POSITION PAPER

February 2017

Epigenetics in plant breeding

The "Groupement d'Intérêt Scientifique Biotechnologies Vertes" (GIS BV), the "Animation Scientifique Biotechnologies Végétales" of the INRA Division "Plant Biology and Breeding" and the "Plant Biology and Biotechnologies" working group (GT4) of the AllEnvi Alliance organized a scientific workshop on November 9, 2015 on the topic of "Taking into account epigenetic processes for plant breeding" in Paris (France). Eighty-eight scientists from the public and private members along with scientists from non-members of the GIS BV gathered at this event. The workshop was organized in two sessions of presentations followed by a global discussion. The first session was dedicated to epigenetic processes and their potential consequences in plant breeding. The second one presented research projects applied to plants of agricultural interest. This paper summarizes the strategies identified and addresses the research opportunities for the French public-private community, in light of the scientific and socio-economic issues regarding the understanding of epigenetic processes in plants and their potential interest in plant breeding.

Scientific background

Conventional plant breeding relies on the transfer of traits of interest, sometimes identified in wild species, to cultivated crops by crossing and recurrent selection. Although not required, the elucidation of the molecular support of such agronomical traits greatly facilitates this process. Until the end of the 20th century, it was thought that isolating the DNA sequences associated with a trait of interest in the plant of origin was sufficient to accurately transfer that trait to crops and to confer the expected phenotype. Since then, it has been demonstrated that DNA provides only part of the information supporting a trait, and that chromatin also contributes. Chromatin is indeed the natural substrate on which DNA transactions take place within the nucleus of eukaryotic cells. Its basic unit is the nucleosome, which contains approximately 150 bp of DNA wrapped around a core of eight histone proteins (2H2A, 2H2B, 2H3 and 2H4). DNA can be methylated, histones can be subjected to hundreds of post-translational modifications (e.g. acetylation, methylation, phosphorylation, etc.) and numerous histone variants also exist. These modifications and variants, together with linker histone H1 as well as non-histone proteins and regulatory RNAs (which do not encode proteins and thus are referred to as non-coding RNAs or ncRNAs), define distinct chromatin/epigenetic states along the genome (the so-called epigenome) as well as between cell types or in response to environmental cues. Thus, chromatin is a highly dynamic structure, which carries variable information in addition to that provided by the (essentially) constant DNA sequence. Nonetheless, chromatin is conditioned in part by the DNA sequence at both local and global levels. Notably, intergenic sequences vary greatly between wild and cultivated species, and these sequences can influence the establishment or the stability of chromatin states over adjacent or distant genes. Moreover, the cellular machinery that produces histones or modifies DNA and histones can also vary between wild and cultivated species. Similarly, ncRNAs are not always conserved between wild and cultivated species. Thus, epigenomes can potentially vary extensively between related species. Consequently, the transfer of a trait of interest from wild to cultivated species not only requires the transfer of the DNA associated with this trait, but also the establishment of the proper chromatin/epigenetic state(s) over this piece of DNA so as to enable the trait to be expressed as in the plant of origin. It is therefore essential to precisely characterize these chromatin state(s) in the plant of origin and to ensure that they are properly re-established in the recipient crop. For example, if a trait of interest identified in a wild species is determined by a gene, the promoter of which is associated with an ncRNA that is absent or not conserved in the recipient crop, it will be necessary to transfer both the gene and the locus

producing the ncRNA. Similarly, if a trait of interest identified in a wild species is determined by a gene the expression of which is somewhat dependent on the epigenetic state of an adjacent transposable element (TE) sequence, both the gene and the adjacent TE will need to be transferred to the recipient crop.

Challenges

The first challenge is to clearly define the basis of an epigenetic state. Whereas a DNA sequence is simply defined by the order of the four bases (A, C, G and T), the exhaustive list of components that define a given chromatin or epigenetic state is yet to be established. These components include: methylation of cytosines and adenines, mono-, di- or tri-methylation, acetylation, phosphorylation, ubiquitination, etc of histones at various positions (e.g.H2AK119, H3K4, H3K9, H3K27, H3K36, etc), and long or short ncRNAs produced in cis or in trans.

A second important challenge is to define how stable is a given chromatin state. Three major levels of stability can be distinguished:

- Transient chromatin states

These types of chromatin states are specific to different cell types or established in immediate response to biotic or abiotic stress, and do not persist after the stimulus is removed.

- Metastable epigenetic states

These types of epigenetic states are induced by particular stress or environmental cues and can persist across multiple cell divisions after induction.

- Inherited epigenetic states

These types of epigenetic states are transmitted across multiple generations and are typically associated with TEs or other repeat sequences. It is still unclear what role the environment plays in inducing or erasing these states.

Specific targets

Increasing epigenetic variability either genome wide or at specific loci

Recent studies on epigenetic recombinant inbred lines in Arabidopsis have demonstrated that epigenetic quantitative trait loci can explain the heritability of complex traits. In crops the number of examples of epigenetically controlled traits is increasing steadily. Recent examples include fruit ripening in tomato and somaclonal variation in oil palm. As epigenetic regulation of gene expression can influence important crop traits, the creation of stably inherited epigenetic diversity could be a very powerful tool in plant breeding. Basically, it enables the modification of traits in crop plants without changing the DNA sequence and it is therefore likely to be most valuable for the further improvement of high performance lines.

Most known epigenetic variants are associated with loss of DNA methylation and correspond to gain of function variants. Although epigenetic marks have so far not been linked to phenotypic variability in a population, this does not necessarily mean that they do not play a role but that other modifications can be more challenging to detect.

There are several possible routes to induce epigenetic variability in plants:

- Regarding drug treatments, there are several well-characterized inhibitors of proteins involved in epigenetic silencing. Treatment of seedlings with DNA methyltransferase inhibitors (e.g. 5-aza-2'-deoxycytidine or zebularine) lead to a global reduction of DNA methylation. However, this reduction appears to be transient only as it is not transmitted to the progeny. Trichostatin A (TSA) is a drug that can inhibit histone deacetylases. Histone deacetylases actively remove acetylation marks on histones and contribute to gene repression. Thus TSA treatment classically results in the reactivation of silenced genes, but here again the effect appears to be transient only.
- A more reliable and demonstrated approach to generate heritable epigenetic variation is through the disruption of genes encoding key epigenetic regulators. Indeed, numerous heritable epialleles have been produced in Arabidopsis plants with mutations in either of two genes (*met1* and *ddm1*) involved in the maintenance of DNA methylation. Here, the problem is to obtain knock-outs in the corresponding genes in crops.
- It has been reported that abiotic and biotic stresses can induce epigenetic changes in plants. While it is currently difficult to control these variations the mobilization of epigenetically silenced and stress-responsive TEs could contribute to stable inheritance of

stress-induced genetic and epigenetic changes from one generation to another.
- Locus-specific DNA methylation and histone modification changes can now be induced using the modified CRISPR-Cas9 technology. In this case a Cas9 protein with an inactivated nuclease domain is fused to an enzyme that modifies DNA methylation or histone modification. This will result in targeted epigenetic changes at the sites of interest.
- Looking for genetic variations in epigenetic regulator genes using populations (natural and breeding) can also help to identify epigenetic variants. These individuals can represent an interesting genetic background to use for crossing and generate new phenotypes in the progeny. Such an approach has so far not been carried out and remains to be tested.

In conclusion, despite evidence that epigenetic variation can affect traits of agronomic importance, the number of reports in crops is still limited to a few species and the stability of this type of variation across generations and environments requires further assessment.

Understanding the origin of phenotypic plasticity and instability

Phenotypic plasticity, the ability of a single genotype to express multiple phenotypes in response to different environments, either external or internal, is widespread amongst both plants and animals. While this plasticity is generally considered to be adaptive and/or advantageous for sessile organisms that have to adapt in place to environmental conditions, it can also represent a major limitation in crop breeding especially when it comes to developmental instability. This instability can be seen as the production of randomly variable phenotypic offspring as well as the within environment phenotypic variance for a given genotype that might result in off-type plants, *i.e.* plants showing a distinct phenotype from the sown cultivar. Improving phenotypic stability is pivotal for coping with the detrimental impacts of climate change and maintaining crop production to feed a growing world population. It is also crucial from a breeder's perspective when considering that phenotypic stability, together with distinctness and uniformity, is a statutory requirement for a variety to be registered.

In this context, understanding the bases of this phenotypic plasticity / instability is crucial for crop breeding. Nevertheless this appears to be a daunting task as the theme encompasses a wide range of cases. Literature on phenotypic plasticity has increased expanding from the initial focus on abiotic factors to that of biotic ones and, in recent years, taking into consideration plant response to global climate change. There is a growing body of evidence that environmental conditions and stress can induce changes at the morphological and physiological levels. However, the rapid establishment of new phenotypes in response to stimuli cannot be explained by genetic mutations only because of the rather low mutation rate. Instead, it has been proposed that plastic changes prefigure genetic ones and that these processes involve molecular epigenetic mechanisms. To disentangle the role of genetics, epigenetics and environment and putatively utilize phenotypic plasticity / instability in crop breeding programs, several questions should be addressed. Below are some examples of research topics in the field of phenotypic plasticity:

- What is the stimulus / stress triggering the phenotypic change?
  Studying the environmental factors triggering phenotypic changes requires the comparison of phenovariants, *i.e.* individuals of the same genotype showing different phenotypes when submitted to different stimuli. This can be done either in controlled conditions or in the field. More challenging is the deciphering of the origin of phenotypic instability in the absence of external stimulus. This is the case for randomly variable phenotypic offspring and off-type plants, where phenovariants appear in similar environmental conditions.
- What is the plant response to stress / stimulus?
- The challenge of organisms is to process environmental information and respond appropriately. Studying responses might help to decipher the molecular basis of plasticity and potentially identify "plasticity genes" as well as traits that are the most likely to show adaptive plasticity.
- Are these effects stable and transmitted across multiple generations?
  Depending on the trait, and when in the lifecycle and for how long the environmental exposure occurs, environmentally induced changes may or may not be reversible. As a

consequence, some effects can show transgenerational heritability, either hard for genetically determined changes or soft for epigenetically determined ones.

- How can phenotypic instability be exploited in crop breeding?
  Plasticity might be exploited to breed for more resilient crops, especially in the context of climate change. For example, priming / hardening has the potential to produce varieties that can maintain production under stressful conditions. However, our ability to harness plant plasticity requires a better understanding of the underlying epigenetic mechanisms.

Epigenetic stress memory for crop improvement

One way for plants to adapt to environmental stress is to develop the ability to "remember" a stress episode and to react more efficiently (faster and more strongly) upon subsequent exposures to stress, that is, with increased resistance. Two types of "memories" seem to co-exist: first, at the level of a population, multiple lines of evidence point towards a trans-generational transmission of the stress memory to the progeny; second, at the level of the plant life-time, short-term memory (also known as "priming"), persisting for days, can also increase the potential of plants to respond to environmental challenges. At the molecular level, this short-term memory could result from a combination of different mechanisms, including modification of the levels of stress-associated receptors, signalling components and transcription factors. However, it has also become clear that both short-term and trans-generational memories largely rely on epigenetic modifications. In the perspective of developing new agricultural practices, exploiting the trans-generational stress memory would be the best option. However, deeper fundamental investigations are still required to clarify whether stress-induced epialleles can be stabilized over several generations and consequently used in breeding programs. For this reason, focusing on short-term memory to unravel the relevant epigenetic mechanisms operating during plant lifetime should be a feasible and promising option to deliver crops with better response to abiotic constraints.

Recent studies have shown that plants can adapt to drought – a major abiotic constraint limiting the productivity of several agricultural crops worldwide – using short-term stress memory. In the model plant *Arabidopsis thaliana*, several genes were found to produce higher transcript levels in subsequent drought exposures as compared to transcriptional level after the initial exposure. The transcription of these "trainable" genes is restored to the basic level in between stress exposures. Importantly, this is associated with specific chromatin marks detected around the transcription start sites of these genes. These marks, and consequently the stress-memory, can persist for 5 days but are lost after 7 days. Similar acclimation to drought has also been described in maize. The goal now is to decipher the molecular mechanisms of short-term drought stress memory in other crops and for other abiotic constraints, which could be manipulated in the future to improve stress tolerance.

Future investment and research targets:
To reach this goal, the following points will need to be addressed:

- Set up experimental designs to investigate stress memory in crops, identify stress trainable genes and determine the chromatin landscape of these genes and test if TEs are involved in these processes.
- Understanding the impact of the environment on epigenetic states. What kinds of natural environmental constraints induce chromatin changes? Are there differences in response between organs/developmental stages?
- Understanding how plants can efficiently integrate environmental signals into phenotypic plasticity (phenotypes for a given genotype in changing environment) through the modification of chromatin marks. What is the general impact of such changes on quality traits, adaptation and phenotypic plasticity? What are the consequences for crop breeding?

Understanding the epigenetic contribution of heritable variability

Assessing the contribution of transgenerational epigenetics to heritable phenotypic variation in either wild species or crops poses major challenges, as most chromatin (DNA methylation) and gene expression variants usually co-segregate with DNA sequence polymorphisms. Nonetheless,

there is now ample evidence that plants have a large potential for bona fide heritable epiallelic variation and that at least part of this potential is realized in nature. Furthermore, we are beginning to understand what determines the variable transgenerational stability of epigenetic variants, which ranges from a few to tens and possibly hundreds or thousands of generations. However, we are far from knowing how much of this additional inheritance system actually contributes to heritable phenotypic variation in natural populations, nor do we know the role of the environment in the creation of environmentally-induced epialleles, which seems less prevalent than was usually anticipated.

Although it has been difficult to affect DNA methylation and more generally chromatin states in a locus-specific manner, the situation is likely to change rapidly with the development of editing tools for the epigenome based on the CRISPR-Cas9 system. Thus, we can anticipate that soon epigenome editing will provide a means to assess for any QTL interval whether allelic and/or epiallelic variants are causal. By the same token, epigenome editing, together with approaches that enable the genome-wide alteration of chromatin states could provide interesting new routes for the improvement of heritable traits.

Cartography of ongoing international initiatives

Overview of the French public and private research on the topic

This section provides an overview of the public laboratories working on plant epigenetics in France. The first group (Table 1) is composed of public research units working on model plants like *Arabidopsis thaliana* and addressing biological questions about fundamental epigenetic processes. In the frame of translational biology, the second group (Table 1) belongs to public units working on crops (maize, rice, sugar beet, wheat, oil palm, poplar, tomato, apple). Both groups have the objectives of unravelling fundamental epigenetic processes to increase our knowledge about the molecular mechanisms involved. Indeed, it is now clear that genome diversity and level of complexity of organisms are associated with marked differences in epigenetic processes. For example, polyploid plants, plants with large genomes and perennials seem to have differ in their epigenetic control. Although Arabidopsis seems to be particularly tolerant to genome-wide modifications of methylation patterns, other plant species and crops (e.g. tomato or maize) in particular are not. In addition, the second group attempts to make use of the knowledge gained on crops for different biological processes (like somatic embryogenesis, heterosis, adaptation to environmental constraints and global changes) to better understand the contribution of epigenetics in general and to design new type of epigenetics molecular makers in particular (complementary to classical genetic markers such as SNPs). Accordingly, the private sector is also developing research programs on epigenetics and plant breeding in collaboration or not with academic partners. Usually, due to confidentiality of this work, it is more complex to evaluate at the present time the contribution of the private sector. Some initiatives between public and private units already exist for maize, oil palm, sugar beet, poplar. However, it seems now essential, to reinforce competitiveness of these excellent units, to better connect the different groups and build up a global national network. Another challenge will be to organize this network at the EU level.

This paper is endorsed by the strategic committee of the GIS BV, the "Plant Biology and Biotechnologies" working group (GT4) of the AllEnvi Alliance.

Annex 1. List of Participants

ACHOUR Zeineb (INRA), ARCHIPIANO Muriel (SOLTIS), BARNECHE Fredy (ENS), BAURENS Christophe (CIRAD), BECKERT Michel (MENESR), BEN Cécile (ENSAT), BOCHARD Anne-Marie (Limagrain), BOUCHE Nicolas (INRA), BUCHER Etienne (INRA), CAIVEAU Sébastien (SYNGENTA), CHABANNES Matthieu (CIRAD), CHAIB Jamila (HM Clause), COLLONIER Cécile (INRA), COLOT Vincent (ENS), DAYDE Jean (ENSAT), DE FARIA MARASCHIN Simone (Nestlé), DELARUE Marianne (UPSud), DEVAUX Pierre (Florimond Desprez), DJENNANE Samia (INRA), DREVENSEK Stéphanie (UPSud), DUGAS Olivier (BIOGEMMA), DURAND-TARDIF Mylène (GIS BV), DUTRIEZ Sophie (CAUSSADES SEMENCES), ENJALBERT Jérôme (INRA), EVRARD Aurélie (INRA), FAIVRE Patricia (INRA), FUDAL Isabelle (INRA), GALLAIS André (INRA), GAMAS Pascal (INRA), GEFFROY Valérie (INRA), GENTY Amélie (SECOBRA Recherches), GENTZBITTEL Laurent (ENSAT), GOLDRINGER Isabelle (INRA), GOLSTEIN Catherine (HCB), GOMEZ Victoria (INRA), GRENIE Alice (Vilmorin), GRIMA-PETTENATI Jacqueline (CNRS), HERBOMMEZ Jean-François (KWS MOMONT), JACOB Emilie (UPJV), JALIGOT Estelle (CIRAD), JAUBERT-POSSAMAI Stéphanie (INRA), JOLY Hélène (CIRAD), JUBAUT Mélanie (Agrocampus Ouest), JUST Jérémy (ENS), KUHN Estelle (INRA), LANCIANO Sophie (IRD), LASARACINA Olivia (Génoplante-Valor), LATRASSE David (INRA), LELANDAIS Christine (CNRS), LE PASLIER Marie-Christine (INRA), LE PROVOST Grégoire (INRA), LECLERC Julie (CIRAD), LECOMTE Sylvain (Linéa Semences de lin), LEGAC Anne-Laure (Université d'Orléans), LEPINIEC Loïc (INRA), LIEGARD Benjamin (INRA), LIONNETON Eric (HM Clause), MANZANARES-DAULEUX Maria (INRA), MARTINANT Jean-Pierre (Limagrain), MAURY Stéphane (Université d'Orléans), MESTIRI Imen (ENS), MEY Géraldine (EURALIS), MIROUZE Marie (IRD), MOQUET Frédéric (GAUTIER SEMENCES), MUNOS Stéphane (INRA), PAUX Etienne (INRA), PETIT Aurélie (Ciref), PICHON Jean-Philippe (BIOGEMMA), PIOVAN Romain (GIS BV), RAYMOND Olivier (ENS), REMAY Arnaud (GEVES), RENARD Michel (INRA), RINCENT Renaud (INRA), ROBERT Olivier (Florimond Desprez), ROMESTANT Michel (RAGT 2n), ROUDIER François (ENS), ROUSTER Jacques (BIOGEMMA), SANTONI Sylvain (INRA), SATGE Carine (INRA), SAUVAGE Christopher (INRA), SZAMBIEN Maxime (GIS BV), TALEB Ghania-Mimia (Limagrain Vegetable Seeds), TATOUT Christophe (Céréales Vallée/Vegepolys/Agri Sud-Ouest Innovation), THIOUNE El Hadji (Nestlé), THIS Dominique (Montpellier SupAgro), TRONTIN Jean-François (FCBA), VAN EX Frederic (Bayer CropScience), VAUCHERET Hervé (INRA), VITTE Clémentine (CNRS), WENDEN Bénédicte (INRA)

Table 1. French public laboratories working on plant epigenetics

Group	Abbreviation	Full-name	Location
Group 1: public research units working on the model plant *Arabidopsis thaliana*	IBENS	Institut de Biologie de l'Ecole Normale Supérieure - UMR8197. Inserm U1024	Paris
	IJPB	Institut Jean-Pierre Bourgin – UMR 1318	Versailles
	IPS2	Institut des Sciences des Plantes de Paris-Saclay – UMR 1403	Orsay
	LGDP	Laboratoire Génome et Développement des Plantes – UMR 5096	Perpignan
	GreD	Génétique Reproduction et Développement – UMR 6293	Clermont-Ferrand
	IBMP	Institut de Biologie Moléculaire des Plantes – UPR 2357	Strasbourg
Group 2: public research units working on crops (maize, sugar beet, wheat, rice, tomato, apple grape, poplar, oil palm…)	GQE	Génétique Quantitative et Evolution Le Moulon - UMR 0320	Gif-sur-Yvette
	LBLGC	Laboratoire de Biologie des Ligneux et des Grandes Cultures EA1207, USC 1328 INRA	Orléans
	GDEC	Génétique, Diversité, Ecophysiologie des Céréales - UMR 1095	Clermont-Ferrand
	IRHS	Institut de Recherche en Horticulture et Semences – UMR 1345	Angers
	IGEPP	Institut de Génétique, Environnement et Protection des Plantes – UMR 1349	Rennes
	AGAP	Amélioration génétique et adaptation des plantes méditerranéennes et tropicales – UMR 1334	Montpellier
	DIADE	Diversité - Adaptation – Développement – UMR 232	Montpellier
	BFP	Biologie du fruit et pathologie – UMR 1332	Bordeaux