientific question largely unsolved (Negin & Moshelion, 2017). In controlled conditions, using lab or greenhouse models, there is tight control of externalities, and high experimental reproducibility. These pre-field phenotyping experiments make a rough selection of candidate phenotypes focusing on plant performance proxies relevant for the agricultural productivity (yield). Subsequently these candidate phenotypes should be validated for their agricultural relevance in dynamic and physiological trait based field experiments (De Boeck et al., 2015; Junker et al., 2015; Moshelion & Altman, 2015; Poorter et al., 2012; Vanhove, Vermaelen, Panis, Swennen, & Carpentier, 2012).

In the first year of this project, we have set up a large scale, high-throughput characterization experiment for banana, a major crop in (sub-) tropical regions (> 144 million tons yearly production). We simulated the climate of the Eastern African highlands. This region is characterized by a dry season of 1-3 months, lower temperatures and lack of nutrients such as nitrogen and potassium (Taulya, van Asten, Leffelaar, & Giller, 2014). In the environment of the highlands, banana requires vast amounts of water (at least 1200 mm / year) and is heavily susceptible to drought: 8 % yield decline per 100 mm water not available for evapotranspiration (Hegde & Srinivas, 1989; van Asten, Fermont, & Taulya, 2011). Breeding new, drought tolerant banana cultivars that are not too much affected by the lower temperatures and can cope with low fertilizer input is a difficult and lengthy process, as such the exploration of the available biodiversity is of real interest (Christelová et al., 2016; Ortiz & Swennen, 2014). The International Transit Center (ITC) of Bioversity International, hosted at the KU Leuven, is the world's largest banana germplasm collection with over 1500 accessions. To increase management efficiency and turnover of the collection both genotypic (Christelová et al., 2016) and phenotypic characterization are crucial. Our phenotypic characterization experiment adds value to the germplasm collection towards a major stressor in the East-African Highland banana cultivation.

Simulating lower temperatures and lack of nutrients is more straightforward in soilless, hydroponics cultivations than simulating drought. Drought simulation implies the establishment of a lowered water

potential, for example by application of osmotic active compounds. Suitable compounds act on the water potential without interference or plant absorption. The most suitable one is the high molar mass polyethylene glycol (PEG, MW>6000 g/mol) (Blum, 2011; Burlyn E Michel & Kaufmann, 1973; Money, 1989; Verslues, Agarwal, Katiyar-Agarwal, Zhu, & Zhu, 2006). PEG in water exists in a very hydrated form through hydration binding on its ether oxygen atoms creating a hydration shell (Burlyn E Michel & Kaufmann, 1973; Steuter, Mozafar, & Goodin, 1981). These hydrogen bonds increase the thermodynamic disorder and lower the free energy of the solution, as such the specific PEG - water interactions cause a sharp increase of osmotic pressure. The osmotic pressure increases quadratically with increasing concentration (B. E. Michel, 1983; Burlyn E Michel & Kaufmann, 1973; Money, 1989; Steuter et al., 1981). This is determined by shape, conformation, and size (molecular weight) of the individual PEG molecules, more than by the PEG concentration (molarity) (Burlyn E Michel & Kaufmann, 1973; Steuter et al., 1981). PEG solutions of higher molecular weight have a higher osmotic pressure at the same concentration (Burlyn E Michel & Kaufmann, 1973; Money, 1989). Plant cell walls are considered impermeable for larger PEG molecules, creating an osmotic potential at the cell wall boundary (cytorhizis) (Verslues, Ober, & Sharp, 1998). This contrasts to other, smaller compounds, like mannitol or sucrose, and salts, like NaCl, which readily penetrate the cell wall creating osmotic pressure at the cell membrane (plasmolysis) and easily enter de cell.

Worldwide, millions of tonnes of these large, water soluble, non-ionic, PEG polymers with a C-C backbone, are produced (Huang et al., 2005). PEG is widely used in for example cosmetics, lubricants, plastics, as anti-freezing agent, as such PEG is ubiquitous in (waste) water. The polymers are highly resistant to normal biodegradation, resulting in their longevity in natural environments (Allen et al., 2002; Huang et al., 2005). In highly specialized conditions such as waste water treatment installations, both aerobic and anaerobic pathways of bacterial PEG-degradation have been found (Bernhard, Eubeler, Zok, & Knepper, 2008; Dwyer & Tiedje, 1986; Kawai, 2002; Marchal, Nicolau, Ballaguet, & Bertoncini, 2008; Takeuchi, Kawai, Shimada, & Yokota, 1993).

Material and methods

Plant growth conditions and osmotic stress treatment in the bananaTainer

In vitro banana plantlets are obtained through the International Transit Center (Bioversity International, Heverlee, BE). The bananaTainer, the plants are grown in, is a climate controlled reefer container, engineered by Urban Crops Solutions (Waregem, BE). Plants grow under 12/12 (day/night) LED powered light regime, temperature and relative humidity are set to 24/13 °C (day/night) and 70% respectively. The CO2 concentration in the bananaTainer is kept at 350-370 ppm. In this hydroponic growth model plants are grown per six in a shallow tray holding 4L of nutrient solution. Nutrient solution is provided by an overflow system, refreshing the tray volume every hour. The total circuit volume is kept constant. The nutrient solution contains: 361 mg/L KNO_3, 121 mg/L K_2SO_4, 176 mg/L $MgSO_4.7H_2O$, 181 mg/L $MgCl_2.6H_2O$, 194 mg/L KH_2PO_4, 398 mg/L $NaH_2PO_4.2H_2O$, 464 mg/L $Ca(NO_3)_2.4H_2O$, 105 mg/L $CaCl_2.2H_2O$, 0.1125 mL/L EDDHSA-Fe, 1.1 mg/L H_3BO_3, 2.7 mg/L $MnSO_4.H_2O$, 0.23 mg/L $ZnSO_4.7H_2O$, 0.16 mg/L $CuSO_4.5H_2O$, 0.07 mg/L $NaMoO_4.2H_2O$, pH = 6, modified from Swennen et al (1986). Analysis reports of the nutrient solution composition are provided by the Belgian Soil Science Service (Bodemkundige Dienst, Heverlee, Leuven). The pH and EC of the nutrient solution are monitored and logged on 1 min interval.

Osmotic stress is applied by adding 50 gram of Polyethylene glycol (PEG-8000) (Carl-Roth) per liter nutrient solution (5%). The osmolality is measured using a Freezing point osmometer (Osmomat-3000, Gonotec GmbH, DE) using 50 µL sample. The osmolality is output in mOsmol/kg, and is converted into the osmotic pressure (MPa) by the Van't Hoff equation (Phi = RTc).

Morphological variables extraction

At the start of the experiment all plants are phenotyped non-destructively (whole plant weight, pseudostem height, and topview leaf area), and the last formed leaf is marked. This phenotyping event at the start of the experiment allows to select the most homogenous subgroup to use in the actual experiment. Based on the total mass, the 24 most homogenous plants per cultivar are selected, excluding those which are too big, too small, or damaged.

After 6 weeks of growth in the bananatainer, all plants are phenotyped again using combined (destructive) measurements and digital imagery. Firstly, the plants are weighed and the pseudostem height measured. Secondly the canopy area is calculated from top view images. Green plant pixels are separated from the blue background by color segmentation. Using a red reference surface of known size (10 x 5 cm) the green area is calculated. Thirdly, the leaves are separated and spread out by increasing age and subsequently imaged. Based on these images the number of leaves, and individual leaf length, width, and area are determined. Lastly, the weight of separated plant parts (root, pseudostem, and leaves) is recorded before and after 14 days of drying at 70°C.

All image analysis is performed using an in house R tool based on the EBImage (Pau, Fuchs, Sklyar, Boutros, & Huber, 2010), and imagemagick packages. The final list of morphological variables (measured and calculated) is taken up in the banana crop ontology database (Banana: CO_325, available at http://www.cropontology.org/terms/CO_325/, accessed Oct. 23, 2018).

Visualization and identification of the microbiota

For visualization and identification purposes the microorganisms are cultured on plates using infected sample from the nutrient solution in the bananaTainer. These plates are aerobically incubated at 20°C. The microbial growth in the nutrient solution is estimated by loading approximately 1.5 mL in a spectrophotometer (Pharmacia BioTech, SE) at λ 600 nm, comparing absorbance to purified water. The standard curve was made by adding increasing volumes of microbiota loaded nutrient solution to 35 mL, nutrient solution (composition same as above) containing 5% PEG. The spectrophotometric measurement were performed after 1h incubation at 24°C.

Identification of the microorganisms was performed at the Belgian Coordinated Collections of Microorganisms at Ghent University. Samples were taken when the infection in the bananaTainer reached its maximum. A serial dilution (10^0 till 10^-8) on 3 different media was established. 1) nutrient solution containing 5 % PEG, solidified by Gelrite (Duchefa Biochemie, NL); 2) trypticase soy agar (TSA) with addition of 5 ppm amphotericin B and 200 ppm cycloheximide and; 3) DYPA with addition of 100 ppm chloramphenicol. After aerobic incubation at 20 °C single colonies were picked and spectra generated using MALDI-TOF MS. Spectral clustering reduces replication and identification is done against an in house LMG reference database. Reliable identifications have log score > 1.7.

Plant growth experiment characterizing chlorine dioxide applicability

The effects of daily ClO_2 addition is assessed in a hydroponics-based growth screening trial. Three plants were grown together in a 3 L tray with the roots freely in hydroponic solution (nutrient content same as reported above). Plants were grown in a growth chamber (12/12 light/dark) at 25 °C and 75 % RH. On a week/daily basis the trays were weighed and refilled up to 3 liters, as such the daily transpiration was determined. Subsequently, non-destructive root phenotyping was done by imaging 1 plant per tray using a side view camera (Canon EOS1300D), making sure a red reference surface (50 cm^2) was pictured simultaneously. Finally, chlorine dioxide was prepared according to manufacturer's instructions (Aqua Ecologic, BE) and supplied at 0, 0.5, 1, 3, or 10 ppm. In total 3 plants were grown per tray, with 3 trays per treatment in 5 treatments (45 plants in total).

Using R (3.4.3, EBImage, and magick packages) root images of 1 plant per tray were segmented from the background using color segmentation. The root area, as seen from the side, is calculated by counting the number of root pixels and comparing this to the number of pixels of the reference surface. A proxy for root damage was calculated by comparing the number of pixels with Hue value between 0 - 15 (from 0-255 HSV color space) to the total number of root pixels.

Results

Introducing the bananaTainer, a container based, tailor made, banana growth chamber to select climate smart varieties

The major aim of the project is to simulate the stresses prone to the East-African highlands and phenotype the Musa biodiversity for its suitability to grow and produce there. This implies many repeated tests have to be performed in a maximally controlled and fully repeatable environment. For this we introduce the bananaTainer, a container based, tailor made banana growth chamber engineered by Urban Crops Solutions (https://urbancropsolutions.com/, Waregem, BE). The specific, 3-layer, vertical farming design allows to grow 504 randomly placed banana plantlets under tightly controlled conditions in hydroponics (Figure 1). Two separate circuits (each total volume 500 L) allow 2 different nutrient media (simulating drought and/or nutrient deficiency) to set up the experimental design. Frequent irrigation events fully refresh and aerate the nutrient solution every hour. There is complete drain reuse and the total water volume per circuit is monitored and kept constant. Nutrient solution filtering is crucial and ensured by a 80 micron particle filter per circuit. Environmental settings (temperature, relative humidity, CO_2 concentration, and day length) are controlled and can be set according to the experimental design. The LED lights provide a specific spectrum relevant for banana (van Wesemael et al. 2018 (ISHS, in revision)). Every run, two well studied reference cultivars (Cachaco (ITC0643) and Mbwazirume (ITC0084)) are taken along to estimate the inter run variability. In total this allows characterization of 19 new cultivars every run under two conditions using 12 biological replicates (504 = 12 * 2 * (19+2)).

Figure 1: The 'bananaTainer' is a modified, climate-controlled reefer container specifically designed to house 504 banana plants. The water circuit contains 2 loops, the main loop supplying the plants with water by an overflow system, passing a particle filter (80 micron) before returning to the nutrient tank. The secondary loop treats the water with a UV source and leads the water over an EC and pH sensor.

The pipeline encompassing plant nursery growth, pre-test phenotyping, bananatainer growth, post-test phenotyping followed by feature extraction and data analysis has successfully been established. So far, we tested 57 different cultivars, of 24 different subgroups in the bananatainer pipeline. In total this represents 631 different cultivars of the ITC Musa germplasm collection (Table 1). The aim is to evaluate the suitability of cultivar growth in the East African highland agro-ecological zone. By expressing the daily growth of the plants relative to the reference cultivar taken along every run

(Cachaco, ITC0643, Bluggoe, ABB) we can rank the cultivar performance (Figure 2). Roughly 50% of the cultivars grow better than the reference cultivar. Cultivars with significant better growth are Uzakan, Silk, Pahang, Lacatan, Raja, Huti Green Bell, and Simili Radjah (p-value<0.05, Tukey HSD test). Ingarama (ITC0160, Mutika/Lugujira, AAA) performs significantly worse than the reference cultivar.

Table 1: 57 used cultivars, representing 24 different subgroups, and a total of 631 different ITC cultivars.

SubSpecies/SubGroup	Genome	# accessions in ITC	Cultivar	ITC code	Exp
banksii	AA	30	Banksii	ITC0623	III
zebrina	AA	6	Zebrina	ITC1177	II
malaccensis	AA	15	Pahang	ITC0609	I
Mshare	AA	14	Huti green bell	ITC1559	I
			Makyughu II	ITC1446	III
Sucrier	AA	9	Pisang Mas	ITC0653	I
Unknown	AA	/	Guyod	ITC0299	I
Unknown	AA	/	Mjenga Gros Michel Diploid	ITC1253	III
Cavendish	AAA	51	Grande Naine	ITC0180	II
			Williams	ITC0365	II
			Lacatan	ITC0768	III
			Poyo	ITC0345	III
Gros Michel	AAA	9	Gros Michel	ITC1122	II
			Highgate	ITC0263	III
Ibota	AAA	7	Khai Thong Ruang	ITC0662	II
Mutika/lugujira	AAA	76	Mbwazirume	ITC1356	I,II,III
			Igisahira Gisanzwe	ITC0083	I
			Ingagara	ITC0166	I
			Ingumba y'Imbihire (Inyamunyu)	ITC0155	I
			Inyoya	ITC0163	I
			Inzirabahima	ITC0150	I
			Mpologoma	Noitc	I
			Nakitengwa	ITC1180	I
			Namunwe	ITC1629	I
			Nyitabunyonyi	ITC1556	I
			Igitsiri (Intuntu)	ITC0081	II
			Guineo	ITC0005	III
			Ingarama	ITC0160	III
			Ingumba y'Inyamunyo	ITC0126	III
			Kibiddebidde	ITC1624	III
			Kisukari usiniguse	ITC1546	III
			Makara	ITC0177	III
			NSH 42	ITC1802	III
Red	AAA	10	Red Dacca	ITC0575	II
Rio	AAA	3	Leite	ITC0277	II
Unknown	AAA	/	Pisang Berangan	ITC1287	II
Unknown	AAAA	/	Fhia-23	ITC1265	I
Iholena	AAB	2	Uzakan	ITC0825	II
Mysore	AAB	10	Pisang Ceylan	ITC1441	II
Pisang Kelat	AAB	5	Pisang Palembang	ITC0450	II
Pisang Raja	AAB	4	Pisang Rajah	ITC0587	II
			Pisang Raja Bulu (Raja)	ITC0843	III
Plantain	AAB	292	Orishele	ITC1325	II
Pome	AAB	25	Foconah	ITC0649	I
			Prata Ana	ITC0962	II
Silk	AAB	14	Kipakapaka	ITC1550	I
			Yangambi n°2	ITC1275	II
			Figue Pomme Géante (Silk)	ITC0769	III
Bluggoe	ABB	16	Cachaco	ITC0643	I,II,III
			Dole	ITC0767	II
			Kivuvu	ITC0157	II
Monthan	ABB	10	Monthan	ITC1483	I
Pelipita	ABB	4	Pelipita	ITC0472	III
Peyan	ABB	2	Simili Radjah	ITC0123	III
Pisang Awak	ABB	17	Kayinja	ITC0087	I
			Namwa Khom	ITC0659	I
			Fougamou	ITC0101	III
Pisang Klutuk Wulung	BB	/	Pisang Klutuk Wulung	ITC1587	III

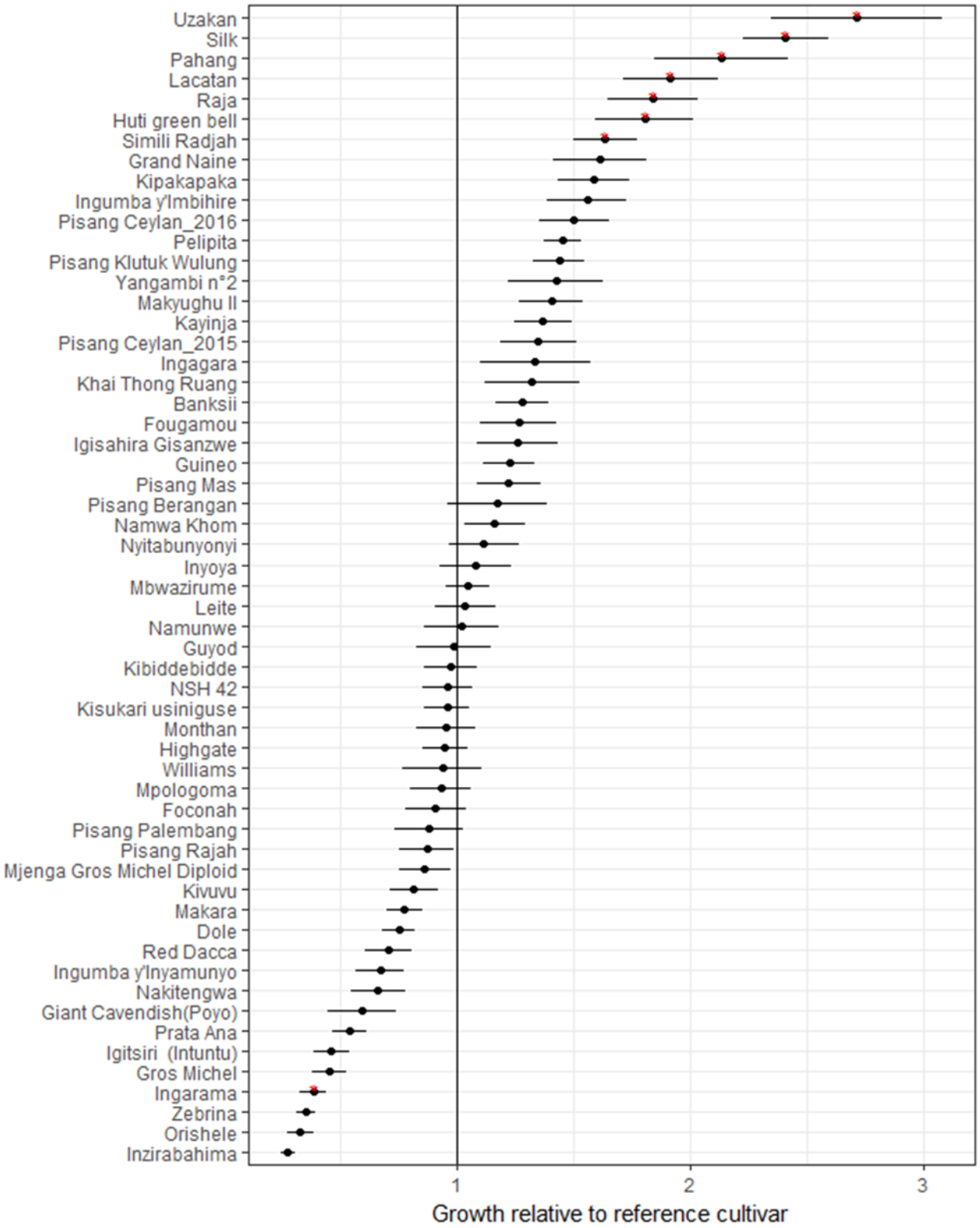


Figure 2: Impact of low temperature on the growth of 57 cultivars evaluated relative to the reference cultivar (Cachaco, Bluggoe, ABB). Mean relative growth with standard error is represented. Significant differences are indicated (p-value<0.05).

The primary output of the bananaTainer growth setup is to obtain a relative ranking of cultivar suitability for the East African highlands (cf. Figure 2). However, we generate additional phenotypic features, complementing this growth ranking in order to get insight in the crucial features differentiating the well suitable cultivars. This phenotyping is for example done via imaging. Per run

over 870 images are taken and processed. An example of the image analysis pathway, for loose leaves is displayed in Figure 3. These images of loose leaves allow to phenotype the evolution of leaf growth throughout the experiment.

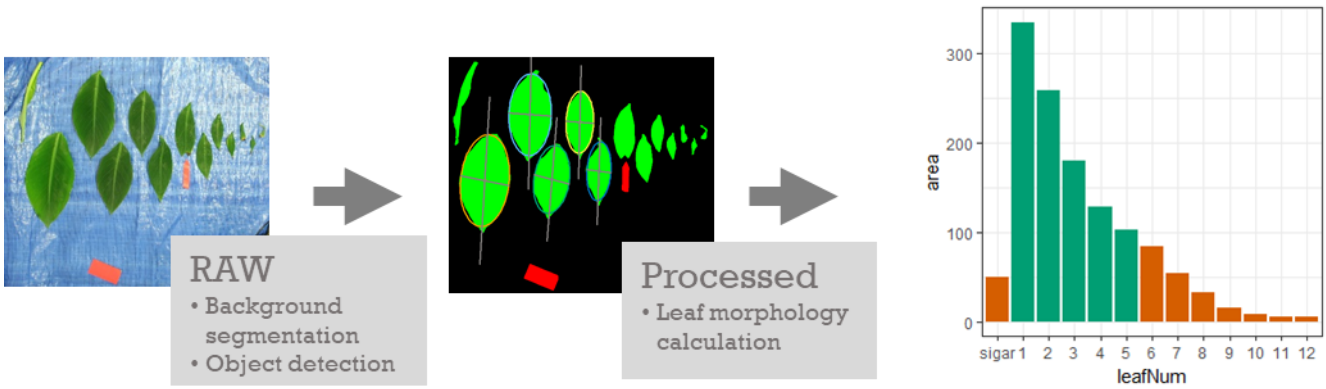


Figure 3: Image analysis pipeline describing evolution of leaf area through the experiment.

A characterization of the microbial biomass growing on PEG

In the PEG circuit runs were infected by micro-organisms. Upon microbial infection, the aerated nutrient solution is characterized by an increase in spectrophotometric turbidity (λ 600 nm). The turbidity is linearly correlated to microbial biomass (Figure 4). The pH of the nutrient solution also increases from pH 6 to pH 7 in 7 days, while the pH of the control nutrient solution does not change (Figure 5A). Nutrient composition analysis shows that the pH change corresponds with a serious shift in N and HCO_3 (Figure 5B). The micro-organisms consume N and produce CO_2 that converts to HCO_3

Via MALDI-TOF MS spectra of isolated colonies three colonies which could grow solely on the PEG substrate, are confidently identified: Cupriavidus sp., possibly C. basiliensis (log score = 2.000), Pseudomonas putida species group (log score = 2.470), Acinetobacter sp., probably A. johnsonii (log score 2.550).

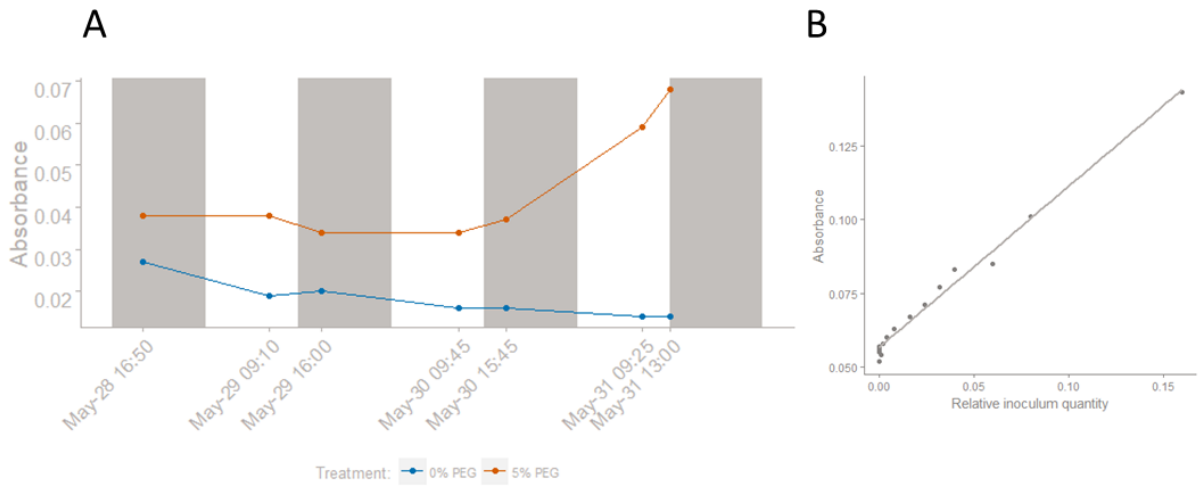


Figure 4: (A) The microbial presence manifests as an increase in spectrophotometric turbidity (OD lambda 600 nm). (B) The turbidity increases linear with the inoculum quantity.

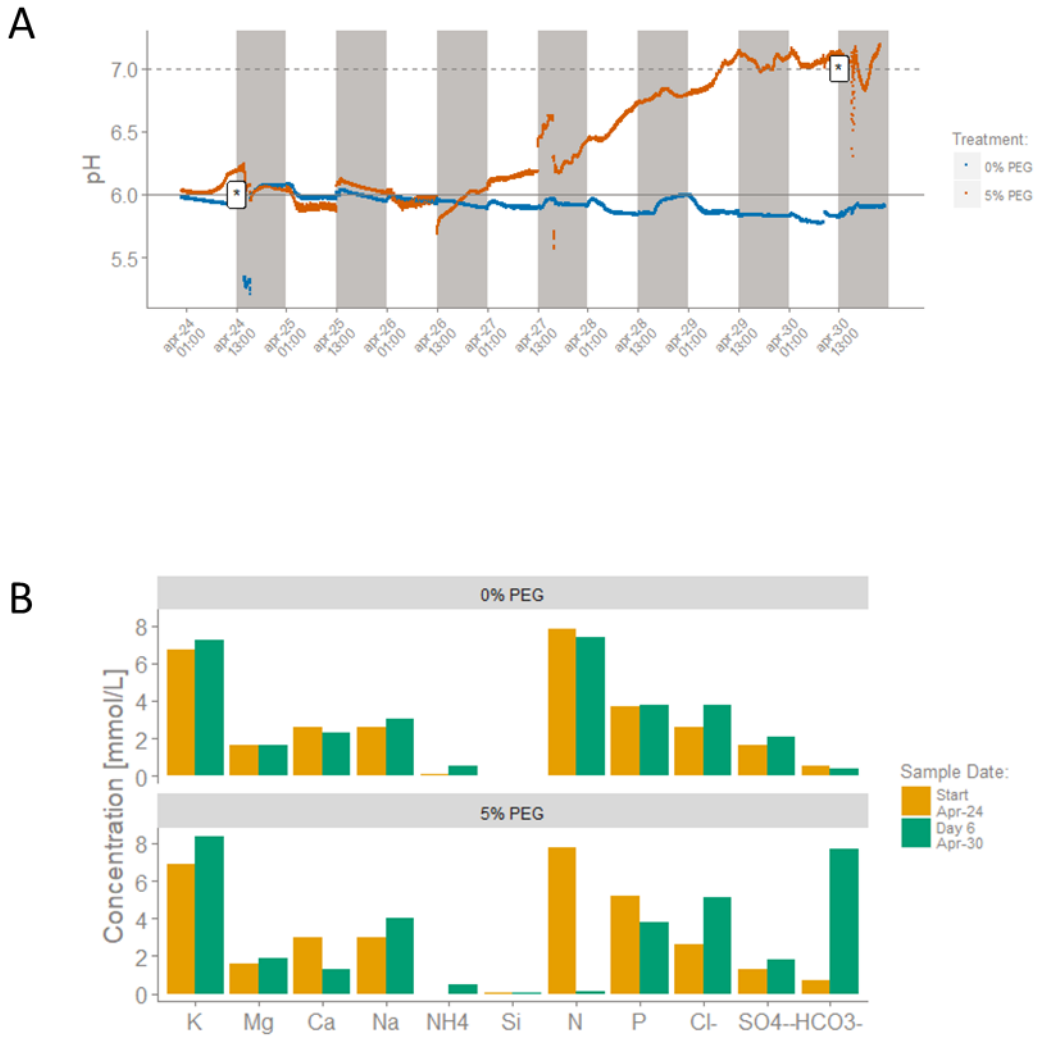


*Figure 5: (A) The microbial infection is manifested by a pH rise in the nutrient medium with addition of 5% PEG. pH measured in both tanks continuously (1 minute interval). Sampling times for nutrient solution analysis indicated by *. Nighttime periods indicated by shading. (B) The microbial infection depletes nutrient solution in the 5% PEG nutrient solution.*

Effects of daily chlorine dioxide application on plant performance traits.

The initial healthy root system shows severe browning under daily addition ClO_2. Especially the root hairs are vulnerable to the non-selective oxidation of chlorine dioxide (Figure 6). Daily application of chlorine dioxide lowers transpiration rate at 0.5, 1, and 3 ppm, but especially at 10 ppm (Figure 7A). The increase in root area is lowered at 3 and 10 ppm applied ClO_2, as visualized by the side view root area over time in Figure 7B. The daily addition of chlorine dioxide causes root browning (Figure 7C) by the relative pixel count of pixels in the red space of HSV color space (H < 15).

Figure 6: The daily addition of chlorine dioxide causes root browning, impeding root growth. The root hair structures are most vulnerable.

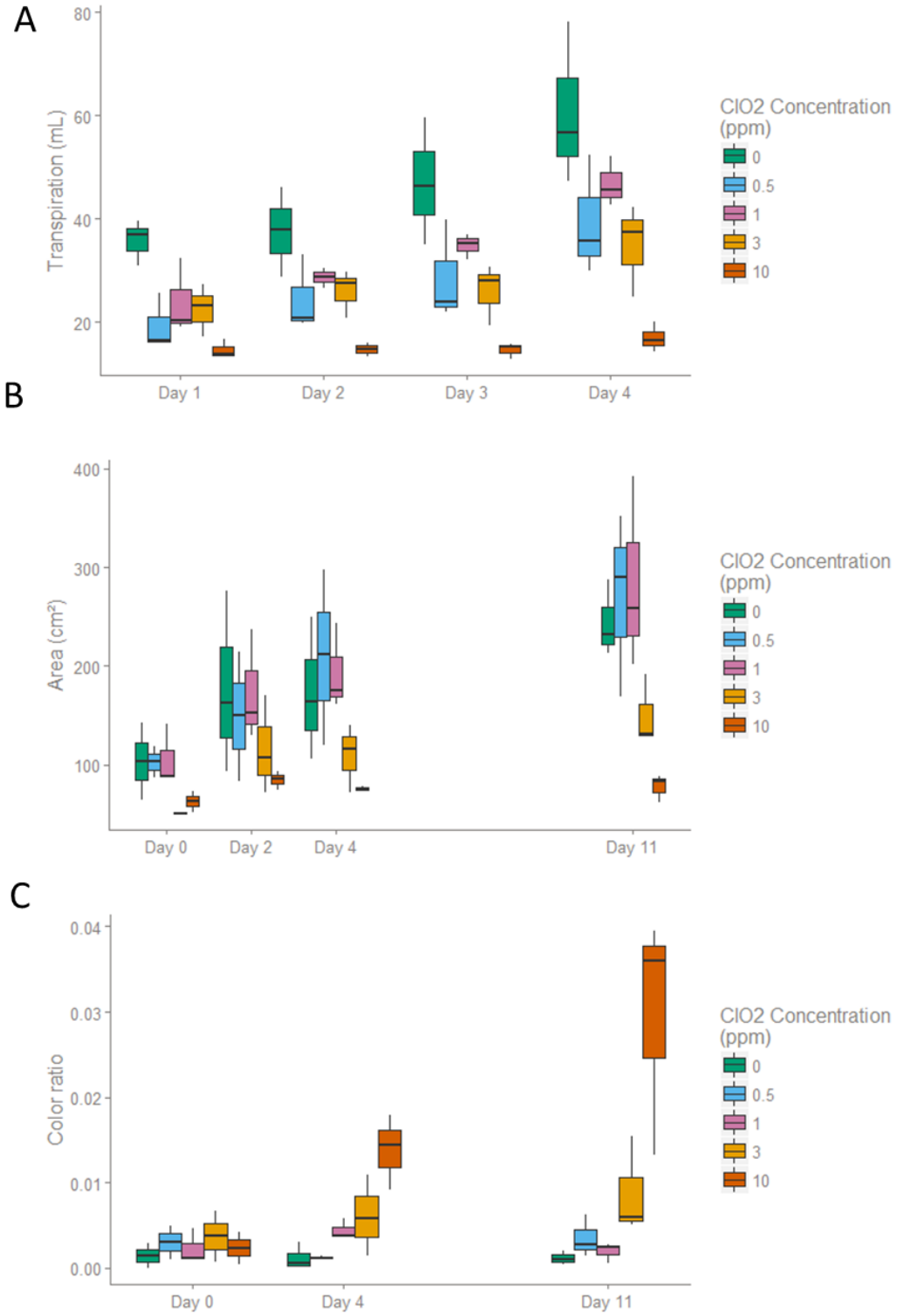


Figure 7: Daily chlorine dioxide applications have impact on plant performance: A) daily transpiration rate, B) root area, C) root brownness. Transpiration is determined per tray, there are three plants grown together in 1 tray with 3 replicate trays per treatment. The root area is calculated for 1 plant per tray. The root brownness variable is also calculated for 1 plant per tray. It is the proportion of root pixels with H<15 (in 0-255 HSV color space).

Discussion

In this work we introduce the bananaTainer, a relatively cheap, large scale banana characterization infrastructure designed for a high throughput growth characterization (Figure 1). In our lab model, vegetative growth is a proxy for the field performance. The aim is to screen for the suitability to grow in the African Highlands. The evaluation in high throughput pre-screening as introduced here is the first high throughput attempt to add phenotypic value to the Musa germplasm collection of ITC. For further studies it allows end users (scientists, farmers, extension workers) to focus on the most promising subset of cultivars. For this pre-screening inspired aim, the bananaTainer has a suitable design (Figure 1), allowing 12 biological replicates, with 2 reference cultivars overcoming the inter-run variability. Water is continuously recycled in the overflow based system and the total running volume is kept constant. Water recycling however, also brings in phytosanitary issues. Only in the PEG stress treatment we observed an increase in spectrophotometric turbidity (Figure 4) and a drastic pH increase (Figure 5A). This coincided with a depletion of nutrients (Figure 5B) and was correlated to microorganism presence. No carbon is added to the nutrient solution, apart from PEG. We validated the presence of 3 PEG metabolizing bacteria in the nutrient solution (Cupriavidus basiliensis, Pseudomonas putida, and Acinetobacter johnsonii). (Marchal et al., 2008) studied specialized microflora from waste water treatment installations and from soils, and found Cupriavidus basiliensis and Pseudomonas putida are able to metabolize PEG 400. In our case a larger PEG-8000 molecule is consumed. Plants normally grow in environments full of microorganisms: the soil. As such non-pathogenic microorganisms are not an issue. However, the bacterial nutrient consumption in our environment reduces the interpretability and experimental reproducibility because it depends on the microbial activity. Colonization speed, growth rate, … can be different from run to run. Prevention of microorganism infection in combination with PEG proves rather difficult, as the addition of chlorine dioxide, a non-selective oxidator of organic material (Scarlett et al., 2016) damages the roots and affects the transpiration (Figure 6, Figure 7). The chlorine dioxide oxidizes in a non-selective fashion, targeting bacteria, roots, iron chelates, … The root hairs, the most vulnerable structures of the root system, are damaged first, attributing to the impeded transpiration. In the most severely damaged root systems there is a 10 fold increase in brownness, compared to the fresh start

The main project objective was to establish a high throughput, pre-field phenotyping infrastructure based on vegetative growth. Models using PEG to apply very precise and repeatable low water potentials in soilless, hydroponic cultivations seem relevant in this regard (Poorter et al., 2012). However, using PEG as a simulation of drought is not possible. Alternative characterization of the drought response is to observe the impact of cessation of water, relative to *ad libidum* water supply. These experiments are highly relevant but more challenging to control.

Update Report Dec 2018-July 2019

Results and discussion

Establishment of relevant temperature scenarios in the bananatainer

In this case study we assess the temperature sensitivity of the different banana cultivars and how it can contribute to climate smart cultivar selection. The growth performance of 35 cultivars has been screened taking into account the spatial heterogeneity in the average daily temperature in the African Highlands (Figure 8). We have extracted 61 phenotypic variables using the bananaTainer pipeline, 19 of which are linked to universal crop ontology terms (CO_325) (Table 2).

Table 2: Variables monitored during the bananatainer phenotyping. Variables are classified per measurement class they attribute to (area, damage, growth, height, mass, mass distribution, water content). The method (image, measured or calculated) as well as the data source and pipeline script used to derive are also indicated.

Variable	Description	CO term	Unit	Class	Method	dataSource	script
circumferenceCanopy	circumference of the canopy in topview		cm	Area	img	topViewImg	topViewImg2data
damageAllLeavesArea	damage (browning) in all formed leaves total damaged area (H<45 in HSV)		cm²	Damage	img	looseLeavesImg	looseLeaves2data
damageAllLeavesRatio	damage (browning) in all formed leaves in ratio (H<45 in HSV)			Damage	img	looseLeavesImg	looseLeaves2data
damageOldLeavesRatio	damage (browning) in all old leaves in ratio (H<45 in HSV)			Damage	img	looseLeavesImg	looseLeaves2data
damageTotLeavesArea	damage (browning) in all leaves total damaged area (H<45 in HSV)		cm²	Damage	img	looseLeavesImg	looseLeaves2data
damageTotLeavesRatio	damage (browning) in all leaves in ratio (H<45 in HSV)			Damage	img	looseLeavesImg	looseLeaves2data
dAreaRelYoungest	relative area difference between youngest 2 leaves			Area	img	looseLeavesImg	looseLeaves2data
ddAreaRelYoungest	change of the area difference			Area	img	looseLeavesImg	
dLengthRelYoungest	relative leaf length difference between youngest 2 leaves			Area	img	looseLeavesImg	looseLeaves2data
dWidthRelYoungest	relative leaf width difference between youngest 2 leaves			Area	img	looseLeavesImg	looseLeaves2data
growthCanopy	increase in canopy area (topview)		cm²	Area	meas	topViewImg	
growthPlantDry	plant growth during the experiment in dry weight	CO_325:0000914	g	Growth	calc	massFiles	
growthPlantFresh	plant growth during the experiment in fresh weight	CO_325:0000915	g	Growth	calc	massFiles	
growthPS	growth of the pseudostem during the experiment		cm	Growth	calc	massFiles	
growthPS_perLeaf	growth of the pseudostem during the experiment per leaf formed	CO_325:0000918	cm/#	Growth	calc	loosLeavesImg + massFiles	
growthRootFresh	root growth during the experiment		g	Growth	calc	massFiles	
growthShootFresh	shoot growth during the experiment		g	Growth	calc	massFiles	
heightPS	height of the pseudostem		cm	Height	meas	massFiles	massFiles2data
heightPSStart	height of the pseudostem at the the start		cm	Height	meas	massFiles	massFiles2data
leafAreaAll	total leaf area formed during the experiment	CO_325:0000883	cm²	Area	img	looseLeavesImg	looseLeaves.R
leafAreaCanopy	area of the whole canopy in topview	CO_325:0000882	cm²	Area	img	topViewImg	topViewImg2data.R
leafAreaCanopyStart	area of the whole canopy in topview at the exp start		cm²	Area	img	topViewImg	
leafAreaDeclineAllLeavesBin	binary variable do we have a leaf area decline in sequence of formed leaves?		0/1	Damage	img	looseLeavesImg	looseLeaves2data
leafAreaSpecific	leaf area per gram leaf weight dry (only when no sigar)		cm²/g	Area	calc	looseLeavesImg + massfiles	
leafAreaYoungest	Area of the youngest leaf	CO_325:0000884	cm²	Area	img	looseLeavesImg	looseLeaves.R
leafLengthAll	summed leaf length of leaves formed during the experiment	CO_325:0000886	cm	Area	img	looseLeavesImg	
leafLengthYoungest	length of the youngest leaf	CO_325:0000887	cm	Area	img	looseLeavesImg	
leafWidthAll	summed leaf width of leaves formed during the experiment		cm	Area	img	looseLeavesImg	
leafWidthYoungest	width of the youngest leaf	CO_325:0000888	cm	Area	img	looseLeavesImg	
malFormAllLeavesBin	binary variable do we have malfunctioned leaf (L/W >4)		0/1	Damage	img	looseLeavesImg	looseLeaves2data
massLeafDry	leaf dry weight		g	Mass	meas	massFiles	massFiles2data
massLeafFresh	leaf fresh weight		g	Mass	meas	massFiles	massFiles2data
massPlantDry	plant dry weight		g	Mass	meas	massFiles	massFiles2data
massPlantFresh	plant fresh weight		g	Mass	meas	massFiles	massFiles2data
massPlantFreshStart	plant fresh weight at the start		g	Mass	meas	massFiles	massFiles2data
massPSDry	pseudostem dry weight		g	Mass	meas	massFiles	massFiles2data
massPSFresh	pseudostem fresh weight		g	Mass	meas	massFiles	massFiles2data
massRootDry	root dry weight		g	Mass	meas	massFiles	massFiles2data
massRootFresh	root fresh weight		g	Mass	meas	massFiles	massFiles2data
massShootDry	shoot dry weight		g	Mass	meas	massFiles	massFiles2data
massShootFresh	shoot fresh weight		g	Mass	meas	massFiles	massFiles2data
maxDamageAllLeavesArea	maximum damage (browning) in all formed leaves total damaged area (H<45 in HSV)		cm²	Damage	img	looseLeavesImg	looseLeaves2data
maxDamageAllLeavesRatio	maximum damage (browning) in all formed leaves in ratio (H<45 in HSV)			Damage	img	looseLeavesImg	looseLeaves2data
numberLeaves	number of leaves formed during the experiment	CO_325:0000900	#	Growth	img	looseLeavesImg	looseLeaves2data
numberLeaves_perGrowthPS	number of leaves formed during the experiment per pseudostem growth	CO_325:0000919	#/cm	Growth	calc	looseLeavesImg + massfiles	
ratioLeafPlantDry	leaf dry weight relative to the whole plant dry weight	CO_325:0000909		Mass d.	calc	massFiles	
ratioLeafPlantFresh	leaf fresh weight relative to the whole plant fresh weight			Mass d.	calc	massFiles	
ratioLengthWidthLeafPreExp	ratio of length to width of the youngest leaf pre experiment			Area	img	looseLeavesImg	looseLeaves2data
ratioLengthWidthYoungest	ratio of length to width of the youngest leaf			Area	img	looseLeavesImg	looseLeaves2data
ratioPSPlantDry	pseudostem dry weight relative to the whole plant dry weight	CO_325:0000910		Mass d.	calc	massFiles	
ratioPSPlantFresh	pseudostem fresh weight relative to the whole plant fresh weight			Mass d.	calc	massFiles	
ratioRootLeafDry	root dry weight relative to the leaf dry weight			Mass d.	calc	massFiles	
ratioRootLeafFresh	root fresh weight relative to the leaf fresh weight			Mass d.	calc	massFiles	
ratioRootPlantDry	root dry weight relative to the whole plant dry weight	CO_325:0000911		Mass d.	calc	massFiles	
ratioRootPlantFresh	root fresh weight relative to the whole plant fresh weight			Mass d.	calc	massFiles	
ratioRootShootDry	root dry weight relative to the leaf + pseudostem dry weight	CO_325:0000894		Mass d.	calc	massFiles	
ratioRootShootFresh	root fresh weight relative to the leaf + pseudostem fresh weight			Mass d.	calc	massFiles	
rwcLeaf	relative water content of the leaves	CO_325:0000895	%	Water c.	calc	massFiles	massFiles2data
rwcPlant	relative water content of the plant	CO_325:0000898	%	Water c.	calc	massFiles	massFiles2data
rwcPS	relative water content of the pseudostem	CO_325:0000896	%	Water c.	calc	massFiles	massFiles2data
rwcRoot	relative water content of the root	CO_325:0000897	%	Water c.	calc	massFiles	massFiles2data

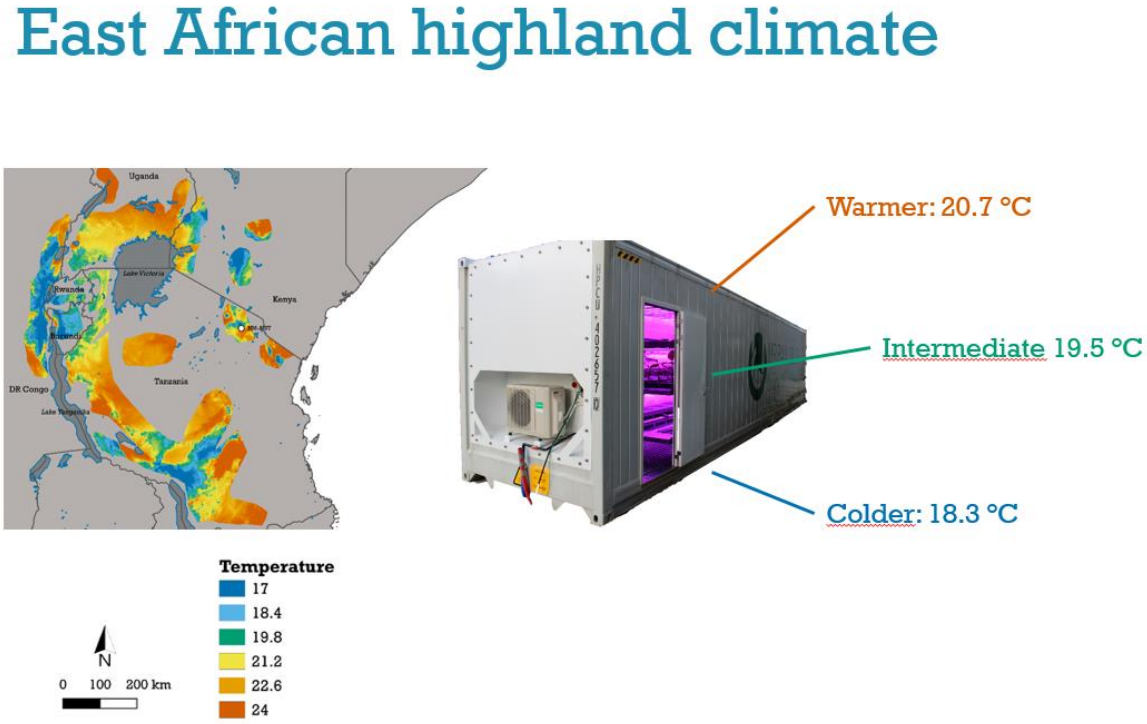


Figure 8: Map of the East-African Highlands banana growth regions displaying the average rain season temperature (Karamura 1998, Fick and Hijmans, 2017). The bananatainer simulates 3 average day temperatures 20.7, 19.5 and 18.3.

In the bananatainer we established three relevant temperature regimes (24h average temperature: 20.7 °C, 19.5 °C, and 18.3 °C (Figure 8)). These regimes represent a span of the existing temperatures in the East African highlands (based on 30 year climate records (Fick and Hijmans, 2017)). As such we assume that our established warm temperature regime proxies for the East-African highlands. More Western production areas have lower average temperatures, these correspond to the lower temperature regimes in the bananaTainer.

The bananaTainer experiments are a relative growth screening and rely on a reproducible experimental setup. The growth of the reference cultivars Cachaco (ITC0643) and Mbwazirume (ITC0084) is compared to take possible external effects into account. Three way Analysis of Variance (ANOVA) of the growth over two experiments, three temperature treatments, and two reference cultivars, show that there are significant cultivar and treatment effects, but there is no significant experimental effect as the repeatability of the experimental setup is ensured over different runs (Table 3).

Table 3 Three-way Analysis of Variance (ANOVA) shows significant treatment and cultivar, but no experimental effect. Variance of growth per day in two replicated experiments is studied for both reference cultivars, Cachaco and Mbwazirume.

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)	
cultivar	1	1.813	1.813	8.227	0.00488	**
Trt	2	8.681	4.341	19.697	4.00E-08	***
exp	1	0.159	0.159	0.723	0.39675	
cultivar:Trt	2	0.776	0.388	1.761	0.1763	
cultivar:exp	1	0.248	0.248	1.128	0.29042	
Trt:exp	2	0.191	0.096	0.434	0.64885	
cultivar:Trt:exp	2	1.217	0.609	2.762	0.06719	.
Residuals	120	26.444	0.22			

Ranking the growth potential at different temperatures

The main bananaTainer output is the objective ranking of the different accession based on growth. This enables a selection of the best performing cultivars per agro-ecological zone. Our aim was to rank the different cultivars for their performance in the major rain season, the most relevant season determining production. In general, average growth is reduced under lower temperature for all cultivars (Figure 9). Hence the lowest temperature regime is below the optimum plant growth levels. For Williams (Cavendish, AAA) this optimum was found at 21.5 (25°C/18°C (day/night average temperature)(Turner and Lahav, 1983)). The cultivars Silk (ITC0769), Raja (ITC0843) and Lacatan (ITC0768) are among the best performing cultivars under all temperature conditions. Cultivars such as Nakitengwa (ITC1180) and Ingarama (ITC0160) perform 4 times less good in all conditions. The growth does not always decline linearly with the temperature: some cultivars show their main growth penalty between 20.7°C and 19.5 °C and some between 19.5°C and 18.3 °C. In other words, there is not one single temperature threshold for all cultivars. This is illustrated in Figure 10 with the pseudostem growth. Mpologoma (no ITC) and Makyughu II (ITC1446) are significantly impacted at the intermediate temperature, while Ingagara (ITC0166) and Foconah (ITC0649) have a similar performance in 20.7°C and 19.5°C regime, and have a significant decline at the 18.3°C regime. Commonly a single base temperature (13 °C) is proposed below which banana growth is fully arrested (Ganry, 1973). This base temperature was established for Cavendish cultivars and is typically used to calculate the heat accumulation throughout growth. Calculated heat units are a commonly used time indication in modelling (Tixier et al., 2004). Our results (Figure 9) show that a common base temperature for all banana cultivars is unlikely. There is diversity within Musa spp. which has consequences for cultivar suitability in the East African highlands and pleas for per cultivar determination of this base temperature prior to modelling studies.

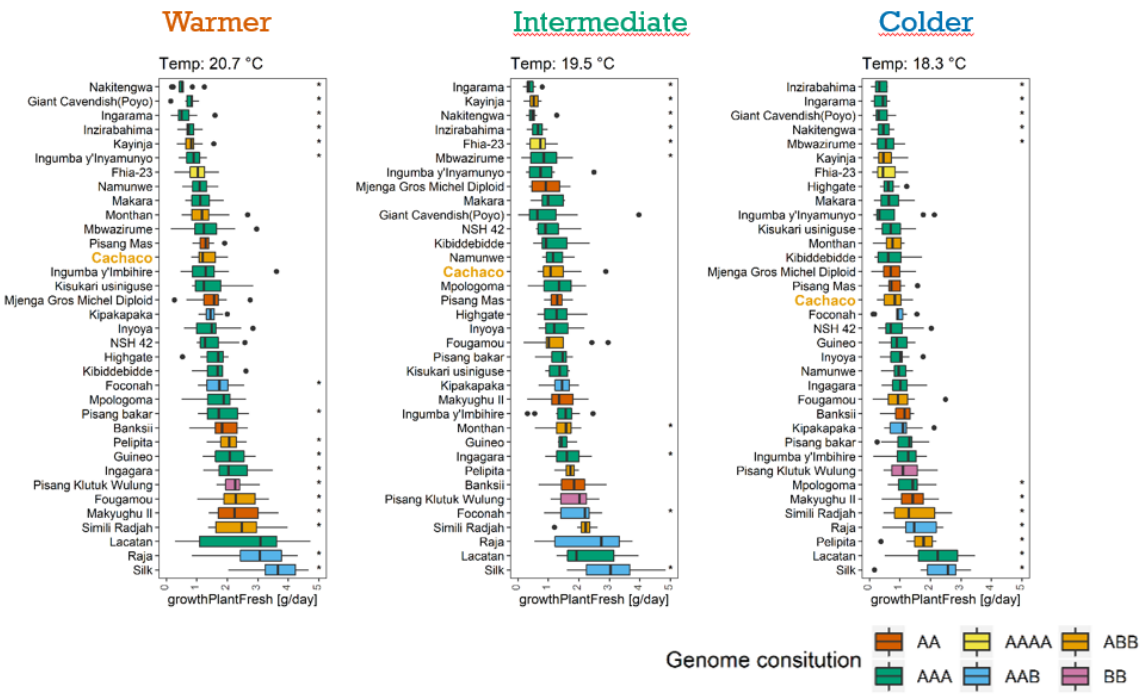


Figure 9: Major output of the bananaTainer system is an agro-ecosystem specific growth rate per cultivar. For three climate conditions the daily growth is ranked. Statistical difference with the experiment specific reference Cachaco (R) growth is indicated (*: $p < 0.05$). Cultivars are coloured per genome constitution.

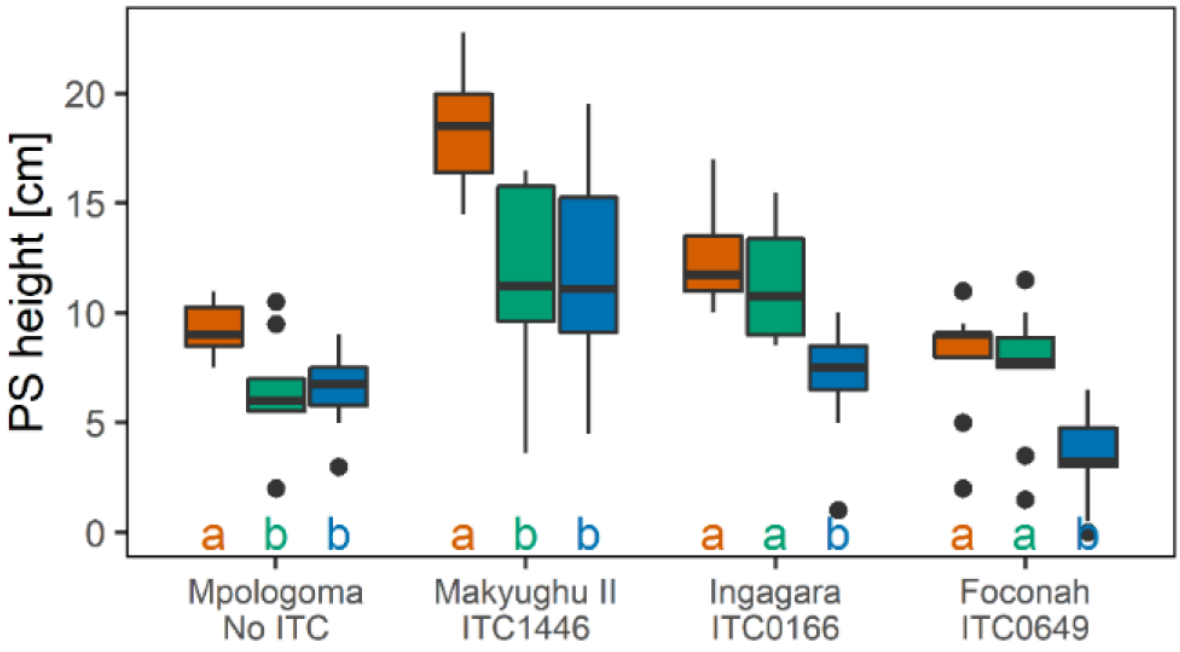


Figure 10: Pseudostem growth for six well growing cultivars. Significant growth differences between temperature treatments are indicated by different letters (Tukey HSD, following cultivar specific 1 way ANOVA (growth ~ treatment), alpha = 0.05).

50 days of growth under diverse temperature regimes leads to cultivar specific performance for example related to damage, leaf area development or plant mass distributions. This explains why

additional variables were measured (Table 2). These allow a better insight into the cultivar physiology. However, not all variables are equally important for every scientific question. Using sparse Partial Least Squares based discriminant analysis (sPLS-DA) the variables explaining most of the temperature treatment specific variance are identified. A new set of variables is constructed based on the original phenotypic variables. The new variable set is ordinated in order to maximize covariance between the phenotypic variables (X) and the output (Y: categorical: treatment) (Figure 11). All plants can be compared in this new set of variables. Especially component 1 (the first new variable) explains a lot of the covariance with the temperature regimes. The phenotypic variables constructing component 1 are mainly related to growth features (height pseudostem, leaf / canopy area). These can be identified in the loading plot. Negatively associated with the temperature treatment are the ratioRootShootFresh and ratioRootLeafFresh. A different investment in root biomass compared to shoot and leaf is a common response to unfavorable root zone temperatures given that the reached root zone temperatures are not detrimental (Koevoets et al., 2016). As such the multitude of phenotypic traits show the plasticity of plant response under different temperature regimes.

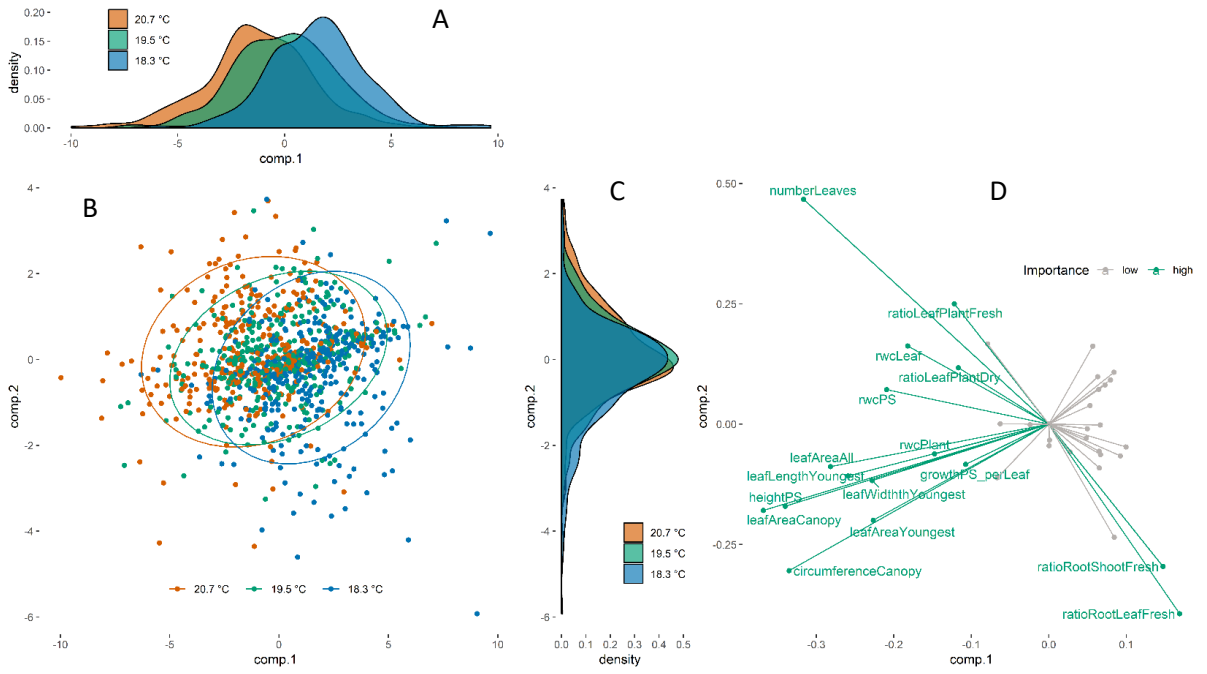


Figure 11: PLS-DA evaluation. Multivariate statistics allows to interpret the whole phenome changes due to the temperature treatment and to select the most important phenotypic variables. Output of sPLS-DA with X –matrix of phenotypic variables, and Y-matrix of temperature treatment is shown through panels A-D. (A) spread of the sample score along sPLS-DA component 1, colored per treatment, (C) spread along sPLS-DA component 2. (B) score plot of the sPLS-DA with 95% confidence interval ellipses drawn per temperature regime. (D) Variable loadings underlying the score plot (B). The most important variables are displayed and named in green, less important in gray

Conclusion

We introduced the bananaTainer as a device to initiate high throughput *Musa* germplasm screening. Germplasm collections such as the ITC (Bioversity International) can efficiently be screened to screen for suitable plant material to meet future challenges. Its flexible, setup is exemplified for growth

performance evaluation in three relevant temperature regimes (average temperature of 20.7°C, 19.5°C, 18.3°C). These results should be validated in experiments closer to the agricultural reality such as in the greenhouse or field, but these validation trials can then be limited to the most promising cultivars. Also additional agro-ecosystem specific growth experiments can be done, targeting banana growth in for example the lowlands.

Climate smart cultivar selection through gravimetric transpiration phenotyping

Introduction

The phenology of banana is a function of its vegetative growth, determined by cultivar/subgroup specific thresholds (Taulya et al., 2014). Water deficit hampers growth by reducing cell expansion in roots, leaves, and fruits (Turner et al., 2007). During the vegetative phase water deficit slows down growth, hence postpones bunch initiation and lowers the yield expressed per unit time (Taulya et al., 2014). The slightest reductions in soil water content already affect stomatal conductance, transpiration, and photosynthesis (Turner et al., 2007). Leaves are formed inside the pseudostem and are gradually pushed out, at which point they enroll and fully gain photosynthetic functionality (Turner et al., 2007). Banana plants tend to avoid the leaf water potential from dropping below a critical water potential, and keep the leaves hydrated. During bunch formation water deficit limits fruit filling and impairs leaf functionality, thereby reducing their photosynthetic capacity (Carr, 2009; Turner et al., 2007).

Improvement in crop management and cultivation can alleviate yield gaps. Still the production remains subject to the inherent capacities of the used cultivars (Godfray et al., 2010; Negin and Moshelion, 2017). For banana, available genetic diversity in wild relatives and/or (landrace) cultivars is collected, for example in the International Transit Center (ITC, Bioversity)(Vandenhouwe et al., 2003). Collections alike should be addressed to increase the adaptive character of (future) banana production systems while reducing the risks (Dempewolf et al., 2017; Lipper et al., 2014). Farmers in low input, agricultural systems such as the East African highlands do risk mitigation by selecting several cultivars adapted to the (local) agronomic requirements (Hauser and Van Asten, 2010; Nsabimana et al., 2008). Our goal is to rationalize these cultivar selection decisions. This requires comprehensive water deficit research in banana, aimed at gaining physiological insight applicable for farmers' field cultivar selection. Upon water deficit, stomatal conductance is reduced, hence drought tolerance research focusses on the antagonistic relationship between growth (carbon uptake) and water loss (Vadez et al., 2013). We analyzed a multitude of possible transpiration (water loss) related traits. The final output is a set of useful traits illustrating the interplay between plant transpiration and vegetative growth. The evaluation of these traits in real agro-ecological settings, and the comparison to other relevant cultivars, will improve the knowledge base for climate smart cultivar selection.

Material and methods

Plant material and growing conditions

Four different experiments were conducted spread over a year in our greenhouse in Leuven. Two cultivars were used: Cachaco, Bluggoe subgroup (ABB, ITC0643) and Mbwazirume, Mutika/Lugujira subgroup (AAA, ITC0084). Plant material was obtained from the banana biodiversity collection of Bioversity International, hosted at KU Leuven. *In vitro* plantlets were acclimated in the greenhouse in small pot soil pots for 7 weeks prior to the experiment. Plants were transferred to bigger (10 L) pots 1 week before the start of the experiment. At the start of the experiments the most homogenous plants per cultivar were selected based on leaf area. During the experiment the plants are kept under sufficient moisture conditions (pF < 2.0) for a period of time (20 - 30 days), after which the soil water content is lowered to a pF of 2.7 through evapotranspiration.

The greenhouse is equipped with a multi lysimeter setup of high precision balances (max weight 12kg, 1g accuracy, Phenospex, Heerlen, NL) each of which loaded with one plant. Every minute balances register and store the system weight. The coupled, balance specific, watering unit ensures irrigation

(5 mL accuracy) per plant on a daily basis at 5:15 AM. The composition of the provided nutrient solution is: 1.53 g/L $Ca(NO_3)2.4H_2O$, 0.33 g/L KNO_3, 0.025 mL/L Fe-EDTA (4.5%), 0.34 g/L K_2SO_4, 0.48 g/L $MgSO_4.7H_20$, 0.27 g/L KH_2PO_4, 0.11 mg/L Molybdate.$4H_2O$, 0.56 mg/L $ZnSO_4(7H_2O)$, 0.16 mg/L $CuSO_4(5H_2O)$, 7.69 mg/L $MnSO_4(H_2O)$, 2.01 mg/L H_3BO_3.

Water conditions

Water lost through transpiration is replaced based on plant specific target weights. These targets are determined weekly by a cumulative breakdown of the whole system weight (Eq. 1). The system (m_{tot}) consists of (1) a plastic pot and tray with a fixed weight ($m_{plastic}$), (2) dry soil (m_{soil}), (3) a plant (m_{plant}), and (4) soil water (m_{water}). The weight of the dry soil is a function of its volume (ρ=0.2267 g/cm³), measured after pot filling at the beginning of each experiment. The soil volume in the pot is considered constant as no soil is removed or added.

$$m_{tot} = m_{plastic} + m_{soil} + m_{plant} + m_{water} \quad \text{(Eq. 1)}$$

The plant weight (m_{plant}) is estimated through measurement of the top view leaf area. The top view leaf area correlates to the aboveground biomass and allows to estimate the plant weight based on a cultivar specific linear regression (*mass [g] ~ a * leaf area [cm²]*). For Cachaco: a=0.0697 g/cm² (R^2_{adj}=0.96, n=123), and for Mbwazirume: a=0.0743 g/cm² (R^2_{adj}=0.94, n=53). These regression values are based on greenhouse and growth chamber grown plants collected over several previous experiments and collected an in-house database.

The soil water treatment (Ψs [pF]) is correlated to the volumetric water content (θ), the volume of water per soil volume, and is a defined property of the used potting soil (Eq 2). During the first phase of the experiment plants are kept under well-watered conditions. Upon initiation of the water deficit treatment, irrigation is withdrawn until plants are at pF 2.7. After which they are kept at pF 2.7 by deficit irrigation. The plants determine their own speed of soil drying by their transpiration rate. All supplied water is assumed plant available. The soil is covered to prevent soil evaporation. On a weekly basis the target weights are updated, incorporating plant growth.

$$\Psi_s = 217.35 \times \theta^4 - 536.57 \times \theta^3 + 484.7 \times \theta^2 - 192.58 \times \theta + 30.429 \quad for\ \theta > 0.3637\ cm^3/cm^3 \quad \text{(Eq. 2)}$$

$$\Psi_s = -4.9875 \times \theta + 4.3369 \quad for\ \theta \leq 0.3637\ cm^3/cm^3$$

Determining whole plant transpiration values

Target weights for irrigation settings are calculated based on weekly measured top view images. After the experiment the daily plant weight is modelled from those weekly measurements in order to recalculate the soil water balance. The leaf area (LA) follows a power law equation over time with experiment specific start values (*dLA/dt = r*LA^b*). The daily modelled leaf area is converted to plant weight using the cultivar specific slopes relating leaf area and plant weight. The soil water weight is determined by breaking down the full system weight (Eq. 1) taking into account the weight of the plastic parts, the dry soil, and the plant weight. As such we obtain a daily volumetric water content which is converted to soil water potential (Ψs [pF]).

System weight values (measured every minute) are smoothed using the Savitzky-Golay filtering method with a window of 91 (signal R package). Daily the data between 5 and 7 AM was removed from the data to exclude effects of water supply. The actual transpiration rate per plant is obtained by differentiation of the smoothed data. From the dynamic transpiration data, measured every minute, the whole day transpiration was extracted as the difference of system weight between 9AM and 4AM. Around both time points the median value in a 20 minute interval was used. Similarly daytime transpiration (9AM-9PM), and night time water loss (9PM-4AM) were derived.

The leaf area increase throughout well-watered conditions was calculated as the difference in modelled leaf area at the start and end. When multiplied by the cultivar specific slope (Cachaco: a=0.0697 g/cm², Mbwazirume: a=0.0743 g/cm²) the leaf area increase is translated to a weight increase (i.e. growth). Leaf area increase and growth are expressed per day (cm²/day, g/day). The transpiration efficiency is calculated as the mass increase relative to the total transpiration volume over the same time period.

Additionally, we used segmented regression (segmented package, R (Muggeo, 2008)) in a time slot of 60 minutes before until 120 minutes after the start of the day. The breakpoint of the segmented regression indicates the timing of increase of transpiration. Only data with significant segmented regression (p-value Davies Test <0.05, segmented R package) and positive slopes were kept.

Environmental data collection

Temperature (°C) and humidity (%RH) data were collected using 6 data loggers (Trotec, DE) collecting data every 5 minutes. This data was averaged per 15 minutes. Vapour pressure deficit (VPD [kPa]) was calculated from temperature and relative humidity values (Eq. 3)

$$VPD = (1 - RH \times 0.01) \times 0.6108 \times \exp\left(\frac{17.27 \times temp}{temp + 237.3}\right) \quad \text{(Eq. 3)}$$

Radiation was collected every five minutes via a sensor (Skye instruments, UK) inside the greenhouse. Supplemental lighting of 14 W/m² at plant level is provided (between 7AM and 7PM) when incoming solar radiation was below 250 W/m². This is controlled by the central system of the greenhouse and is added to the sensed data. The light sensor was not illuminated by supplementary lighting.

Daily summary values were calculated like the daily transpiration values (24h: 6AM-6AM, day: 6AM-6PM, night: 6PM-6AM). Plants are linked to the data of the closest located sensor. Days with outlying (extreme) average environmental values were discarded. Table 4 shows the low and high boundaries between which data is kept.

Table 4: Environmental data boundaries. Days with environmental data between these boundaries were examined. Extremities were selected rank based (0:5 % and 95:100 %).

	Low	High
Maximal radiation [W/m²]	39	665
Average daytime temperature [°C]	22.5	26.1
Average daytime vapour pressure deficit [kPa]	0.66	1.51
Midday soil water content [pF]	1.7	2.8

Image analysis

Weekly images were taken using a top view digital camera (Canon, EOS 1200) using a blue background and a 150cm² red reference surface. Using in house build software the plant is separated from the background and the leaf area is calculated relative to the reference surface.

Time lapse top view images were additionally taken for 12 plants throughout the August 2018 experiment. These were taken every five minutes using cameras (Foscam FI9800P). For the daytime images (7AM-5PM), the plant was segmented from the background using a k-nearest neighbor algorithm. Hereby we used a lookup table with manual identifications of plant and background pixels. The lookup tables are constructed per plant based on 16 images. The width of the balance, visible on every image was taken as a scale to convert pixels to cm^2. Per plant, per day, the area was regressed to fit a second order polynomial function (*leaf area ~ intercept + a*time + b*time²*). From this function we derived convexity (b/intercept). Only days with a positive b-parameter were taken into account to avoid days in which a leaf was enrolling and the leaf area increased too steeply, masking the convexity.

Data analysis

All feature extraction and data analysis was done using R (V 3.4.3). Statistical interpretation of results took into account repeated measurements by incorporating a plant specific factor as random effect in a linear mixed model. The time effects were taken into account by using an autocorrelation structure. Sub experiments June 2018, Aug 2018, and Oct 2018 were conducted on 6 plants per cultivar. After a well-watered phase 5 plants per cultivar were brought to water deficit conditions, maintaining 1 in well-watered conditions. The Oct 2017 was slightly different using 12 plants per cultivar. After a well-watered phase, 6 were brought to water deficit conditions, while 6 were maintained under well-watered conditions.

In all displayed boxplots the bold middle line represents the median value. The box itself is bound by the first and third quartile and the whiskers extend to 1.5 times the interquartile distance from the box. Data points that fall outside the whiskers are considered as outliers and are represented as dots.

Results

Whole plant water relations on a multi-lysimeter installation

The major aim of this work is to study plant transpiration in a range of different environmental conditions (Table 4) relevant for the agro-ecosystem of interest, the East African highlands. We are interested in transpiration and growth aspects favoring one cultivar over the other for production in a certain agro-ecological zone. The accurate multi lysimeter system consists of 24 balances with a typical raw and smoothed data output represented in Figure 12. The soil moisture content is controlled gravimetrically, based on weekly defined target weights (Eq 1). In practice rewatering to these fixed target weights is slightly offset due to the plant growth between target weight updates. This deviation is small, and slightly more pronounced under water deficit. After the experiment plant growth is incorporated on a daily basis and the soil water content is corrected accordingly allowing proper interpretation.

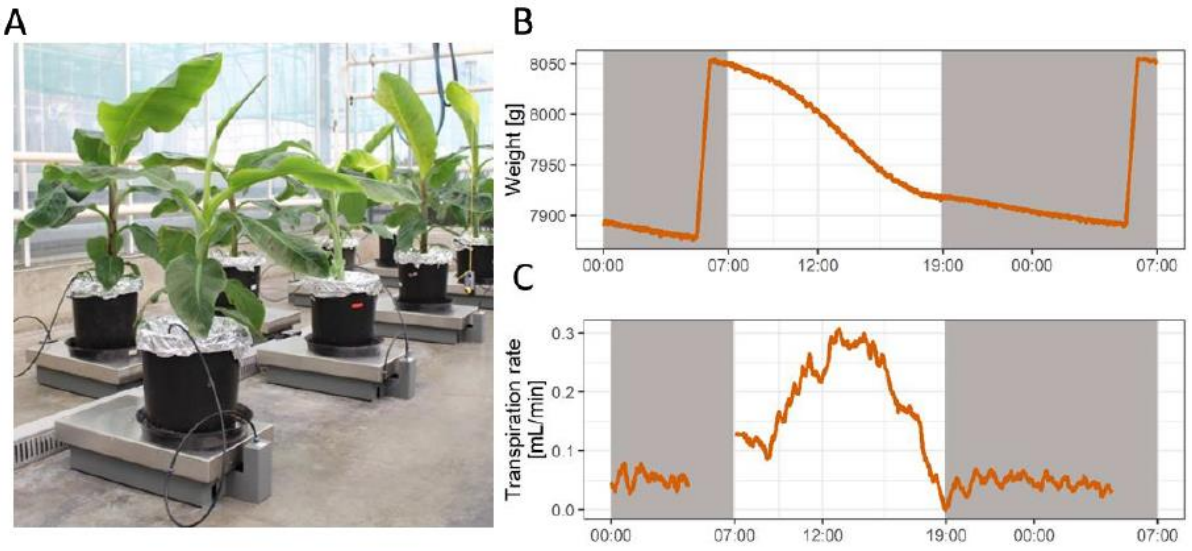


Figure 12: Lysimeter setup in the greenhouse. The setup consists of 24 balances, each loaded with a plant in a 10 L pot, coupled to an irrigation system. B) Raw data example for 1 plant (Mbwazirume cultivar). C) Savitzky-Golay smoothed transpiration rate of the same plant. Nighttime (radiation < 5W/m²) shown by grey background. Data around the irrigation events (5-7 a.m.) is excluded from the smoothed data.

In every sub-experiment plants were kept in well-watered conditions for 20-30 days, after which they were subjected to water deficit. Plants determine their own soil moisture levels by transpiration. The pot volume (hence soil water volume) is restricted, not all plants reach water deficit (pF 2.7) at the same time. This introduces a level of heterogeneity which is dealt with by pooling experimental results over the four sub experiments, creating one dataset of 1411 daily plant transpiration profiles. The data volume compensates for the variability introduced through the environment and the water treatment. In well-watered conditions Mbwazirume shows significantly more total transpiration (x1.61) and more growth per day (x1.93) compared to Cachaco (Table 5). The transpiration per leaf area [$mL*h^{-1}*m^{-2}$] cannot be distinguished significantly.

Table 5: Summary description of transpiration and growth-related traits. Mean (± SD) values calculated for well-watered (WW: 1.7 < pF < 2.0) and for water deficit (WD: 2.4 < pF < 2.8) conditions separately. Per cultivar significant differences ($p < 0.05$) between water regimes are indicated in bold. Significant differences between cultivars per water regime are indicated (*, $p < 0.05$).

Trait	Description	Cultivar	Watering regime			
			WW		WD	
TR_{DN}	24h transpiration [mL m^{-2} h^{-1}]	Cachaco Mbwazirume	37.2 ± 17.61 37.2 ± 12.981	nS	17.2 ± 5.414 19.8 ± 7.216	nS
TR_D	Day transpiration [mL m^{-2} h^{-1}]	Cachaco Mbwazirume	48.2 ± 21.194 49.3 ± 16.76	nS	24.3 ± 7.647 28.7 ± 10.541	nS
TR_N	Night transpiration [mL m^{-2} h^{-1}]	Cachaco Mbwazirume	18.5 ± 13.375 16.4 ± 9.51	nS	4.9 ± 2.418 4.7 ± 2.283	nS
TR_D/TR_N	Day/ night transpiration rate	Cachaco Mbwazirume	3.4 ± 1.55 2.6 ± 12.289	nS	6 ± 3.046 6.3 ± 3.341	nS
GR	Growth rate [g/day]	Cachaco Mbwazirume	6.3 ± 2.217 12.5 ± 6.531	*		
relGR	Relative growth rate	Cachaco Mbwazirume	1.7 ± 2.004 2.2 ± 1.839	nS		
LAI	Leaf area increase rate [cm^2/day]	Cachaco Mbwazirume	91.2 ± 31.828 168 ± 87.872	*		
TE	Transpiration efficiency [g/mL]	Cachaco Mbwazirume	0.052 ± 0.015 0.069 ± 0.019	*		

Whole day integrative responses of transpiration rate in relation to the environmental conditions

Progressive soil drying causes a decline in transpiration per leaf area, characterized by a sharp threshold, at high soil water contents estimated at pF 2 - 2.1) (Figure 13). This soil moisture content is cultivar specific: Mbwazirume (AAA) has a significantly higher transpiration during 0.1-0.2 pF units compared to Cachaco (ABB). The transpiration rate at pF 2 and 2.1 differs significantly ($p<0.1$) for these two genotypes. Compared to well-watered conditions (1.7 < pF < 2.0), the mean transpiration rate of Mbwazirume at 2.4 < pF < 2.7 is reduced by 47% and is even reduced to 53% for Cachaco (Figur 13)

Under well-watered conditions, an increased temperature or radiation leads to a higher transpiration rate for Mbwazirume and Cachaco plants (Figure 13). Increased radiation also increases transpiration under water deficit conditions. However, under water deficit, starting from pF 2.4, higher temperature no longer influences the transpiration rate (slope of regression ~0).

During the light period well-watered plants transpire 3.4 – 2.6 (Cachaco – Mbwazirume) times their nocturnal water loss. Under water deficit this increases significantly ($p<0.05$) to 6 – 6.3 (Cachaco – Mbwazirume) Significant cultivar differences in this ratio cannot be found.

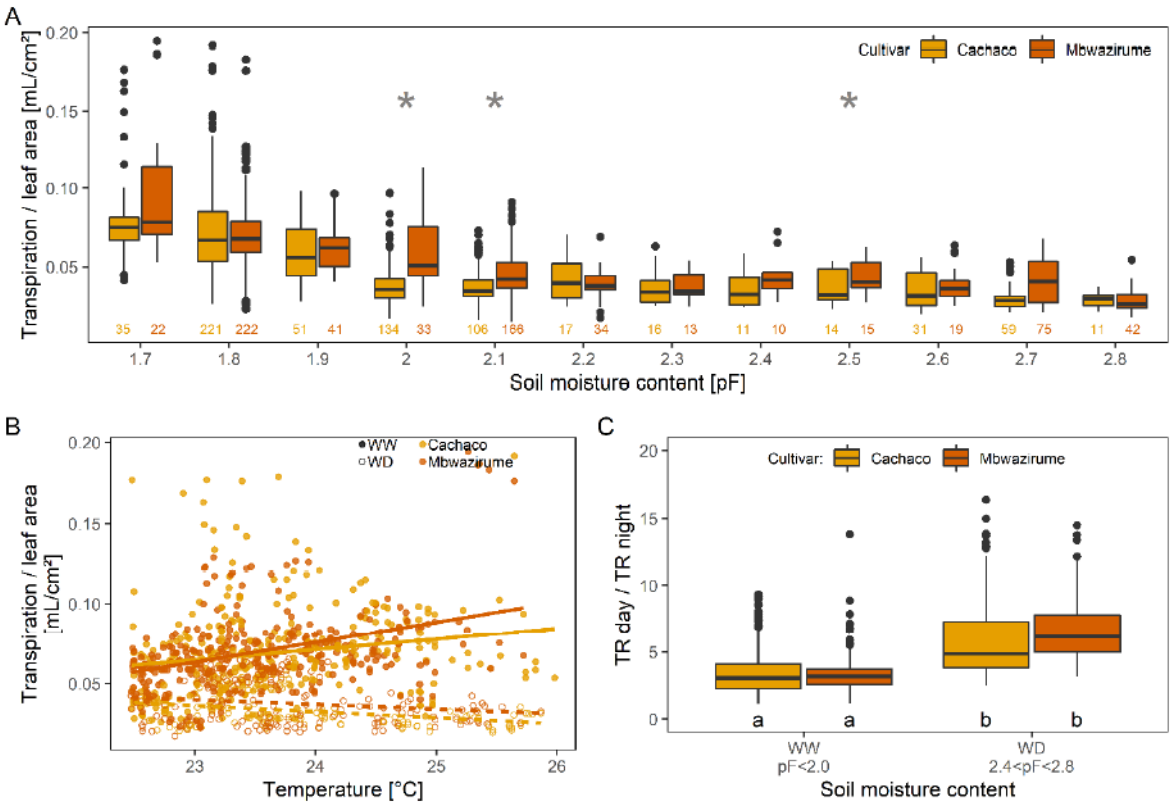


Figure 13: A) The whole day (24 h) transpiration per leaf area declines under increasingly dry soil conditions. Significant cultivar differences ($p < 0.1$) of linear mixed effects model per soil moisture level are indicated (*) with the number of unique plant - day combinations per cultivar. B) The daily (daytime, 12 h) transpiration per leaf area is influenced by the average daytime temperature. C) The daytime transpiration rate relative to the rate of nightly water loss increases under water deficit (WD) conditions compared to well-watered (WW). Significant differences ($p < 0.1$) of Tukey multiple comparisons of a linear mixed model are indicated by different letters. The model used cultivar * soil water content as fixed effects. A,C) plant/day combinations as random effects to account for repeatedly measured plants per day. B,C) WW: pF < 2.0, WD: pF > 2.4

The response to increasing radiation at the start of the day is cultivar specific

At the start of the day (radiation > 5W/m²) plants adapt to the changed (light) conditions. The transition upon irradiation is not immediate. Plants maintain their nightly water loss rate up to more than 1 hour, changing into a more rapid water loss rate only after a cultivar specific breakpoint. At the breakpoint, the transition from low water loss rate to a higher rate is sudden. This breakpoint occurs 20-30 minutes (well-watered – water deficit) earlier in Mbwazirume compared to Cachaco (Figure 14)

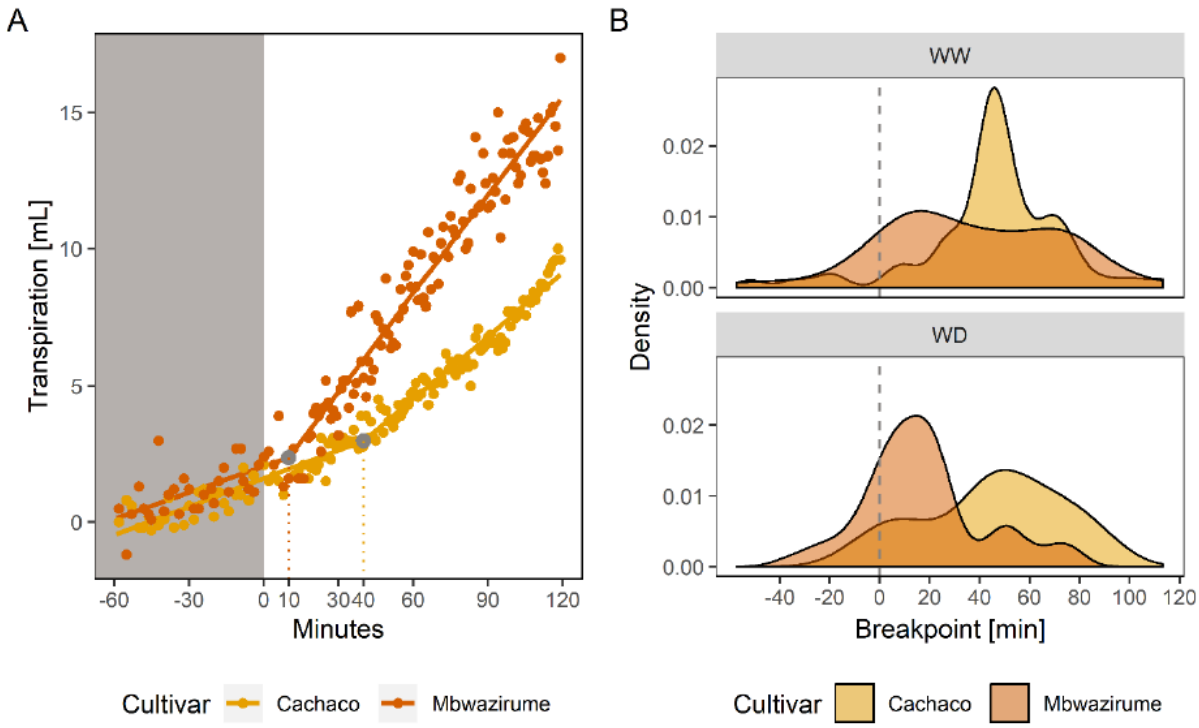


Figure 14: Cultivars differ in their onset of transpiration at the start of the day. Mbwazirume shows quicker increase of the transpiration when the day starts. A) Exemplified for two well-watered plants on a selected day. Segmented regression (lines) and their breakpoint (grey circles) indicated. B) The transpiration increase is postponed on average 30 minutes in Cachaco relative to Mbwazirume. WW: 1.7 < pF < 2.0, WD: 2.4 < pF < 2.8.

Discussion

Identification of transpiration related traits characterizing the Musa response to water deficit in East African highlands

Closing the yield gap for banana cultivation in the East African highlands relies on climate smart farm management and the choice of cultivars has a crucial role to play. There is a need for rational, knowledge-based cultivar selection so that farmers can do risk mitigation ensuring production stability. The major bottleneck is fast and reliable phenotypic evaluation (Pereyra-Irujo et al., 2012). In an ideal situation plants are tested in the prevailing environmental conditions on site. However, field trials are besides laborious and expensive also very difficult to control. Suitability characterization of a set of cultivars would need multi-location, multi-year studies without guarantee of success. Pre-field screening experiments in greenhouses or growth chambers save time and money by selecting the most promising cultivars for field validation trials (Junker et al., 2015; Negin and Moshelion, 2017; Vanhove et al., 2012). Growth chamber experiments offer great control over the environmental conditions, but are the most distant from agricultural reality. Greenhouse based experiments are more realistic but still offer control over the growing environment. Plants grow in soil and under actual water deficit in non-steady state conditions.

Not all plants reach the same water content simultaneously because they have a slightly different leaf area. The multi-lysimeter installation allows to precisely replace transpired water on an individual basis. In general, Mbwazirume transpired more than Cachaco, while it also grew more than Cachaco. Target weights are determined by the cumulative breakdown of the system weight (Eq 1).

Our aim is to dissect the complex relationship between transpiration and a fluctuating environment, without getting lost through the data heterogeneity. To do so, variables of different scales of detail were used throughout this study:

- Transpiration under progressive soil drying and its response to fluctuating temperature and light conditions
- Night water loss in relationship to the daytime transpiration
- Specific transpiration behavior around an external impulse such as the start of the day

Evaluation of the most important transpiration related traits for use in farmers' field

Plants are sessile organisms, but they act upon their environment. Responses take place on different hierarchical levels, in different tissues and in different time frames (Dhondt et al., 2013). We separate integrative responses to the prevailing environmental conditions from dynamic responses to fluctuations in external conditions (Figure 13 and 14). Plant functionality in a dynamic environment depends on balancing carbon fixation and water loss, a process essentially under stomatal control. In the search for climate smart cultivars for the East African highlands we study one side of this antagonistic relationship, water loss. Water deficit is a major environmental constraint for banana production, and transpiration cannot be neglected in this regard (van Asten et al., 2011). But the plant response to lower water content interferes with other management factors. For example the efficacy of fertilizer treatments is lowered under dry conditions (Okech et al., 2004).

We have extracted a few interesting transpiration related traits based on a year of phenotyping in fluctuating conditions in a greenhouse mimicking the environmental conditions of East African highlands. These traits will be discussed aimed at the applicability in farmers' fields.

Impact of reduced soil moisture content

Banana is sensitive to decline in soil water content, drought responses are already found in conditions which are mild for other species (Turner et al., 2007). At pF 2.0-2.1 there is a sharp decrease of transpiration (Figure 13). Decreases in transpiration such as these are observed in tomato, but at lower water contents (Halperin et al., 2017). This proves banana is a sensitive to water deficit, the critical soil moisture content for a response is cultivar specific (Figure 13). Cachaco has more strict, conservative, control over the transpiration under progressive soil drying, closing stomata at higher water potentials compared to Mbwazirume (Figure 13). Assuming both cultivars have the same potential to extract water from a stored water volume, Mbwazirume is predicted to reach its proper growth limiting soil potential earlier and so Cachaco will be able to continue to grow longer during the dry season. This is a benefit for Cachaco if the dry season is short and the growth retardation due to reduced transpiration moderate. A cultivar specific response to drought, potentially improving the water extraction potential, is to be expected. Cachaco has been reported to invest more in root biomass upon mild water deficit (van Wesemael et al 2019). Improved water uptake in general, or as a response to drought, allows plants to use otherwise inaccessible water (Tardieu et al., 2017).

Saving water for later is a reasonable drought avoidance strategy (Passioura, 2012). However, in a heterogeneous field, stored water is lost through competition by weeds, other (banana) plants, or the intercrop and by soil evaporation. Quick crop establishment suppresses growth of competitors by reducing available light through canopy closure. Major differences in establishment exist within the Musa biodiversity. Although field validation is required, Mbwazirume grows faster than Cachaco. Management practices minimizing soil evaporation and maximizing water infiltration such as mulching or rain water harvesting enlarge the stored water volume (Makurira et al., 2009; van Asten et al., 2011). This provides the plants with a bigger reservoir to continue growing when rainfall, or irrigation, is absent. However, also with more available water, maximizing carbon gain per mL of water consumption (transpiration efficiency) is crucial (Kissel et al., 2015; Passioura, 2012). Under well-watered conditions, the transpiration efficiency of Mbwazirume is significantly higher than Cachaco (TAB). We hypothesize that this is mainly due to the enhanced growth speed. As such, in well-watered conditions Mbwazirume shows a better use of water resulting into more growth, and is hence preferred over Cachaco.

Temperature is one of the main phenological drivers of banana growth (Turner et al., 2016; Turner and Lahav, 1983). We observe a linear transpiration increase with increase in temperature when ample water is available (Figure 13). The sensitive banana response reduces transpiration to a minimum under water deficit and an increase in temperature no longer causes increased transpiration. This fierce transpiration control occurs at mild soil water conditions (2.4<pF<2.8) and is in correspondence to Turner and Thomas (1998). They observed a linear increase of transpiration with VPD condition up to 2.8 kPa while no increase under water deficit. There is a strong, probably ABA-related, control on the transpiration, an increased evaporative demand by higher temperature cannot break this control.

The night is an important water loss factor

Banana plants lose a lot of water during the night (we observed around 30 % of the daytime transpiration in well-watered conditions, Figure 13). Nightly water loss is costly water spillage: water is lost without carbon fixation through (light 2606 dependent) photosynthesis. As such, reduced nightly water loss contributes to beneficial improved water use efficiency (Coupel-Ledru et al., 2016). Consequently we plea to still take night transpiration into consideration for water use efficiency improvement, although no significant cultivar differences were identified between Cachaco and Mbwazirume (Figure 13). In an evolutionary context, plants originating from niches without limitations

on available water benefit from nightly water loss (Chaves 2611 et al., 2016). It maintains transport and allows reallocation of nutrients over the plant on a 24/7 basis (Caird et al., 2007; Drake et al., 2013). Sources of nightly water loss are incomplete stomatal closure, cuticular water loss, guttation, … Incomplete stomatal closure at night is common in crops (Caird et al., 2614 2007). Banana is characterized by a huge root pressure, maintaining (nutrient) transport in humid air conditions (Turner et al., 2007). Guttation contributes to the nightly water loss through. We hypothesize that the proportion of nightly to daily water loss is strongly influenced by the daily radiation, increasing transpiration. A positive relation between daily radiation and TRD/TRN is expected. This could not be proven in our dataset, most probably because of the reached light intensities in Belgium that are rather low.

Upon water deficit (pF > 2.4) the ratio of day to night transpiration changes significantly (Figure 13). The day time transpiration is reduced by lower available moisture, but the nighttime water loss is decreased even more. This is the opposite response of what was observed in wheat (Claverie et al., 2018) and grapevine (Coupel-Ledru et al., 2016). In banana, Thomas and Turner (1998) observed a change in leaf elongation ratio between day and night upon water deficit. These authors attribute this to a decreased turgor, especially during the day for non-irrigated plants. We observe less night transpiration, which is essentially turgor driven, and hence complement these data. Water deficit reduces the water transport, mainly controlled by ABA. But this also reduces the water available for (nocturnal) root pressure built up and explains (part) of the bigger nocturnal transpiration decrease upon water deficit.

Speedy stomata at dawn

The dynamics of transpiration deserve attention as they show the adaptability of the cultivars in response to changes. A daily recurring game changer is the response to light at dawn. Plants start to receive radiation and shift from their nightly water loss to a higher transpiration rate. This transpiration transition is sudden and cultivar specific (Figure 14). The transpiration profile shows a breakpoint within 60 minutes after dawn. Mbwazirume starts on average 30 minutes before Cachaco. On a daily basis, both cultivars have the same average transpiration rate per leaf area. The major difference is the timing of the transpiration onset. On single leaf level (Drake et al., 2013) observed species specific, stomatal conductance patterns upon dawn, which can be interpreted by a breakpoint. However, our study seems to have a more sudden increase.

Stomatal conductance is controlled by ABA, environmental cues, mesophyll signals,… (Matthews et al., 2018). Water deficit influences the ABA content in the plant and reduces the average transpiration in our plants but water deficit does not change the transpiration onset. As plants are grown in the same conditions, differences in environmental cues as such cannot explain the observed cultivar specific response. The cultivar differences are to be found somewhere in the perception-to-action pipeline of stomatal opening. Stomatal opening is a function of hydraulic conductance, coordinated through a network involving ion channels, pressure gradients, enzyme (kinase/phosphatase) activity, … (Kollist et al., 2014; Lawson and Blatt, 2014). The energy (ATP) required for active stomatal opening through K+ transport is derived initially from lipid metabolism in the guard cells, after dawn this quickly transitions into ATP generated in guard cell specific photosynthesis (Geilfus et al., 2018). Hence this pipeline offers many levels for cultivar specific differences to be exploited for climate smart cultivar selection.

Additionally, anatomical differences between Cachaco and Mbwazirume could explain the difference in transpiration onset. Smaller stomata respond quicker, attributed to the higher membrane surface

to volume ratio (Drake et al., 2013). Cultivars with a high density of small, speedy stomata can couple carbon fixation and water loss more tightly in a dynamic environment (Drake et al., 2013; Lawson and Vialet-Chabrand, 2019). Stomatal conductance (gS) responds within minutes to hours upon the first light, while photosynthesis (carbon fixation) is photon flux mediated and starts within seconds or minutes. This time lag causes an imbalance between carbon uptake (controlled by gS) and carbon fixation by photosynthesis which decreases the intrinsic water use efficiency (Matthews et al., 2017). This makes that slower stomata are less optimal in natural, dynamic environments (Lawson and Vialet-Chabrand, 2019). Quick metabolic increase by photosynthesis and stomatal conductance allows plants to make most benefit of the usually lower morning VPD conditions (Chaves et al., 2016). Stomatal opening is also the result of the interplay between guard cells and the surrounding epidermal cells (Franks and Farquhar, 2007). Guard cells which are small compared to the surrounding epidermal cells have to overcome a relatively big turgor (back)pressure in order to increase stomatal conductance (Franks and Farquhar, 2007). Using microscopic images of imprints of the stomatal complex this could be studied.

We hypothesize that the quick stomatal opening in relatively low light (morning) conditions is negative for water use efficiency. Quick opening causes a relatively high water loss but in the limited morning light conditions carbon fixation is not as efficient. The coupling between stomatal conductance and photosynthesis is related to water use efficiency and could be worked out more using gas exchange measurements. The stomatal response largely determines the plants' life strategy. Quick stomatal openings are typical of more risk taking, for example ruderal, plant species. These maximize carbon fixation in order to fulfill the life cycle as quickly as possible (Merilo et al., 2014). Banana cultivars can be seen in this regard too: selected to produce as fast as possible, even in a variable, but mild, environmental condition, a benefit because of the strict correlation between vegetative growth and yield (Taulya et al., 2014). However, in stress prone conditions this risky strategy could lead to major yield losses when conditions are bad. This indicates the complexity of climate smart cultivar selection.

Conclusion

Whole plant responses in agronomic reality are a complex interplay between genotype, environment, and management. Dissecting these multifactorial interactions allows to study more comprehensible traits with more straightforward (environmental) responses (Tardieu and Tuberosa, 2010). However, it requires a posteriori field validation. Studying the water use of field grown bananas throughout a season in the East African highlands would be largely complex. Rather, we demonstrate the use of a gravimetric setup in a greenhouse to assess whole plant transpiration in response to environmental variability and cultivar specific differences among this.

Mbwazirume is in our setup the preferred cultivar with a more risk taking, opportunistic transpiration behavior. This cultivar grows best, has a more efficient water use. Mbwazirume postpones the transpiration rate decline upon progressive soil drying, which is a risky strategy if stress would be more severe or longer. Cachaco is a more conservative cultivar with a poorer growth performance in our greenhouse setup. The response to progressive soil drying and the postponed transpiration initiation upon dawn illustrates that this conservative cultivar potentially has a good potential to avoid a water deficit in more severe or longer stress conditions.

We are interested in the performance of Cachaco and Mbwazirume in more severe or longer stress conditions. If risk taking cultivars such as Mbwazirume appear vulnerable, management practices like irrigation are the only way alleviate the impact of the stronger stress. Otherwise production could be impacted seriously and other cultivars would have been a better choice. More conservative cultivars like Cachaco, could be able to avoid and or tolerate more severe stresses, and could be suitable in

more extensive farming setups. There is a certain, yet postponed, bunch production. However, these recommendations need further evaluation in greenhouse and in field trials. Nevertheless, the work presented here contributes to a comprehensive knowledge base upon which climate smart cultivar selection is rationalized. We think trials like these presented here provide a suitability ranking based on transpiration behavior. Once validated in situ, serving as (performance) benchmarks for further cultivars

References

Allen, C., Dos Santos, N., Gallagher, R., Chiu, G. N. C., Shu, Y., Li, W. M., … Bally, M. B. (2002). Controlling the Physical Behavior and Biological Performance of Liposome Formulations through Use of Surface Grafted. Bioscience Reports, 22(2), 225–250. https://doi.org/10.1023/A:1020186505848

Bernhard, M., Eubeler, J. P., Zok, S., & Knepper, T. P. (2008). Aerobic biodegradation of polyethylene glycols of different molecular weights in wastewater and seawater. Water Research, 42(19), 4791–4801. https://doi.org/10.1016/j.watres.2008.08.028

Blum, A. (2011). Plant Breeding for Water-Limited Environments. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-7491-4

Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M., … Doležel, J. (2016). Molecular and cytological characterization of the global Musa germplasm collection provides insights into the treasure of banana diversity. Biodiversity and Conservation, 1–24. https://doi.org/10.1007/s10531-016-1273-9

Dalal, A., Attia, Z., & Moshelion, M. (2017). To Produce or to Survive: How Plastic Is Your Crop Stress Physiology? Frontiers in Plant Science, 8(December), 1–8. https://doi.org/10.3389/fpls.2017.02067

De Boeck, H. J., Vicca, S., Roy, J., Nijs, I., Milcu, A., Kreyling, J., … Beier, C. (2015). Global Change Experiments: Challenges and Opportunities. BioScience, 65(9), 922–931. https://doi.org/10.1093/biosci/biv099

Dwivedi, S. L., Ceccarelli, S., Blair, M. W., Upadhyaya, H. D., Are, A. K., & Ortiz, R. (2016). Landrace Germplasm for Improving Yield and Abiotic Stress Adaptation. Trends in Plant Science, 21(1), 31–42. https://doi.org/10.1016/j.tplants.2015.10.012

Dwyer, D., & Tiedje, J. M. (1986). Metabolism of Polyethylene Glycol by Two Anaerobic Bacteria, Desulfovibrio desulfuricans and a Bacteroides sp. Applied and Environmental Microbiology, 52(4), 852–856.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., … O'Connell, C. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337–342. https://doi.org/10.1038/nature10452

Hegde, D. M., & Srinivas, K. (1989). Effect of soil matric potential and nitrogen on growth, yield, nutrient uptake and water use of banana. Agricultural Water Management, 16(1–2), 109–117. https://doi.org/10.1016/0378-3774(89)90045-0

Huang, Y. L., Li, Q. B., Deng, X., Lu, Y. H., Liao, X. K., Hong, M. Y., & Wang, Y. (2005). Aerobic and anaerobic biodegradation of polyethylene glycols using sludge microbes. Process Biochemistry, 40(1), 207–211. https://doi.org/10.1016/j.procbio.2003.12.004

Junker, A., Muraya, M. M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A. E., … Altmann, T. (2015). Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. Frontiers in Plant Science, 5(January), 1–21. https://doi.org/10.3389/fpls.2014.00770

Kawai, F. (2002). Microbial degradation of polyethers. Applied Microbiology and Biotechnology, 58(1), 30–38. https://doi.org/10.1007/s00253-001-0850-2

Langridge, P., & Reynolds, M. P. (2015). Genomic tools to assist breeding for drought tolerance. Current Opinion in Biotechnology, 32, 130–135. https://doi.org/10.1016/j.copbio.2014.11.027

Lipper, L., Thornton, P., Campbell, B. M., Baedeker, T., Braimoh, A., Bwalya, M., … Torquebiau, E. F. (2014). Climate-smart agriculture for food security. Nature Climate Change, 4(12), 1068–1072. https://doi.org/10.1038/nclimate2437

Marchal, R., Nicolau, E., Ballaguet, J. P., & Bertoncini, F. (2008). Biodegradability of polyethylene glycol 400 by complex microfloras. International Biodeterioration and Biodegradation, 62(4), 384–390. https://doi.org/10.1016/j.ibiod.2008.03.013

Michel, B. E. (1983). Evaluation of the water potentials of solutions of polyethylene glycol 8000 both in the absence and presence of other solutes. Plant Physiology, 72(1), 66–70. https://doi.org/10.1104/pp.72.1.66

Michel, B. E., & Kaufmann, M. R. (1973). The Osmotic Potential of Polyethylene Glycol 6000. Plant Physiol, 51, 914–916. https://doi.org/10.1104/pp.51.5.914

Mickelbart, M. V., Hasegawa, P. M., & Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nature Reviews Genetics, 16(4), 237–251. https://doi.org/10.1038/nrg3901

Money, N. P. (1989). Osmotic Pressure of Aqueous Polyethylene Glycols. Plant Physiology, 91(2), 766–769. https://doi.org/10.1104/pp.91.2.766

Moshelion, M., & Altman, A. (2015). Current challenges and future perspectives of plant and agricultural biotechnology. Trends in Biotechnology, 33(6), 337–342. https://doi.org/10.1016/j.tibtech.2015.03.001

Negin, B., & Moshelion, M. (2017). The advantages of functional phenotyping in pre-field screening for drought-tolerant crops. Functional Plant Biology, 44(1), 107–118. https://doi.org/10.1071/FP16156

Ortiz, R., & Swennen, R. (2014). From crossbreeding to biotechnology-facilitated improvement of banana and plantain. Biotechnology Advances, 32(1), 158–169. https://doi.org/10.1016/j.biotechadv.2013.09.010

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBImage-an R package for image processing with applications to cellular phenotypes. Bioinformatics, 26(7), 979–981. https://doi.org/10.1093/bioinformatics/btq046

Poorter, H., Fiorani, F., Stitt, M., Schurr, U., Finck, A., Gibon, Y., … Pons, T. L. (2012). The art of growing plants for experimental purposes: A practical guide for the plant biologist. Functional Plant Biology, 39(11), 821–838. https://doi.org/10.1071/FP12028

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE, 8(6). https://doi.org/10.1371/journal.pone.0066428

Scarlett, K., Collins, D., Tesoriero, L., Jewell, L., van Ogtrop, F., & Daniel, R. (2016). Efficacy of chlorine, chlorine dioxide and ultraviolet radiation as disinfectants against plant pathogens in irrigation water. European Journal of Plant Pathology, 145(1), 27–38. https://doi.org/10.1007/s10658-015-0811-8

Steuter, A. A., Mozafar, A., & Goodin, J. O. E. R. (1981). Water potential of aqueous polyethylene glycol. Plant Physiology, 67(1), 64–67. https://doi.org/10.1104/pp.67.1.64

Takeuchi, M., Kawai, F., Shimada, Y., & Yokota, A. (1993). Taxonomic Study of Polyethylene Glycol-Utilizing Bacteria: Emended Description of the Genus Sphingomonas and New Descriptions of Sphingomonas macrogoltabidus sp. nov., Sphingomonas sanguis sp. nov. and Sphingomonas terrae sp. nov. Systematic and Applied Microbiology, 16(2), 227–238. https://doi.org/10.1016/S0723-2020(11)80473-X

Taulya, G., van Asten, P. J. A., Leffelaar, P. A., & Giller, K. E. (2014). Phenological development of East African highland banana involves trade-offs between physiological age and chronological age. European Journal of Agronomy, 60, 41–53. https://doi.org/10.1016/j.eja.2014.07.006

van Asten, P. J. A., Fermont, A. M., & Taulya, G. (2011). Drought is a major yield loss factor for rainfed East African highland banana. Agricultural Water Management, 98(4), 541–552.

van Wesemael, J., Kissel, E., Eyland, D., Lawson, T., Swennen, R., & Carpentier, S.C. (2019). Using growth and transpiration phenotyping under controlled conditions to select water efficient banana genotypes. Frontiers in Plant Science 10, 352.

Vanhove, A.-C., Vermaelen, W., Panis, B., Swennen, R., & Carpentier, S. C. (2012). Screening the banana biodiversity for drought tolerance: can an in vitro growth model and proteomics be used as a tool to discover tolerant varieties and understand homeostasis. Frontiers in Plant Science, 3(176), 1–10.

Verslues, P. E., Agarwal, M., Katiyar-Agarwal, S., Zhu, J., & Zhu, J. K. (2006). Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. Plant Journal, 45(4), 523–539. https://doi.org/10.1111/j.1365-313X.2005.02593.x

Verslues, P. E., Ober, E. S., & Sharp, R. E. (1998). Root Growth and Oxygen Relations at Low Water Potentials. Impact of Oxygen Availability in Polyethylene Glycol Solutions. Plant Physiology, 116(4), 1403–1412. https://doi.org/10.1104/pp.116.4.1403

Genomic constitution characterization in banana ABB allotriploids

Abstract

Bananas (*Musa* spp.*)* are a major staple food for hundreds of millions of people in developing countries. The cultivated varieties are seedless parthenocarpic clones of which the origin is still not completely understood. The most important cultivated banana varieties are triploids with an AAA, AAB, or ABB genome constitution, the A and B genomes being provided by *Musa acuminata* and *M. balbisiana*, respectively. Prior to the availability of the banana reference sequence, it was shown that A/B inter-genome recombination was relatively common in banana cultivars and that cultivated triploids were more likely to have passed through an intermediate hybrid.

Using single-nucleotide polymorphism (SNP) markers called from RADseq data, we studied the chromosome structure of 29 ABB genotypes spanning these subgroups. Several inter-genomic replacements were identified, where the A genome was substituted by the B genome counterpart and vice versa. The replacement patterns indicate at least 7 founding events at the origin of the ABB cultivars. Most of the subgroup assignations were confirmed, but several inconsistencies were highlighted that necessitate a revision of the subgroup classification.

The results of this study represent an important step in unravelling the origin of polypoid bananas, which will help review the structure of the subgroups and the classification of ABB cultivars. The observed frequent recombination between A and B genomes suggests *M. balbisiana* as a source of potential useful variability for new cultivars better adapted to environments and able to answer the challenges of climate changes, e.g. increased drought.

Introduction

Bananas (*Musa* spp.) are a genus of herbaceous monocotyledons belonging to the Zingiberales order. *Musa* originated in Southeast Asia where it was domesticated; then during the period of European colonization, the banana culture was spread to tropical areas of Africa and the Americas.

The large majority of cultivated banana (cultivars) are triploid (2n=3x=33), i.e. composed of three sets of 11 chromosomes. These cultivars are sterile due to the triploidy that hampers the production of balanced gametes and, consequently, their propagation is clonal, i.e. based on multiplication of vegetative tissues. The combination of sterility and parthenocarpy ensures the production of seedless fruits that are edible. Two species, *M. acuminata* and *M. balbisiana*, are involved in large scale cultivation contributing their genomes, A and B respectively, to the subgenomes of triploid cultivars. Triploidy establishment occurs by combination of a normal gamete (n=x=11) and a gamete with double set of chromosomes (n=2x=22) originated from normal meiosis of a tetraploid plant (2n=4x=44) or from an unreduced meiosis by a diploid genotype (e.g. an AB hybrid).

Triploid cultivars are classified as either autotriploid, composed of three A subgenomes (AAA), or allotriploid, composed of A and B subgenomes (AAB and ABB). A secondary level of classification for allotriploid banana cultivars is mainly based on the similarity of morphological parameters to traits differentiating *M. acuminata* and *M. balbisiana* (Simmons and Shepherd 1955). Molecular analysis of organelle genomes allowed the characterization of cultivar cytotypes and hypothesis of the crossing pathways that led to their origin (Boonruangrod et al. 2008). Secondary levels of classification were proposed based on inferred parental origins (Simmonds, 1962) and on allotriploid origin inference based on cytotypes (i.e. on the mitochondrial and plastidial genomes) (De Langhe et al. 2010). These studies pointed out the genomic complexity of the allotriploid cultivars and suggested the occurrence of backcrosses with parental species based on the observed residual fertility in some allotriploid cultivars (De Langhe et al. 2010).

Flow cytometry increased the resolution of genomic investigations, allowing accurate measure of the ploidy level and molecular markers such as Single Sequence Repeats (SSRs) and providing a multi-locus survey of the parental alleles' contribution (Christelova et al. 2016). Recently, new generation sequencing (NGS) technologies led to a genome wide detailed survey that provided evidence of the origin of a small portion of chromosomes (Baurens et al. 2018).

A former genome wide study highlighted that allotriploids are not mere additions of sets of 11 chromosomes; regions with variable subgenome ratios were detected along the chromosomes (Baurens et al., 2018), confirming the occurrence of recombination between subgenomes.

The goal of this study is to survey a wide range of ABB allotriploids by using SNP frequency data to check the possible presence of regions with allele frequency ratios different from the expected B2:A1 and to evaluate the usefulness of genome wide analysis in cultivar classification.

Material and methods

Cultivars analyzed

Leaf samples from banana genotypes were supplied by the Bioversity International *Musa* Transit Centre hosted at KU Leuven, Belgium. The passport data of these accessions with reported ABB genome constitution was retrieved from the *Musa* Germplasm Information System (MGIS) (Ruas et al., 2017) (**Table 6**).

Accession code	Accession name	Group	Subgroup	Geographical origin
ITC0245	Safet Velchi	AB	Ney Poovan	India
ITC1034	Kunnan	AB	Ney Poovan	
ITC1747	Agniswar	AB	Kunnan	Indonesia
ITC1729	Padali Moongil	AB	Kunnan	
ITC1752	Poovilla Chundan	AB	Kunnan	
ITC1751	Adukka Kunnan	AB	Kunnan	
Sum002	Muku Bugis	AB	-	Indonesia
Sum004	Mu'u Seribu	AB	-	Indonesia
ITC1880	Mu'u Pundi	AB		Indonesia
ITC0026	Sabra	ABB	Unknown	Gabon
ITC0632	Cachaco enano	ABB	Bluggoe	Colombia*
ITC0643	Cachaco	ABB	Bluggoe	Colombia*
ITC1728	Sambrani Monthan	ABB	Bontha	India
ITC1746	Bankel	ABB	Pisang Awak	India
ITC1748	Boddida Bukkisa	ABB	Pisang Awak	India
ITC1483	Monthan	ABB	Monthan	India
ITC0767	Dole	ABB	Bluggoe	France*
ITC0361	Blue Java	ABB	Ney Mannan	Fiji*
ITC1750	Ney Vannan	ABB	Bontha	India
ITC0123	Simili Radjah	ABB	Peyan	India
ITC1600	INIVIT PB-2003	ABB	Saba	Cuba*
ITC1138	Saba	ABB	Saba	France*
ITC0659	Namwa Khom	ABB	Pisang Awak	Thailand
ITC0339	Pisang Awak	ABB	Pisang Awak	Australia*
ITC0472	Pelipita	ABB	Unknown	Philippines
ITC0053	Bom	ABB	Pisang Awak	France*
ITC0087	Kayinja	ABB	Pisang Awak	Burundi*
ITC0101	Fougamou 1	ABB	Pisang Awak	Gabon
ITC0526	Kluai Namwa Khom	ABB	Pisang Awak	Thailand*
ITC1599	Kambani	ABB	Pisang Awak	Tanzania
ITC1721	Karpuravalli	ABB	Pisang Awak	India
ITC1719	Chinia	ABB	Pisang Awak	India
ITC1749	Vananthpurani	ABB	Pisang Awak	India
ITC0396	Pelipita	ABB	Unknown	Costa Rica*

ITC0397	Pelipita Majoncho	ABB	Unknown	Costa Rica*
ITC0652	Kluai Tiparot	ABB	Unknown	Thailand
ITC1700	Kepok Kuning	ABB	Saba	Indonesia
ITC1745	Kepok Tanjung	ABB	Saba	Indonesia

Table 6: List of 38 accessions analyzed. * origin of the previous *ex situ* collection

DNA extraction and RADseq data generation

Genomic DNA was extracted using a modified MATAB method (Risterucci et al. 2000). DNA libraries were constructed and sequenced using the HiSeq2000 (Illumina) technology at Beijing Genomics Institute.

Detection of Homoeologous Recombinations (HR)

Two approaches have been used to survey the subgenomic structure of the cultivar genomes. The first one was based on the method of Baurens and al (2018). Based on known sequence variability in the A and B genomes, SNP variants were assigned to the parental genome based on known sequence variability and its frequency was reported in a graphical representation of each chromosome, along with the total count of SNPs.

In the second approach, SNP variants unknown on the A genome (established on AAA cultivars 'Grande Naine' and 'Mbwazirume' and the A genome reference) were assigned to the B genome and used to measure the B alleles' frequencies along the chromosomes.

Results

Genome wide survey of A and B assigned SNP allele frequency in allotriploid ABB

SNPs detected in RADseq datasets of 29 genotypes were surveyed with the approach 1 and 2 to check if, in the 11 chromosome groups, the presence of regions where the A/B subgenome balance was different from the expected one based on known genomic structure (i.e. B2:A1 for the allotriploid ABB) (**Figure 15**). Results were consistent between the two approaches.

Only two genotypes did not show any regions with unexpected subgenome balance: 'Kepok Kuning' (ITC1700) and 'Kepok Tanjung' (ITC1745). Among the remaining 27 ABB genotypes, several regions with subgenome balance B3:A0 or B1:A2 were detected, whereas no region with B0:A3 was found (e. g. in Figure 15, chromosomes 4, 9 and 11 have unbalanced regions). Genome comparisons among ABB cultivars revealed that chromosome regions with deviating subgenome balances (hereafter called deviating regions) are in some cases shared among the analyzed genotypes. Identical or very similar patterns of deviating regions were found among the 27 genotypes that were then grouped into seven groups (**Table 1**).

The genotypes in each group have the same pattern of deviating regions. Three deviating regions appear to be shared by three groups (1a, 1b and 1c): in large terminal regions of the long arm of both chromosomes 4 and 11 the detected structure was B3:A0, and in the short arm of chromosome 9 an

interstitial region was B1:A2. Groups 1a and 1c also shared the B3:A0 terminal region of chromosome 11's short arm. Additional deviating regions were specific to each group.

Group 4, including genotypes classified as 'Pelipita', had two whole chromosomes (i.e. chr02 and chr11) with pattern B3:A0 and the chr03 has a B3:A0 pattern but an interstitial region in its long arm with the expected B2:A1 pattern.

'Kluai Tiparot' (ITC0652), uniquely representing group 5, had a particular deviating pattern that was not shared with any of the other studied genotypes. In this cultivar, all the detected unbalanced regions were all B3:A0 and three chromosomes (i.e. chr02, chr08 and chr10) resulted completely in the pattern B3:A0, in addition to several chromosome parts.

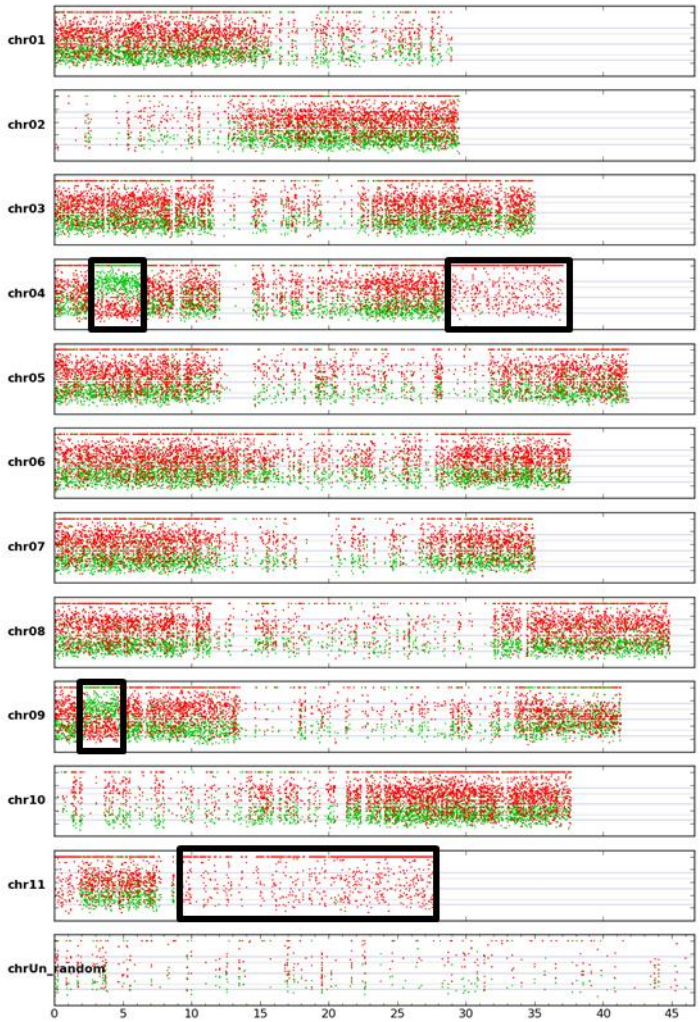


Figure 15: Mosaic genome structure of A/B interspecific ABB genome for the 1a group along the 11 chromosomes. Coverage ratios displayed by red and green dots represent allelic variation corresponding to B and A genomes respectively. Open black boxes indicate chromosome segments that differ from the expected distribution in a ABB cultivar. The scale is indicated in Mb.

Specific deviating regions were found in particular genotypes included in some groups. In many cases, these deviating regions are characterized by the presence of only two chromosome regions (e.g. B2:A0 or B1:A1), indicating the loss of whole or partial chromosomes (**Figure 16**). In Group 1a, 'Dole' (ITC0767) missed the A chromosome 8, resulting in a B2:A0 pattern along the whole chromosome (**Figure 16b**). From a technical point of view, regions B2:A0 and B3:A0 have similar patterns in both survey methods used in this study: only B variants were detected. However, the absence of one chromosome in 'Dole' can be detected by observing the distribution of heterozygous B variants. In fact, the point distribution is unimodal around a frequency value of 0.5 when two B chromosomes are present, whereas it is bimodal around 0.33 and 0.67 frequency values in the case of three chromosomes B. Coverage data are also used to corroborate the chromosome loss inference.

The short arm of the chromosome 6 of 'Sabra' (ITC0026, group 1a) is characterized by a B1:A1 pattern, whereas in the long arm four regions can be distinguished with the following patterns of B2:A1, B3:A1, B1:A1 and B2:A1 (from the centromere to the telomere) (Figure 16). The pattern of region B3:A1 can be distinguished by the B2:A1 due to the more separated variant frequency distribution (0.25 and 0.75 for A and B variants, respectively), and to the higher coverage of the corresponding region (**Figure 16c**)

In group 1b, chromosome group 5 of 'Simili Radjah' (ITC0123) showed unique deviations: 1) a small terminal region of the short arm has only B genome derived variants; 2) a whole long arm was missing, resulting in B1:A1 configuration along most of the arm; 3) B2:A0 in the terminal region. B1:A1 regions can be distinguished from the B2:A1 due to the frequency of A and B variants around the 0.5 value vs two distinguished distributions around 0.33 and 0.67 (A and B variants, respectively) (**Figure 16d**).

In group 1c, the genotype 'INIVIT PB-2003' (ITC1600) showed a unique interstitial region B2:A2 in the chromosome 10 long arm (Figure 16e). In group 2, the genotype 'Pisang Awak' (ITC0339) has a unique B1:A1 configuration in the terminal part of the chromosome 7 short arm and one (or possibly two) interstitial regions B3:A0 in the chromosome 10 short arm (**Figure 16f**).

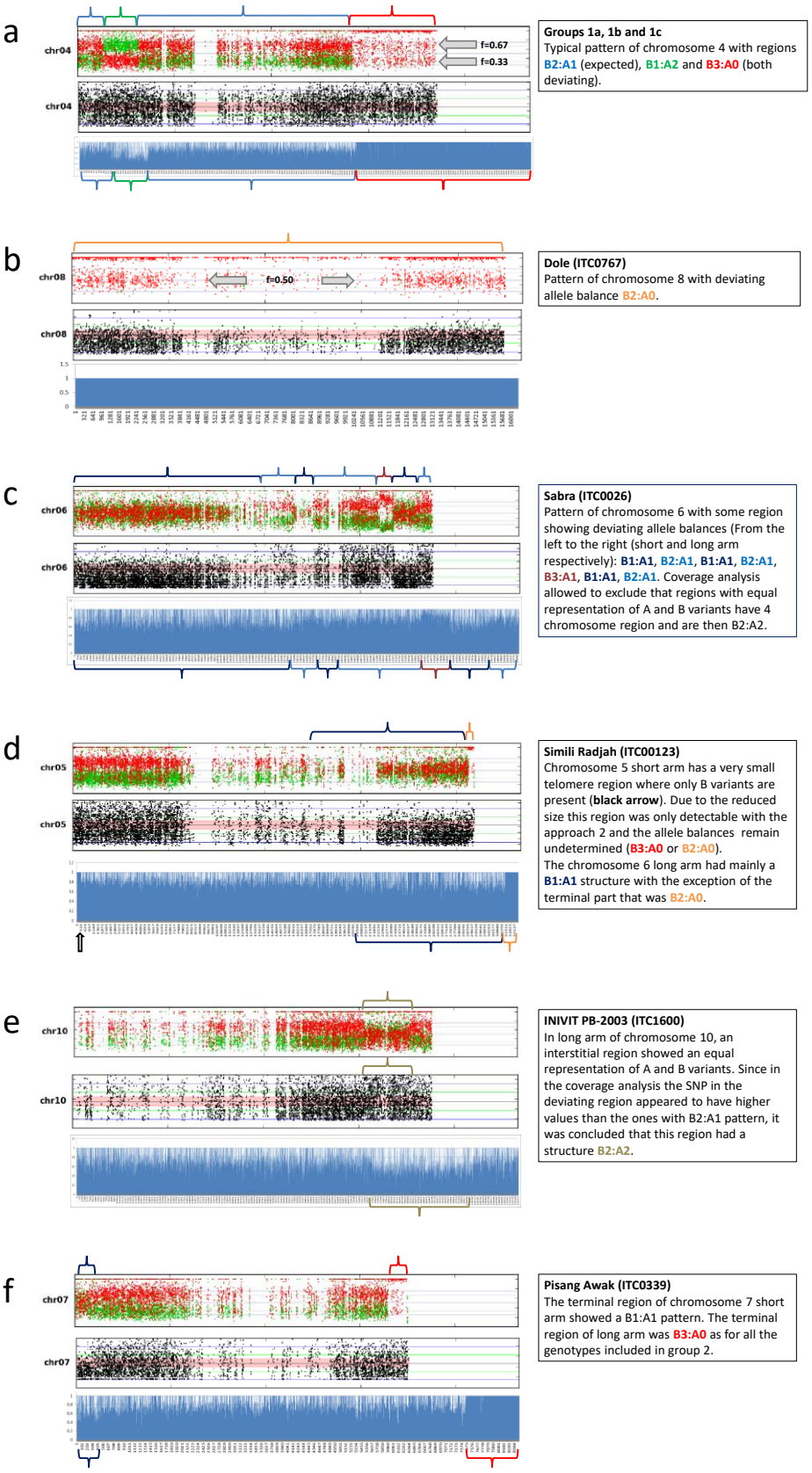


Figure 16: Aneuploid patterns for 5 subgroups in ABB. Euploid recombinant pattern (a) compared with aneuploid patterns.

Genome wide survey of A and B assigned SNP allele frequency in hybrids AB

SNPs detected in RADseq datasets of 9 hybrid AB genotypes were surveyed with the approach 1 to verify the uniformity of the heterozygous genome structure.

Three Indonesian hybrids (Muku Bugis, Mu'u Seribu and Mu'u Pundi) showed the expected presence of A and B genomes along all 11 chromosomes. In the remaining six hybrids, chromosome regions without B assigned SNPs were found. Four patterns were observed in the six hybrids (**Figure 17**). Since the frequency of many A genome assigned SNPs is still around 0.5, the presence of two A homologous regions is inferred, which implies the lack of B genome assigned SNPs is not due to deletions but to replacement with A genome counterparts by recombination.

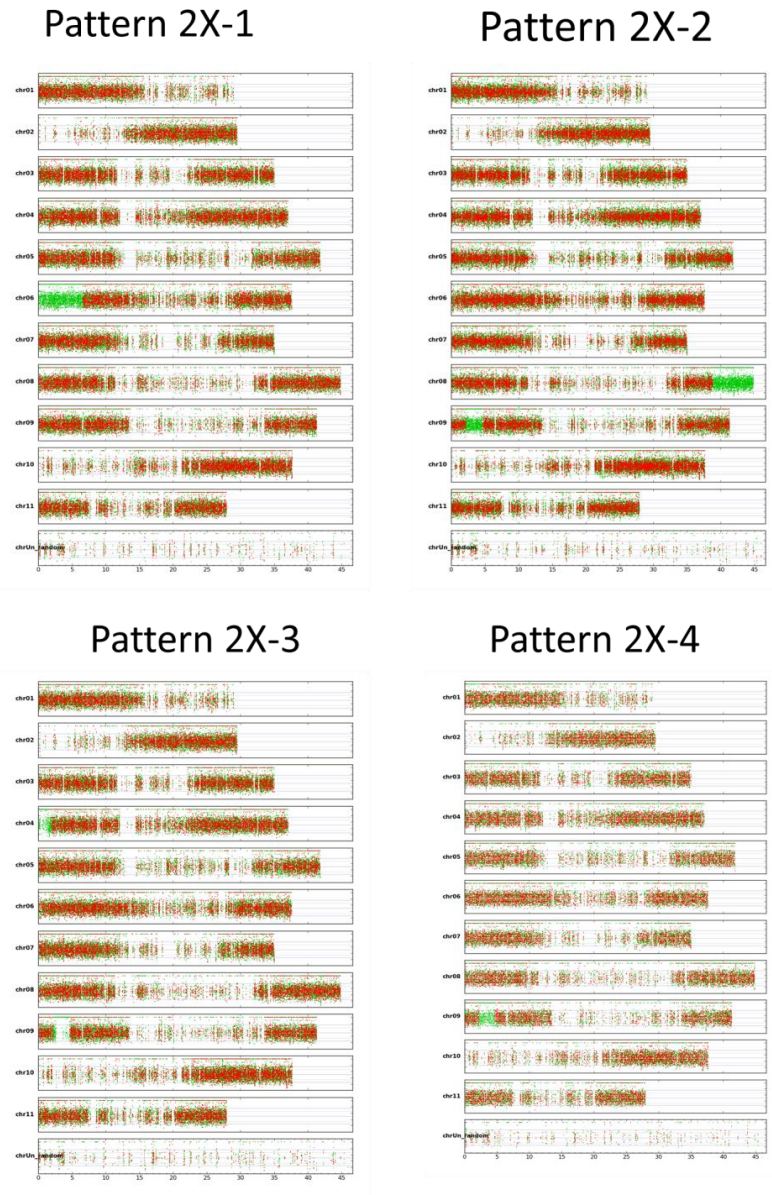


Figure 17: Aneuploid patterns for 5 subgroups in ABB. Euploid recombinant pattern (a) compared with aneuploid patterns.

Discussion

In this study, we take advantage of NGS technologies, and in particular RADseq data, to survey the genomic structures of a panel of allotriploid (ABB) banana cultivars. Homoeologous recombination (HR) was inferred when, in a given chromosome, a change in A/B allele contribution was observed between adjacent regions.

Genome wide analyses were performed with two different approaches. In the first, SNP variants were assigned to respective A or B subgenomes based on known variability in both genepools. In the second approach, only the better-known A variability is taken into account to detect and remove the SNPs assigned to the A subgenome. All variants alternatives to the A reference and not known in the A genome variability were assigned to the B subgenome. In both approaches, the SNP variant frequencies were used to infer the local genomic constitution.

Results obtained with both approaches were perfectly consistent. The interpretation is easier with the first approach, especially in regions characterized by changes in copy number as partial aneuploidy due to segmental duplications or deletions (**Figure 16c**). However, the second approach can allow the detection of small deviating regions as in the short arm of chromosome 6 of 'Simili Radjah' (**Figure 16d**).

Globally, all accessions were confirmed as having an ABB genome constitution, i.e. the ratio between variants assigned to B and A subgenomes was B2:A1. Among the 29 genotypes surveyed, 27 had at least two chromosome portions where the observed ratio was B1:A2 or B3:A0.

The occurrence of regions with deviating ratios between B and A SNP variants indicates the B subgenome substituted its A counterpart (B3:A0) or vice versa (B1:A2). We conclude that the involved chromosomes underwent recombination that resulted in chromatin exchanges between subgenomes. Since sexual reproduction is supposed to be abolished by triploidy, which disturbs correct meiosis processes and hampers the production of viable gametes with balanced chromosomes, the recombination should have taken place before the allotriploid formation. In AB hybrids, the presence of B chromosomes with portions replaced by A chromatin indicates that the *M. balbisiana* parents of hybrids contained A genome introgressions.

Some of the genotypes were found sharing the whole pattern of unbalanced regions. This observation strongly suggests that all the genotypes with the same deviating pattern originated by vegetative propagation from a common ancestor having the same deviating pattern.

The observation of genotype groups sharing only some unbalanced regions (i.e. 1a, 1b and 1c) is more intriguing. In fact, identical or very similar deviating patterns (three unbalanced regions in the case of mentioned groups) obtained by independent recombinations is unlikely. Consequently, we hypothesize that unbalanced regions were generated in a two-step process, the oldest one common to the ancestors of the three groups. The shared events could be already present in the AB hybrid, transmitted by parental A and B gametes where interspecific introgressions were present; this hybrid then contributed with unreduced gametes (having new recombinations originated in independent meiosis) to the allotriploid ancestors of the three groups 1a, 1b and 1c. The presence in three diploid hybrid recombinant patterns of an identical or very similar recombination (interstitial introgression of A chromatin on first arm of chromosome 9B) strongly supports the hypothesis that parents of triploid ABB hybrids contained introgressions of the other parental species. Another possibility is that a

common allotriploid ancestor was able to produce descendants by sexual reproduction, then create new and independent recombinations. In the case of the above-mentioned groups, 1b could be the first generation of allotriploid, whereas 1a and 1c could be derived from 1b by production of 2x gametes where new independent recombination events take place, and then create new unbalanced regions.

The genotype 'Kluai Tiparot' (ITC0652) can be distinguished from all other surveyed accessions for two features: all deviant regions were B3:A0 (no B1:A2 were observed) and three entire chromosomes (2, 8 and 10) were found to be B3:A0. The observed larger contribution of *M. balbisiana* to 'Kluai Tiparot' could be explained by a backcross of an AB hybrid that produced haploid gametes with a normal *M. balbisiana* gamete. In the derived diploid genotype (backcrossed hybrid), the contribution of the B genome is expected to be ¾ of the genome. Consequently, an unreduced gamete from this backcrossed hybrid combined with a normal *M. balbisiana* gamete is expected to have a higher *M. balbisiana* contribution that an allotriploid whose unreduced gamete was provided by a non-backcrossed AB hybrid.

A similar explanation can be proposed for the cultivars included in group 4 ('Pelipita' (ITC0396 and ITC0472) and 'Pelipita Majoncho' (ITC0397)) whose entire chromosomes 2 and 11 have a B3:A0 structure and chromosome 7 is mainly B3:A0 with the exception of a long arm interstitial region. However, in these patterns, two regions in chromosome 9 (interstitial in the short arm and terminal in the long arm) have a B1:A2 structure.

Revisiting subgroups

Genome wide surveys highlighted some inconsistencies in composition of subgroups used to classify ABB cultivars or on the attribution of cultivars to the subgroups. For example, the genotypes representing Monthan and Bontha subgroups share their deviating patterns with genotypes included in Bluggoe subgroups (Genomic group 1a). Moreover, two cultivars assigned to Pisang Awak subgroups ('Bankel' and 'Boddida Bukkisa') had the genomic group 1a deviating pattern, whereas, other cultivars classified as Pisang Awak share the genomic group 2 or 3 patterns. Another cultivar classified in the Bontha subgroup showed the genomic group 2a pattern as well as 'Blue Java' (classed in subgroup Ney Mannan) and 'Simili Radjah' (subgroup Peyan). On the contrary, all the three accessions belonging to the subgroup 'Pelipita' shared an identical and specific deviating pattern (group 4).

These patterns should be used to update the classification of ABB allotriploids. In fact, the genome structure has not only an evolutionary interest, but also can significantly influence the phenotype. Even if somaclonal selection has emphasized the most visible traits, a common clonal origin implies higher similarity on alleles and gene regulation.

Aneuploid accessions

Almost all the surveyed accessions were euploid (entirely triploid). However, in specific genotypes, aneuploidy was detected involving the disappearing of an entire chromosome, chromosome arm or chromosome portion, suggesting the occurrence of chromosome loss or partial deletion. Few regions were also detected with supernumerary regions, likely resulting from duplication of chromosome portions. Since these aneuploidy were individuated in single genotypes (not shared with cultivars having the identical deviating pattern), we suggest that they originated from mutation events more recently than the original triploidization.

Conclusion and perspectives

The results of this study demonstrate multiple origins of ABB cultivars with different and sometime complex routes. It is possible that extension of RAD-Seq genome surveys to other cultivars will allow the discovery of new homoeologous exchanges (HE) patterns, both in AB hybrids and in allotriploids. The frequent occurrence of HEs indicate that A and B subgenomes are prone to recombine, making the *M. balbisiana* genome a source of potential useful variability to create new cultivars able to answer the numerous challenges of banana breeding, such as abiotic stress tolerance associated with climate change.

Phenotypic variability in the triploid cultivars is mainly attributed to somatic and epigenetic variations. However, allopolyploidization, combined with diverse HEs leading to changes in DNA methylation and gene expression, may have contributed to the phenotypic novelty observed in the diverse group of ABBs, and as observed in other crops (He et al. 2017; Lloyd et al. 2018; Li et al. 2019).

Compared with previous analytical methods to assess genetic diversity of the banana genepool, the whole genome survey based on NGS (RAD-Seq in this study) provides a detailed and exhaustive picture of the genome composition. Therefore, we recommend complementing SSR based characterization with NGS technologies such as RAD-Seq as a standard method to characterize banana genebank accessions and to classify them in an objective, early and repeatable way.

References

Baurens F.-C., Martin G., Hervouet C., Salmon F., Yohomé D., et al. (2018) Recombination and Large Structural Variations Shape Interspecific Edible Bananas Genomes? Molecular Biology and Evolution, 36, 97-111. doi: 10.1093/molbev/msy199

Boonruangrod R., Desai D., Fluch S., Berenyi M., Burg K. (2008). Identification of cytoplasmic ancestor gene-pools of Musa acuminata Colla and Musa balbisiana Colla and their hybrids by chloroplast and mitochondrial haplotyping. Theor Appl Genet 118, 43-55. doi: 10.1007/s00122-008-0875-3

Christelová P., Valárik M., Hribová E., De Langhe E., Doležel J. (2011). A multi-gene sequence based phylogeny of the Musaceae (banana) family. BMC Evolutionary Biology 11, 103. doi: 10.1186/1471-2148-11-103

De Langhe E., Hřibová E., Carpentier S., Doležel J., Swennen R. (2010). Did backcrossing contribute to the origin of hybrid edible bananas? Annals of Botany 106, 849-857. doi: 10.1093/aob/mcq187

He Z., Wang L., Harper A.L., Havlickova L., Pradhan A.K., Parkin I.A.P., Bancroft I. (2017). Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnology Journal 15: 594–604. doi: 10.1111/pbi

Li N., Xu C., Zhang A., Lv R., Meng X., Lin X., Gong L., Wendel J.F., Liu B. (2019). DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. The New Phytologist 223, 979-992. doi: 10.1111/nph.15820

Lloyd A., Blary A., Charif D., Charpentier C., Tran J., Balzergue S., Delannoy E., Rigaill G., Jenczewski E. (2018). Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. The New Phytologist 217, 367–377. doi: 10.1111/nph.14836

Risterucci A.M., Grivet L., N'Goran J.A.K., Pieretti I., Flament M.H., Lanaud C. (2000) A high-density linkage map of Theobroma cacao L. Theor Appl Genet 101, 948. doi: 10.1007/s001220051566

Ruas M., Guignon V., Sempere G., Sardos J., Hueber Y., et al. (2017) MGIS: managing banana (Musa spp.) genetic resources information and high-throughput genotyping data. Database 2017, bax046. doi: https://dx.doi.org/10.1093%2Fdatabase%2Fbax046

Simmonds N.W. (1962) The evolution of the bananas. Harlow, Essex. Longmans

Simmonds N.W., Shepherd K. (1955). The taxonomy and origins of the cultivated bananas. Botanical Journal of the Linnean Society, 55, 302-312. doi: 10.1111/j.1095-8339.1955.tb00015.x

Update Report Dec 2018-July 2019

Impact of genomic diversity on transcriptome expression in banana allopolyploid cultivars

Introduction

Phenotypic diversity in banana cultivars is the result of human selection on spontaneous mutations occurred during the evolution of *Musa* spp. Mutations can concern gene in coding sequences or regulatory regions, but also, at a more large scale, involve chromosomes (aneuploidy) or entire genomes (polyploidy).

In the case of polyploidy, genes are present in supernumerary allele copies and new gene interactions can occur in cells and tissues inducing phenotypic diversity. The larger is the diversity among the genomes, the stronger are the expected effects of interaction among the genes. Allopolyploidy , i.e. the presence of subgenomes originating from different species in a polyploid plant, is consequently expected to induce large diversity. An indirect measure of the genomic diversity impact on phenotype can be obtained by the analysis of their transcriptome.

Next generation sequencing (NGS) data allow comparing whole transcriptomes and individuate single genes whose expression is modified according to different conditions (e.g.: different genome structure or in response to stressing treatments).

It was shown that the presence of B genome (provided by *Musa balbisiana*) improves the growth performances under water limiting conditions (van Wesemael et al. 2019). In the precedent chapter it was shown that genotypes supposed to have the same genome constitution could have actually very different genomic structures due to A/B inter-genome recombinations. Since ABB allo-triploids have different homeoalleles doses according to the region where they lie, one can hypothesize that the variability in genome structure has an impact on transcriptomic variability.

In the present study, transcriptomes of cultivars with or without B genome were analyzed to detect differentially expressed genes (DEGs), then some promising genes detected consistently overexpressed under condition of water deficit in allotriploid ABB were described and discussed.

Materials and Methods

Plant material and growth conditions

In vitro banana plants were supplied by the Bioversity International Musa Transit Centre (ITC) hosted at KU Leuven, Belgium (**Table 7**) and were grown for five weeks in autotrophic conditions as described in Zorrilla-Fontanesi et al.(2016) using the following medium: 361 mg/L KNO3, 121 mg/L K2SO4, 176 mg/L MgSO4. 7H2O, 181 mg/L MgCl2.6H2O, 194 mg/L KH2PO4, 398 mg/L NaH2PO4.2H2O, 463 mg/L Ca(NO3)2.4H2O, 105 mg/L CaCl2.2H2O, 60 mg/L Sequestrene, 1.1 mg/L H3BO3, 2.7 mg/L MnSO4.H2O, 0.23 mg/L ZnSO4.7H2O, 0.16 mg/L CuSO4.5H2O, 0.07 mg/L NaMo4.2H2O, pH = 6 (modified from Swennen et al. 1986). Then, at the start of the experiment, the plants were transferred to 1 L PP containers in a Bronson incubator (Bronson Incubator Services B.V., Nieuwkuijk, Netherlands) with a 12h photoperiod, a temperature of 25 °C and a relative humidity of 75 %. The autotrophic growth medium was refreshed on a weekly basis ensuring that the volume of medium never dropped below 55%. The osmotic stress treatment, mimicking a mild drought situation (pF 2.7), was performed by adding 50 g/L PEG (polyethylene glycol 8000, Carl Roth, EG 500-038-2, CAS 25322-68-3) to the growth medium during three weeks to the group of stressed plants while control plants were kept in medium without PEG. Eight biological replicates per genotype and treatment were used. At the end of the experiment, root segments of ~4 cm from the tip were collected, snap frozen in liquid nitrogen and stored at -80 °C for RNA-sequencing.

RNA extraction and cDNA library sequencing

Total RNA was extracted, quantified and its purity assessed as described in Zorrilla-Fontanesi et al. (2016). In total, 60 RNA samples were isolated from 3 biological replicates per genotype and treatment (control vs. stress) (**Table 7**). Prior to cDNA library construction, RNA integrity was checked by ExperionTM (Bio-Rad Laboratories, Inc. USA; RNA Quality Indicator: RQI>8.7 and by Fragment Analyzer (Advanced Analytical Technologies, Inc. USA; RNA Quality Number: RQN>5.8) (**Table 7**). Multiplex sequencing (100 bp single reads) was performed on an Illumina HiSeq 2500 at the Montpellier GenomiX facility as described in Zorrilla-Fontanesi et al. (2016).

Accession Code	Sample Name	Group	Subgroup	Sample Name Abbreviation
ITC0643	Cachaco	ABB	Bluggoe	C
ITC0767	Dole	ABB	Bluggoe	D
ITC0101	Fougamou1	ABB	Pisang Awak	F
ITC0652	Kluai Tiparot	ABB	Klue Teparod	KT
ITC1483	Monthan	ABB	Monthan	Mo
ITC0123	Simili Radjah	ABB	Peyan	SR
ITC1441	Pisang Ceylan	AAB	Mysore	PC
ITC1122	Gros Michel	AAA	Gros Michel	GM

ITC1482	Poyo	AAA	Cavendish	P
ITC0575	Red Dacca	AAA	Red	RD

Table 7: List of banana cultivars selected for the current transcriptomic study including passport data information based on MGIS.

RNA-seq quality validation, read mapping and SNP calling

Between 13.1 and 52.2 million single end reads were obtained per biological replicate. Reads contained in raw FASTQ files (one per sample) were checked using FastQC. Reads were then cleaned to remove adapter sequences and low quality ends (Phred score > 20) with Cutadapt. After trimming, reads inferior to 30 bp were discarded. Reads were then aligned against the Musa acuminata genome of reference v2 (DH Pahang) downloaded on the Banana Genome Hub (http://banana-genome.cirad.fr/)16 using the splice junction mapper for RNA-Seq STAR in a 2-pass mode with the genome annotation file from the reference sequence DH Pahang v2 with the default parameters. Between 67.1 and 90.1% of reads have been uniquely mapped per genotype.

Differential gene expression analyses

RNA-Seq analysis was performed on the ten *Musa* cultivars to compare the root transcriptomes under osmotic treated (stress) and normal (control) conditions. In each case, 3 biological replicates were assessed per genotype and treatment (6 samples per accession). The number of reads in each gene was counted with HTSeq-count using the corresponding gene annotation file and the “union” mode, and also with Salmon. Differential gene expression between stress and control conditions was evaluated using edgeR in the R statistical environment (R Core Team, 2013), where the p-value was adjusted for multiple testing by controlling the false discovery rate (FDR) at ≤5%. Data were normalized using RLE.

Results

Correlation between the presence of B genome and differentially expressed genes

The transcriptome comparison between six ABB (and one AAB) allotriploid and three AAA autotriploid cultivars revealed a variable number of differentially expressed genes (DEGs) correlated to the B genome content. Kluai Tiparot, the ABB cultivar with the higher B genome content, has the higher number of DEGs (1661), whereas the Pisang Ceylan, the unique AAB cultivar, has the lower one (131). The DEGs of the other ABB cultivars spans between 547 and 987 (**Table 8**).

Since the sampled ABB cultivars were shown to have A/B inter-genome recombination generating chromosome regions with variable doses of B homeoalleles, the chromosome distribution of the detected DEGs was analyzed. Regions having higher doses of B genome (B3:A0 in ABB cultivars or B2:A1 in Pisang Ceylan) were found significantly enriched of DEGs (**Table 8**), whereas regions with lower doses of B homeoalleles (B1:A2 in ABB cultivars) appear having lower DEG number (**Table 8**).

	Genome	B3:A0 regions			B1:A2 regions		
Genotype	DEGs	Genes	DEGs	*P* value	Genes	DEGs	*P* value
Cachaco	987 (0.03)	3105 (0.09)	168 (0.17)	3.0^{e-17}	861 (0.03)	12 (0.01)	< 0.01
Dole*	577 (0.02)	3105 (0.10)	111 (0.14)	< 0.01	861 (0.03)	8 (0.01)	< 0.01
Fougamou	823 (0.02)	3483 (0.10)	188 (0.23)	7.1^{e-32}	1226 (0.04)	11 (0.01)	< 0.01
Kluai Tiparot	1661 (0.05)	22,105 (0.66)	1249 (0.75)	5.29^{e-16}	N/A	N/A	N/A
Monthan	558 (0.02)	3105 (0.09)	102 (0.17)	1.11^{e-11}	861 (0.03)	4 (0.01)	< 0.01
Simili Radjah[a]	547 (0.02)	1847 (0.06)	97 (0.16)	1.05^{e-25}	393 (0.01)	1 (0.00)	< 0.05
	Genome	B2:A1 regions			B0:A3 regions		
Genotype	DEG	Genes	DEGs	*P* value	Genes	DEGs	*P* value
Pisang Ceylan	131 (0.004)	3116 (0.09)	28 (0.21)	2.8^{e-06}	803 (0.02)	2 (0.002)	Not significant

Table 8: Enrichment of DEGs in regions with higher content of B genome doses. (Table from Cenci et al. 2019).

Promising DEGs associated to drought stress response in ABB allotriploids

Among the 207 genes detected as differentially expressed in roots under osmotic stress in Cachaco (ABB) but not in two AAA cultivars (Grand Naine and Mbwazirume) (Zorrilla-Fontanesi et al. 2016) a number of genes were detected differentially expressed in an independent experiment on which root transcriptomes of six ABB, one AAB and three AAA cultivars were compared in control condition and under osmotic stress. Most of these genes were found upregulated in Cachaco, Fougamou, Simili Radjah and Red Dacca.

The orthology detection allowed the picking up of a set of genes of potential interest for drought tolerance in banana. First of all, those genes do not have paralogs or paralogs with an antagonist effect in banana genome. Their (co-)orthologues in other plant species were then characterized for gene functions related to abiotic stresses.

With five DEGs identified (Ma01_g21350, Ma06_g23580 and Ma05_g08780; Ma08_g22940 and Ma08_g06080), the polygalacturonase gene superfamily is the most represented gene family. Polygalacturonases are often down regulated genes in plants experiencing water deficit stress (Bray, 2004) and plants overexpressing them are more susceptible to drought (Liu et al., 2014). However, given the complexity of the polygalacturonase superfamily, gene members acquired various functions

in different metabolic pathways. One of these, the D-galacturonic pathway of the L-ascorbic acid, could be promoted in response to the oxidative stress being the L-ascorbic acid a major antioxidant molecule in plants (Venkatesh and Park, 2014). Three *Arabidopsis thaliana* polygalacturonase genes, orthologous to the ones found up-regulated in banana, are predicted to be targeted to mitochondria and all have a similar and specific abiotic stress response pattern, being up-regulated by salt, ABA, hypoxia and osmotic stresses, but not in cold stress (Cao, 2012).

Another candidate gene belongs to the ABC1 kinase (Ma01_g00220), an atypical protein kinase family that is induced by salt stress (Manara et al., 2015). Along with ABC1K8, it was shown to mediate lipid membrane changes in response to stress. In A. thaliana, the abc1k7/abc1k8 double mutant makes the plants highly sensitive to oxidative stress. Protein kinases were already reported as being involved in drought stress tolerance in banana (Muthusamy et al., 2016).

The thiosulfate sulfurtransferase (Ma05_g21270) could be involved as well because one gene falls into a sulfotransferase group where two Oryza sativa members (Os04g0249600 mediated by Apetala transcription factors (AP37 and AP59; and Os02g059600) are up-regulated under drought stress (Oh et al., 2009; Kim et al., 2014).

The 1-aminocyclopropane-1-carboxylate oxidase-like (Ma07_g19730) gene family was also detected up-regulated in stressed banana roots. Although being involved in different functions as ethylene biosynthesis and iron uptake, this gene family shows responses to root abiotic stresses (Gómez-Lim et al., 1993; García et al., 2011; Van de Poel and Van Der Straeten, 2014).

The Outer envelope pore protein (Ma01_g17550) is also an interesting example. Its ortholog in *Lotus japonicus* was observed to be the most up-regulated DEG in drought stressed leaves (Betti et al., 2012). The authors suggested that the up-regulation of OEP16-2 gene could facilitate the throughput of cytosol-synthesized proteins into the chloroplasts and mitochondria.

Finally, we found a drought responsive ATP-binding motif containing protein (Ma01_g17560). This gene contains the domain Usp (Universal stress protein family, pfam00582), that characterizes a gene family responsive to drought (Isokpehi et al., 2011). We observed that the B homeoalleles of Ma01_g17560 is constitutively up-regulated and induces a higher expression in both control and stress samples of 'Cachaco'. The up-regulation is confirmed in 'Fougamou', 'Red Dacca' and 'Simili Radja'.

Discussion

In order to enlarge the adaptability of the cultivation to new geographical areas and to face the challenge of the global climate change, water scarcity tolerance is one of the most important target for banana breeding. New sources of variability need to create new cultivars combining higher adaptability to agronomical desirable traits.

Banana cultivars with genome allotriploid contitution ABB (i.e. containing two chromosome sets derivating from *M. balbisiana* and one from *M. acuminata*) appeared globally performing better under water limiting conditions (van Wesemael et al. 2019). However, wide genome analysis showed that among ABB cultivars the genome diversity is large and mainly generated by A/B inter-genome recombination that generates variable doses of B homeoalleles. This study provide evidences that the genome diversity inside the ABB banana cultivars is reflected in the gene expression and, consequently, in the phenotypic diversity. The deep knowledge of this diversity will increase the possibility of its use in breeding strategies.

On the other hand, the understanding of the physiology in plants exposed to abiotic stresses is required to set strategy for molecular breeding. In this study, genes consistently overexpressed in different ABB cultivars submitted to water deficit were identified. The orthology of some of them to genes already known to be involved in stress response in other plants makes these genes promising candidates for clarifying how the more tolerant plants change their gene expression to face the environmental conditions. The correct establishment of orthology relationships between model and non-model plants is required to take advantage from the already established knowledge in vegetal kingdom.

References

Betti M., Pérez-Delgado C., García-Calderón M., Díaz P., Monza J., and Márquez A. (2012). Cellular Stress Following Water Deprivation in the Model Legume Lotus japonicus. Cells 1, 1089–1106. doi: 10.3390/cells1041089

Bray E.A. (2004). Genes commonly regulated by water-deficit stress in Arabidopsis thaliana. J. Exp. Bot. 55, 2331–2341. doi:10.1093/jxb/erh270

Cao, J. (2012). The Pectin Lyases in Arabidopsis thaliana: Evolution, Selection and Expression Profiles. PLoS ONE 7, e46944. doi:10.1371/journal.pone.0046944

García M. J., Suárez V., Romera F.J., Alcántara E., and Pérez-Vicente R. (2011). A new model involving ethylene, nitric oxide and Fe to explain the regulation of Fe-acquisition genes in Strategy I plants. Plant Physiol. Biochem. 49, 537–544. doi: 10.1016/j.plaphy.2011.01.019

Gómez-Lim M. A., Valdés-López V., Cruz-Hernandez A., and Saucedo-Arias L. J. (1993). Isolation and characterization of a gene involved in ethylene biosynthesis from Arabidopsis thaliana. Gene 134, 217–221. doi: 10.1016/0378-1119(93)90096-L

Kim J. S., Park H.-M., Chae S., Lee T.-H., Hwang D.-J., Oh S.-D., et al. (2014). A Pepper MSRB2 Gene Confers Drought Tolerance in Rice through the Protection of Chloroplast-Targeted Genes. PLoS ONE 9, e90588. doi: 10.1371/journal.pone.0090588

Isokpehi R.D., Simmons S.S., Cohly H.H.P., Ekunwe S.I.N., Begonia G.B., and Ayensu W.K. (2011). Identification of Drought-Responsive Universal Stress Proteins in Viridiplantae. Bioinforma. Biol. Insights 5, BBI.S6061. doi: 10.4137/BBI.S6061.

Liu H., Ma Y., Chen N., Guo S., Liu H., Guo X., et al. (2014). Overexpression of stress-inducible OsBURP16, the β subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice: OsBURP16 overexpression in rice. Plant Cell Environ. 37, 1144–1158. doi:10.1111/pce.12223

Manara A., DalCorso G., Guzzo F., Furini A. (2015). Loss of the Atypical Kinases ABC1K7 and ABC1K8 Changes the Lipid Composition of the Chloroplast Membrane. Plant Cell Physiol. 56, 1193–1204. doi:10.1093/pcp/pcv046

Muthusamy M., Uma S., Backiyarani S., Saraswathi M. S., and Chandrasekar A. (2016). Transcriptomic Changes of Drought-Tolerant and Sensitive Banana Cultivars Exposed to Drought Stress. Front. Plant Sci. 7, 1609. doi: 10.3389/fpls.2016.01609

Oh S.-J., Kim Y. S., Kwon C.-W., Park H. K., Jeong J. S., and Kim J.-K. (2009). Overexpression of the Transcription Factor AP37 in Rice Improves Grain Yield under Drought Conditions. PLANT Physiol. 150, 1368–1379. doi:10.1104/pp.109.137554

Swennen R., De Langhe E., Janssen J., and Decoene D. (1986). Study of the root development of some Musa cultivars in hydroponics. Fruits 41, 515–524.

Van de Poel B., and Van Der Straeten D. (2014). 1-aminocyclopropane-1-carboxylic acid (ACC) in plants: more than just the precursor of ethylene! Front. Plant Sci. 5, 640. doi: 10.3389/fpls.2014.00640

Van Wesemael J., Kissel E., Eyland D, Lawson T, Swennen R., Carpentier S. (2019). Using Growth and Transpiration Phenotyping Under Controlled Conditions to Select Water Efficient Banana Genotypes. Front. Plant Sci. 10, 352. doi: 10.1111/j.1365-313X.2005.02593.x

Venkatesh J., and Park S. (2014). Role of L-ascorbate in alleviating abiotic stresses in crop plants. Bot. Stud. 55, 38. doi:10.1186/1999-3110-55-38

Zorrilla-Fontanesi Y., Rouard M., Cenci A., Kissel E., Do H., Dubois E., Nidelet S., Roux N., Swennen R., Carpentier S.C. (2016). Differential root transcriptomics in a polyploid non-model crop: the importance of respiration during osmotic stress. Sci Rep. 6, 22583. doi: 10.1038/srep22583

Pilot study to select alleles related to drought tolerance

Context of the study

In banana, most of the edible cultivars are hybrids from one or both major wild diploid ancestors: *Musa acuminata* (A genome, 2n=*2x*=22) and *Musa balbisiana* (B genome, 2n=*2x*=22). These cultivars have been assigned to different genomic groups according to the number of chromosomes sets and the species that donated them, i.e. AAA, AAB and ABB (2n=3x = 33). Variability in genome contribution to triploid clones makes it pertinent to investigate the differential gene expression of *Musa* cultivars, in particular since genotypes with the B genome are more tolerant to abiotic stresses than those containing only the A genome.

Due to the increasing power of high-throughput genotyping and transcriptomic methods, associated with the development of genetic and statistical analysis tools, significant gains have been generated for diploid species, (Desta and Ortiz 2014), but on polyploids species it is still challenging (Comai 2005; Grandke et al. 2016). Polyploidy can affect the phenotype, creating complex interactions between loci or alleles such as dominance or epistasis, or through allelic dosage. This represents an additive effect of multiple copies of the same allele. Collecting information on allelic dosage allows researchers to describe a more realistic representation of the effect of each genotype.

Single Nucleotide Polymorphisms (SNPs) are mostly used in allelic dosage because of their high abundance. A SNP database developed under normal and drought stress conditions has provided an opportunity for such allele detection and to study the nonsynonymous SNPs, or substitutions (altering the amino acid sequence of a protein) and synonymous SNPs (which do not affect amino acid sequence). The nonsynonymous and synonymous substitution rates (Ka/Ks) have showed the direction of natural selection and these nonsynonymous substitutions could affect the response to a stress such as drought. The prediction of both positive and negative evolutionary selection pressures on genes could provide a detailed picture of evolutionary selection pressure linked to the allele dosage.

The present study pilot describes the methodology to analyse and select alleles, SNP alleles in two conditions stressed by drought and control genotype on triploid *Musa acuminata* AAA, ABB, AAB genotypes, in order to detect an eventual correlation with the phenotype and list of candidate alleles. Here we present the description of the methodology to analyse data and report preliminary results.

Bioinformatic pipeline development

To identify alleles related to drought tolerance, we used vcf (variant call format) files obtained in the "Impact of genomic diversity on transcriptome expression in banana allopolyploid cultivars" for the differential analyses with RNAseq data. The material and methods are the same. Each position represents a variant along the genome. We have built a bio-informatic pipeline to select the SNP position corresponding to an allele coming from the A or B genome under positive or negative selection. This pipeline is divided in four steps: 1) apply several filtered to clean the vcf files, 2) count the frequency of the alleles, 3) calculate the Ka/Ks ratio to determine the direction of gene selection, and 4) in order to refine the annotation of each gene, use software such as PANTHER and SIFT to detect deleterious alleles.

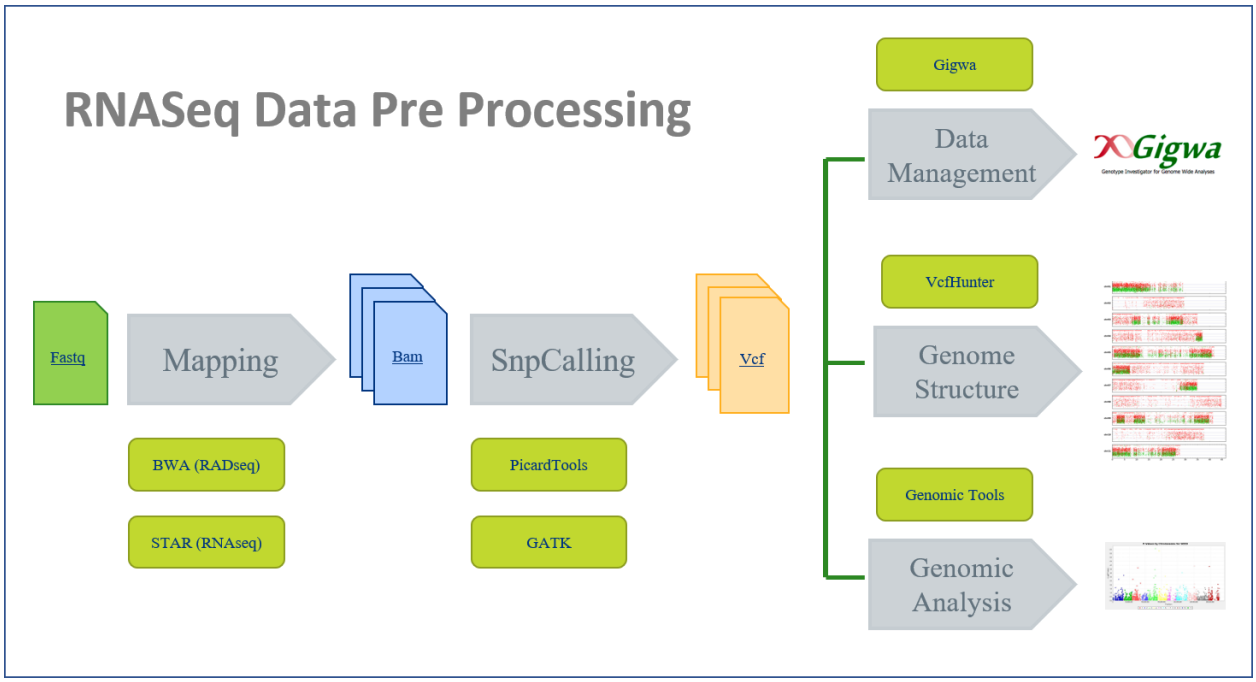


Figure 18: The pipeline used to analyse RNAseq dataset, for each library, then merged for each banana cultivars was divided into three parts. The first step was to map the reads on the *acuminata* reference with the software STAR which detected the splicing events. Bam files were generated, and GATK was used to decipher the SNP. The third part of the pipeline represents the data management with Gigwa, and the genome analysis is described below.

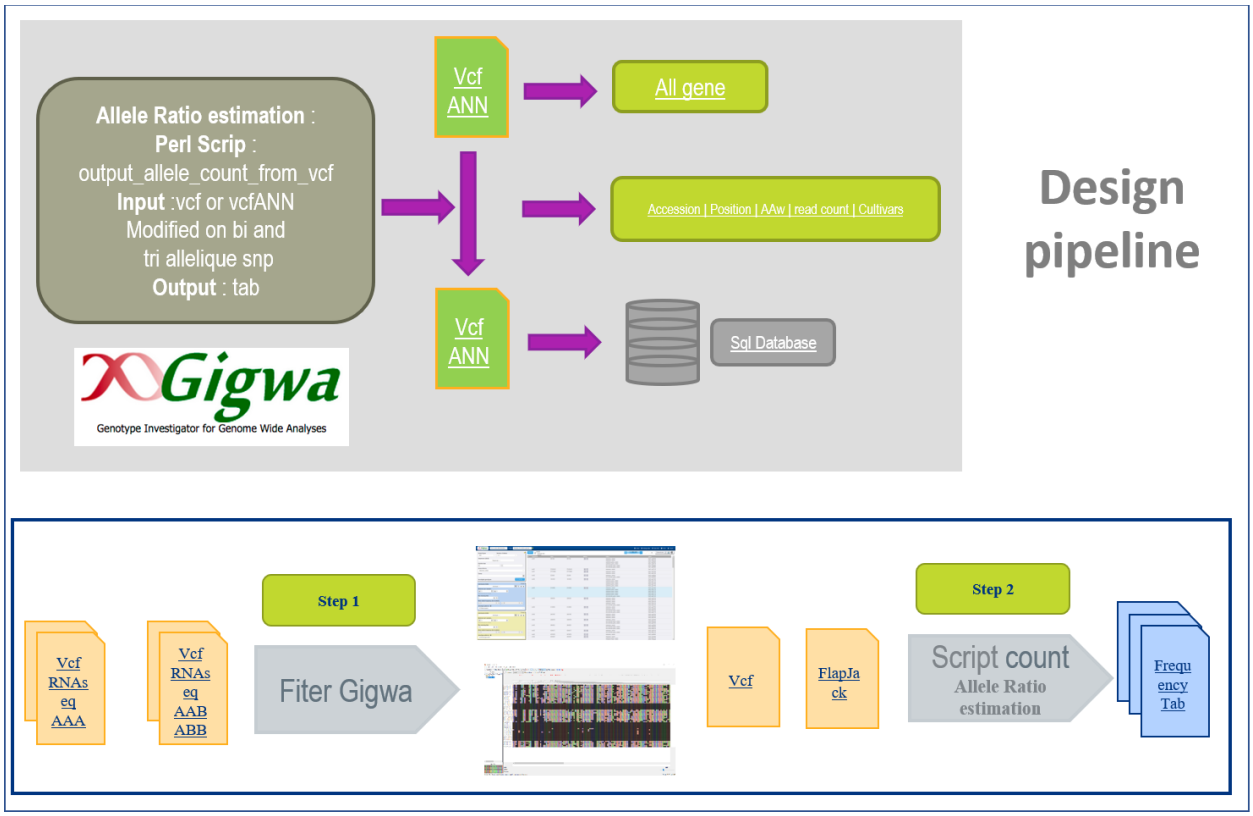


Figure 19: In order to select SNPs on the basis on the allele dosage, the first step based on 4 filters (0 missing data, missense-variant which represent the non-synonymous, DP 10 and QG 20) is done with the software Gigwa (step 1)

Step 1: Filter vcf file

In order to select alleles, several filters were applied on the vcf previously obtained for each repeat of the different RNAseq samples libraries. The vcf file was annotated with the tools SnpEff with the *acuminata* genome as reference. This step could also be done in a global vcf file for each genotype in order to increase the depth for the allele selection. In this case, the SNP calling described in the RNAseq part has to be done with the exact same parameters. This vcf could be uploaded in the Gigwa interface, which is a management SNP database software, developed to facilitate the selection of the SNP according to different quality and annotation criteria.

Main filters applied on the vcf were the nonmissing value, **depth of coverage** (DP> 10) and Genotype Quality (GQ>=20), which were default parameters. **DP** is the filtered depth, at the sample level. This gave the number of filtered reads that support each of the reported alleles. The Genotype Quality represents the Phred-scaled confidence that the genotype assignment (GT) is correct, derived from the genotype Phred-scaled likelihoods (PLs). **PL is "Normalized"** Phred-scaled likelihoods **of the possible genotypes**. Specifically, the GQ is the difference between the PL of the second most likely genotype, and the PL of the most likely genotype. All those filters were applied in the software Gigwa developed to manage data from vcf annotated files. A vcf file was obtained (Figure 19).

Step 2: Allele count

For each genotype present on the vcf filtered file, the perl script “output_allele_count_from_vcf.pl” gave us the frequency of each allele, the position of the SNP on the genome and the gene annotation on a tabular table. The verification of the genome origin of allele, either *acuminata* or *balbisiana,* will be done with the script VcfHunter (Baurens et al.), then the RNAseq reference for A and B genomeare under construction (Figure 19).

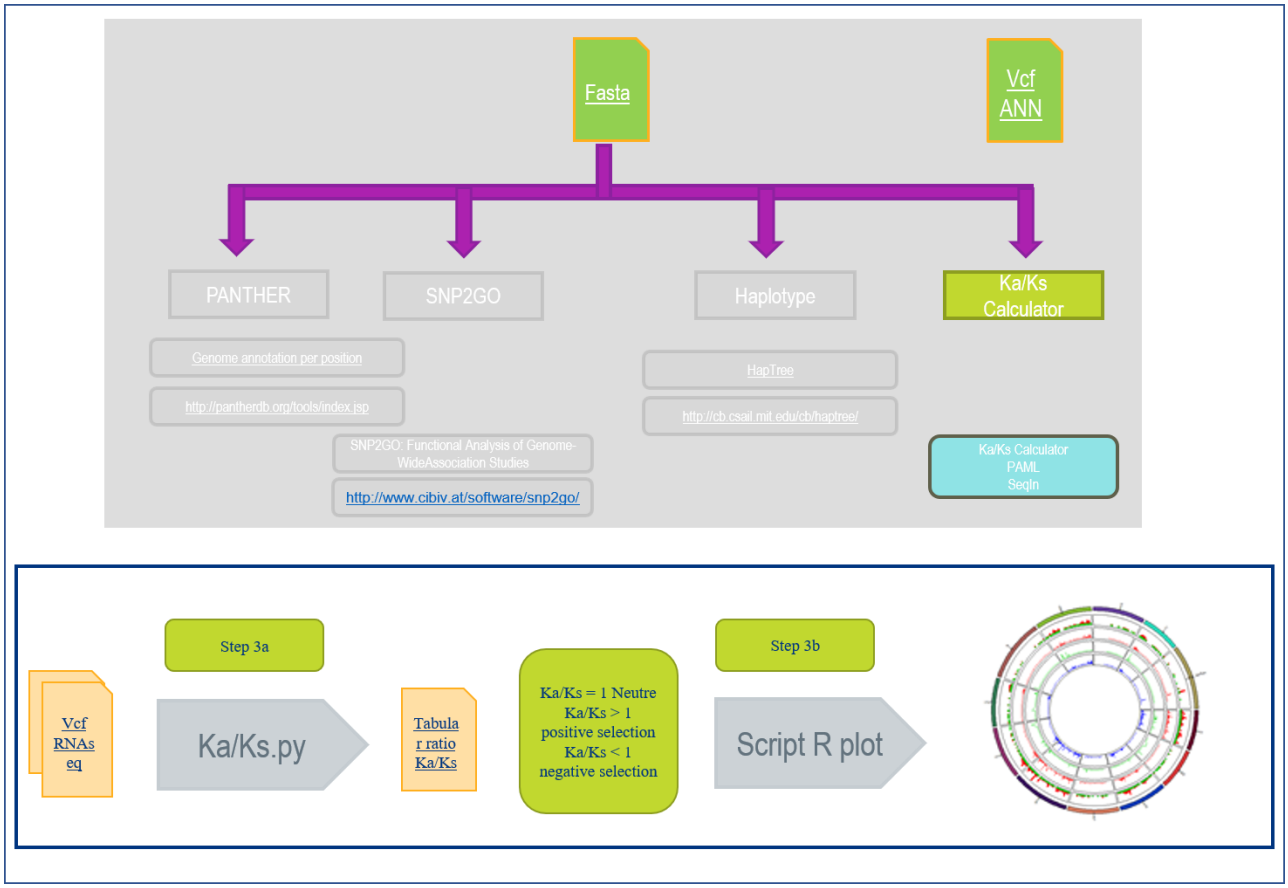


Figure 20: A vcf and a flapJack file were exported and used to calculate the ratio Ka/Ks and the allele frequency count for each SNP. This tabular was used with R Circlize package to represent the circle ka/ks distribution along the genome.

Step 3: Ka/Ks ratio and graph representation, classification of purified genes during selection pressure of SNPs.

For each vcf file generated and filtered the ratio Ka/Ks was calculated according to the pipeline in Figure 20. It was used to estimate the balance between neutral mutations, purifying selection and beneficial mutations acting on a set of homologous protein-coding genes. It was calculated as the ratio of the number of nonsynonymous substitutions per nonsynonymous site (K_a), in a given period of time, to the number of synonymous substitutions per synonymous site (K_s), in the same period. The latter were assumed to be neutral, so that the ratio indicated the net balance between deleterious and beneficial mutations. Values of K_a/K_s significantly above 1 are unlikely to occur without at least some of the mutations being advantageous. If beneficial mutations are assumed to make little contribution, then K_s estimates the degree of evolutionary constraint.

In order to select SNPs under selection, the ratio Ka/Ks, was calculated with a python script directly on the vcf file. Then the distribution of the ratio per SNP along the genome was represented on a circle graph made with the package R Circlize. The two groups of SNPs are planned to be analyzed in the fourth step.

Step 4: Panther and SIFT Annotation, distinction of tolerable SNPs from Deleterious SNPs

In order to add another layer to the refinement SNP as the potential functional effect of amino acid substitution on corresponding proteins, SIFT (Sorting Intolerant from Tolerant) and PANTHER (Protein Analysis Through Evolutionary Relationship) are planned to be used to distinguish the tolerable SNPs from the deleterious SNPs to find the deleterious allele.

Results

To test the pipeline, the SNP selection was done on the global vcf containing all samples analyzed in the RNAseq study. Three libraries for the stressed condition, and three libraries for the control condition were done for each sample, then the selection was done on a total of 60 libraries.

In order to select alleles coming from the A genome, the selection process through Gigwa was done with the filters, 0 missing data, missense-Variant (corresponding to the nonsynonimous) on the group samples AAA heterozygote and AAB-ABB homozygote. A total of 445 SNPs was selected and exported in vcf format in order to use the KaKs.py script to calculate the Ka/Ks ratio. On this tabular file a custom R script to visualize the distribution along the genome was applied (Figure 21). Then we selected the SNP with a Ka/Ks ratio > 1 for the position under a positive selection.

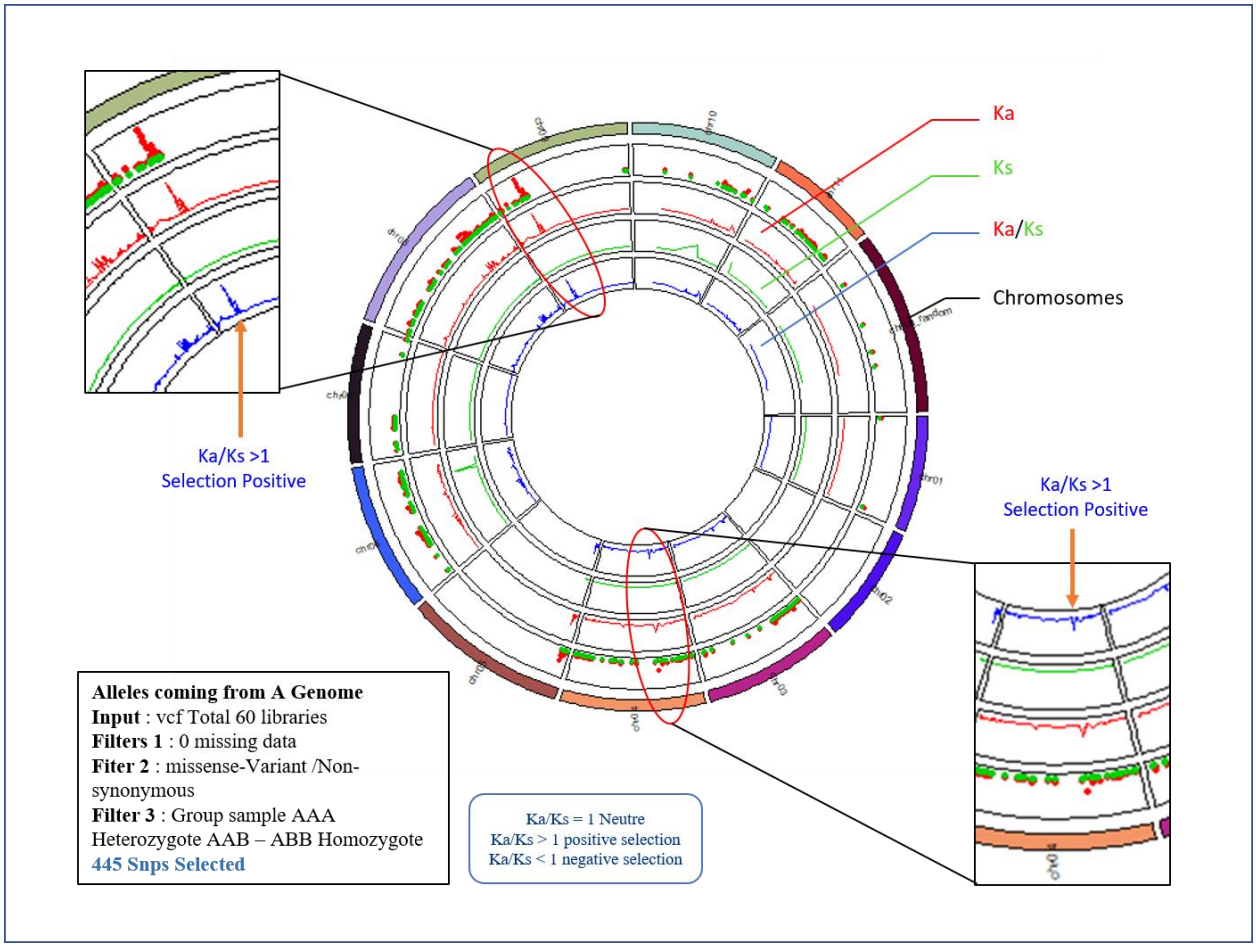


Figure 21: Distribution of the Ka, Ks, and Ka/Ks ratio along the chromosome for a selected pool of alleles coming from the A genome. This circle represents the distribution of each SNP filtered according the parameters for all SNP positions along each chromosome. Each track represents a datum. The first track represents Ka/Ks points. The second track represents Ka for each SNP selected, the third represents Ks for each position and the fourth third the Ka/Ks ratio.

In order to select alleles coming from B genome, the selection process through Gigwa was done with the filters, 0 missing data, missense-Variant (corresponding to the nonsynonimous) on the group samples AAA homozygote and AAB-ABB heterozygote. A total of 2812 SNPs was selected and exported in vcf format in order to use the KaKs.py script to calculate the Ka/Ks ratio. On this tabular file a custom R script to visualize the distribution along the genome was applied (Figure 22). Then we select the SNP with a Ka/Ks ratio > 1 for the position under a positive selection.

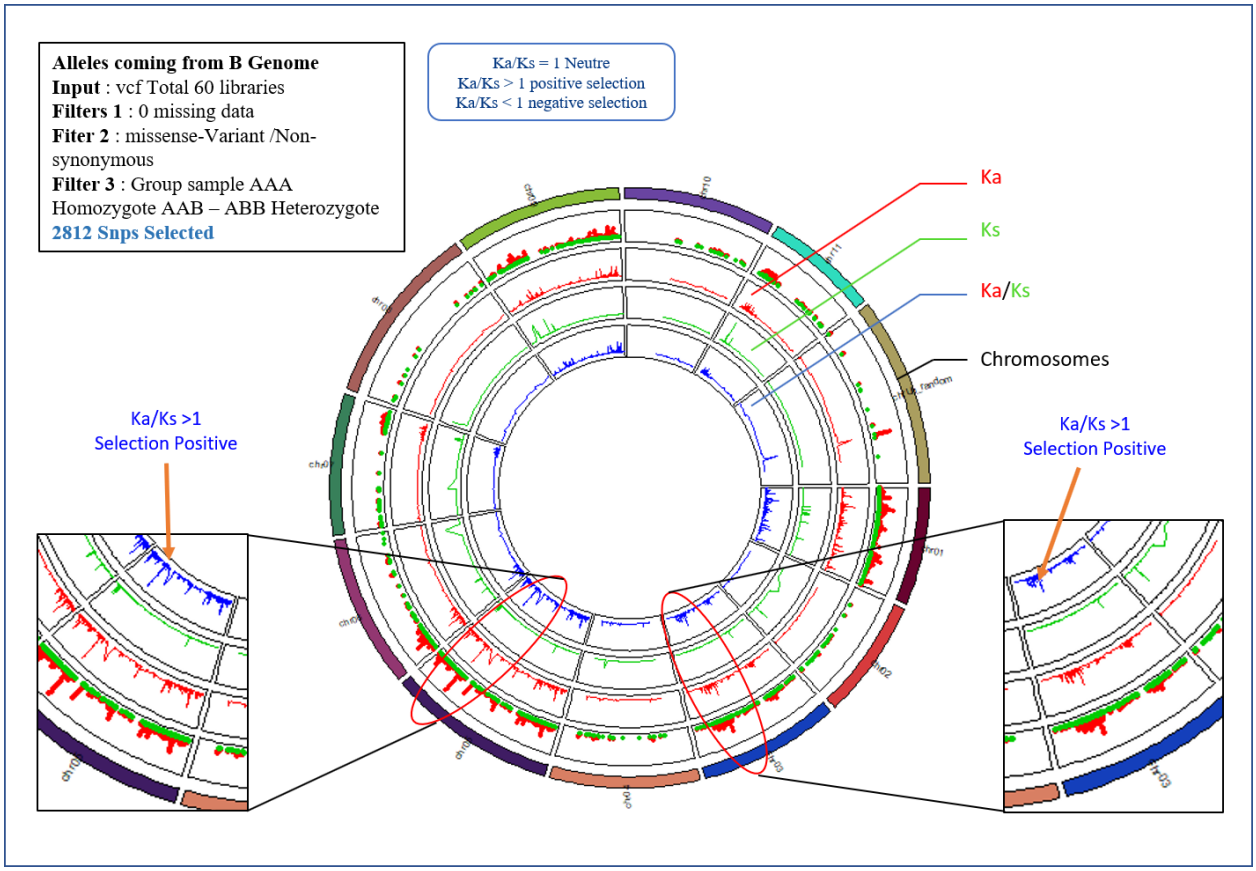


Figure 22: Distribution of the Ka, Ks, and Ka/Ks ratio along the chromosome for a selected pool of alleles coming from the B genome. This circle represents the distribution of each SNP filtered according the parameters for all SNP positions along each chromosome. Each track represents a datum. The first track represents Ka Ks points, the second track represents Ka for each SNP selected, and the third represents Ks for each position and the fourth third the Ka/Ks ration.

Conclusions and perspectives:

To select alleles related to drought tolerance, this pipeline will be applied on the global vcf file containing all the different cultivars AAA, AAB and ABB. Firstly, the selection will be done for the two groups and then a comparison will be done for the two conditions, control and stressed, to obtain a list of candidate alleles. Furthermore, some other tests, such as Test of Tajima, will be performed to confirm the Ka/Ks ratio. Then a cluster of genes will be performed to reveal the deleterious and tolerable status with two annotation tools PANTHER and SIFT. This list of alleles corresponding to different genes will be compared with the proteomic and transcriptomic results to confirm the results and to discover new genes involved in drought tolerance and resistance pathways.

References

Anders S., Pyl P.T., Huber W. (2014) *HTSeq — A Python framework to work with high-throughput sequencing data.* Bioinformatics 31,166-169 doi:10.1093/bioinformatics/btu638

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Baurens F.-C., Martin G., Hervouet C., Salmon F., Yohomé D., et al. (2018) Recombination and Large Structural Variations Shape Interspecific Edible Bananas Genomes? Molecular Biology and Evolution, msy199. doi: 10.1093/molbev/msy199

Comai L. (2005) The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6, 836–846. doi: 10.1038/nrg1711

Desta Z. A., Ortiz R. (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 19, 592–601. doi:10.1016/j.tplants.2014.05.006

Dufresne F., Stift M., Vergilino R., Mable B. K. (2014) Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Mol. Ecol. 23, 40–69. doi:10.1111/mec.12581

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21. doi: 10.1093/bioinformatics/bts635

Grandke F., Singh P., Heuven H. C., De Haan J.R., Metzler D. (2016) Advantages of continuous genotype values over genotype classes for GWAS in higher polyploids: a comparative study in hexaploid chrysanthemum. BMC Genomics 17, 672. doi:10.1186/s12864-016-2926-5

Gu Z. (2014) circlize implements and enhances circular visualization in R. Bioinformatics 30, 2811–2812. doi:10.1093/bioinformatics/btu393

Hieter P., Griffiths T. (1999) Polyploidy–more is more or less. Science 285, 210–211. doi:10.1126/science.285.5425.210

Kumar P., Henikoff P., Ng P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols 4, 1073–1081. doi: 10.1038/nprot.2009.86

Martin M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10-12. doi: 10.14806/ej.17.1.200.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A., (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20, 1297-303

Mi H., Muruganujan A., Ebert D., Huang X., Thomas P.D. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucl. Acids Res. 47, D419–D426 doi: 10.1093/nar/gky1038

Patro R., Duggal G., Love M.I., Irizarry R.A., Kingsford C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 14, 417-419. doi: 10.1038/nmeth.4197

Robinson M.D., McCarthy D.J., Smyth G.K. (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics* 26, 139-140. doi: 10.1093/bioinformatics/btp616.

Sempéré G., Philippe F., Dereeper A., Ruiz M., Sarah G., Larmande P. (2016) Gigwa-Genotype investigator for genome-wide analyses. Gigascience5, 25. doi: 10.1186/s13742-016-0131-8

Slater A.T., Cogan N.O., Forster J.W., Hayes B.J., Daetwyler H.D. (2016) Improving genetic gain with genomic selection in autotetraploid potato. Plant Genome 9. doi:10.3835/plantgenome2016.02.0021

Spoelhof J. P., Soltis P. S., Soltis D. E. (2017) Pure polyploidy: closing the gaps in autopolyploid research. J. Syst. Evol. 55, 340–352. doi:10.1111/jse.12253

Thomas P.D. Kejariwal A., Campbell M.J., Mi H., Diemer K., Guo N., Ladunga I., Ulitsky-Lazareva B., Muruganujan A., Rabkin S., Vandergriff J.A., Doremieux O. (2003) PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 31, 334–341. doi: 10.1093/nar/gkg115