Comment about the pathotype identification key proposed in Gascuel et al. (2016) (DOI: 10.1371/journal.pone.0148513) :

Looking for association between the phenotypic response of sunflower lines to their infection by different pathotypes of *P.halstedii* and the genotypic profile of some pathotype effectors:

Author: Patrick VINCOURT, patrick.vincourt@orange.fr

Introduction:

In their paper, Gascuel et al. (2016) discussed *("Are effector genotyping data linked to virulence profiles of Pl. halstedii pathotypes?")* the possibility to infer, from the joint profiles of sunflower phenotypic response to the infection and of the effector genetic polymorphism, which effector could play a particular role in some {sunflower * *P.halstedii* pathotype} interaction, i.e. which (set of) resistance gene allele(s), on the plant side, is interacting with which (set of) effector allele(s), on the pathogen side, to result in a resistance or susceptibility pattern. Gascuel et al. (2016) provided two quite clear examples (PhCRN33 and PhRXLR15, see Fig. 5 of the original manuscript) where such association could be hypothesized. They proposed an identification key of *P. halstedii* pathotypes which might appear limited, as this key was not able to identify each analyzed pathotype by a straight forward, unique genotypic profile based on the effector polymorphism. The purpose of this comment is to provide the readers with some indications suggesting that this result might be due to some limitations of the genotyping technology we used to characterize the whole set of *P.halstedii* samples rather than to a lack of functional relationship between the candidate effectors we mentioned and the role they play in the {sunflower * downy mildew} interaction. Moreover, based on data available from the S1 table of the original manuscript and from Tourvieille et al. (2012), a way to extend this approach is proposed with the purpose to identify other similar, however probably more intricate, situations.

Material and Methods :

Each sunflower genotype can be characterized by its specific response (i.e compatible or incompatible interaction) when faced to the reference *P.halstedii* pathotypes. Each *P.halstedii* polymorphic effector can be also described by the presence or absence of a given allele in these reference pathotypes. When the dataset is adequately built, it is therefore possible to correlate the profile response of sunflower genotypes with the genotypic pattern of these effectors. A full association i.e. same profile for the sunflower genotypes and for the effector genotypic means, for example, that (i) the sunflower genotypes are susceptible to pathotypes A and B and (ii) these A and B pathotypes are exhibiting a specific amino acid profile for these effectors. As a help to identify such situations, we performed a Factorial Correspondence Analysis (FCA, program *dudi.coa*, *ade4* package in R) with an table built from data of S1 table (original manuscript), with *P.halstedii* reference pathotypes in column, and both phenotypic sunflower responses, including the differential lines (Tourvieille et al., 2012), and amino acid effector profiles in rows (Table 1). Then the association, and mostly the lack of association cases, can be identified, looking at the coordinates on the different axis of the FCA.

Effector allele or Sunflower Phenotypic Response	Type	100_Y	304_Y	710_Y	730_Y	703_Y	700_Y	334_Y
CRN33_A	E	0	0	1	1	0	0	1
CRN33_B	E	1	1	0	0	1	1	0
D4_R	S	1	1	0	0	1	1	0
D4_S	S	0	0	1	1	0	0	1
D7_R	S	1	1	1	1	0	1	1
D7_S	S	0	0	0	0	1	0	0
D8_R	S	1	1	1	1	0	1	1
D8_S	S	0	0	0	0	1	0	0
RXLR15_A	E	0	0	0	0	1	0	0
RXLR15_B	E	0	0	0	0	0	0	1
RXLR15_C	E	1	1	1	1	0	1	0

Table 1: Subset of the dataset used to perform the Factorial Correspondence Analysis. The full dataset comprises 124 rows[66 phenotypic sunflower responses, Type=S for Sunflower + 58 effector profiles, Type=E for Effector] and 7 columns for the 7 reference pathotypes] (Download the full dataset here). In the full dataset, D1 to D9 are the sunflower differential lines as usually designated. The response of sunflower infection was described with four classes according to Tourvieille et al. (2012), from S1 to R1. As examples: the data of {CRN33_A,710_Y} = 1 means that the reference isolate of the pathotype 710 has the allele A for the effector CNR33. Similarly, while when the {D4_R, 100_Y} combination has the value 1, it means that the sunflower line D4 has the phenotype R (fully resistant) when fronted to the *P.halstedii* reference pathotype 100.

Results and Discussion:

Only a relatively limited fraction of the SNP polymorphisms initially detected on the seven reference pathotypes allowed detecting the polymorphism on the whole set of *P.halstedii* accessions. More importantly, some of the discriminant SNP – based on the initial data on the seven reference pathotypes – were not able to be converted into KASPar ™ markers for the analysis of the whole set of samples. As a first example: the proposed key does not allow distinguishing the pathotypes 100 and 304. However, as shown in the S1 table , a member of the CRN17 effector family (CRN17.3, see predicted protein alignment at position 236, [R/K]), a non synonymous SNP would be of interest to distinguish the 100 and 304 pathotypes. As a second example: using the proposed key, the pathotype 710 would be identified only at the step 2 (cf. Figure 4 of the original manuscript).However, it appears that among the 3 SNP detected from the sequences of the reference pathotypes, the only one which was able to be used in the frame of the KASPar ™ technology did not permit to exhibit a "710" specificity (see https://www.heliagene.org/P.halstedii/effector_polymorphisms/Plhal027443_to_PLHAL.all.AA.gif for details). This case should be interesting to analyze because, in reference to the discussion about Programmed Ribosomal Frameshifting (PRF), there is two different cases (334 and 710 pathotypes) where a hypothetical PRF could account for the result of the interaction. All together, this suggest that the lack of discriminant power of the proposed key is due to the difficulty to define discriminant KASPar ™ markers, rather than due to an intrisic lack of relationship between effector genotypic profiles and sunflower genotype specific responses.

As shown in Figure 1, some associations were found in comparing the profiles of effectors amino acid sequences and the resistance profiles of sunflower lines. As already described in Gascuel

et al. (2016), a quite simple association of {sunflower response pattern * effector genotype} could be hypothesized in two cases: (D4, PhCRN33) and (D7-D8, PhRXLR15). No other hypothesis could be easily formulated for other cases, probably indicating that the phenotypic result of the interaction might depend from a combination of genes on plant side and on pathogen side, rather than resulting from a simple the “gene for gene” model.

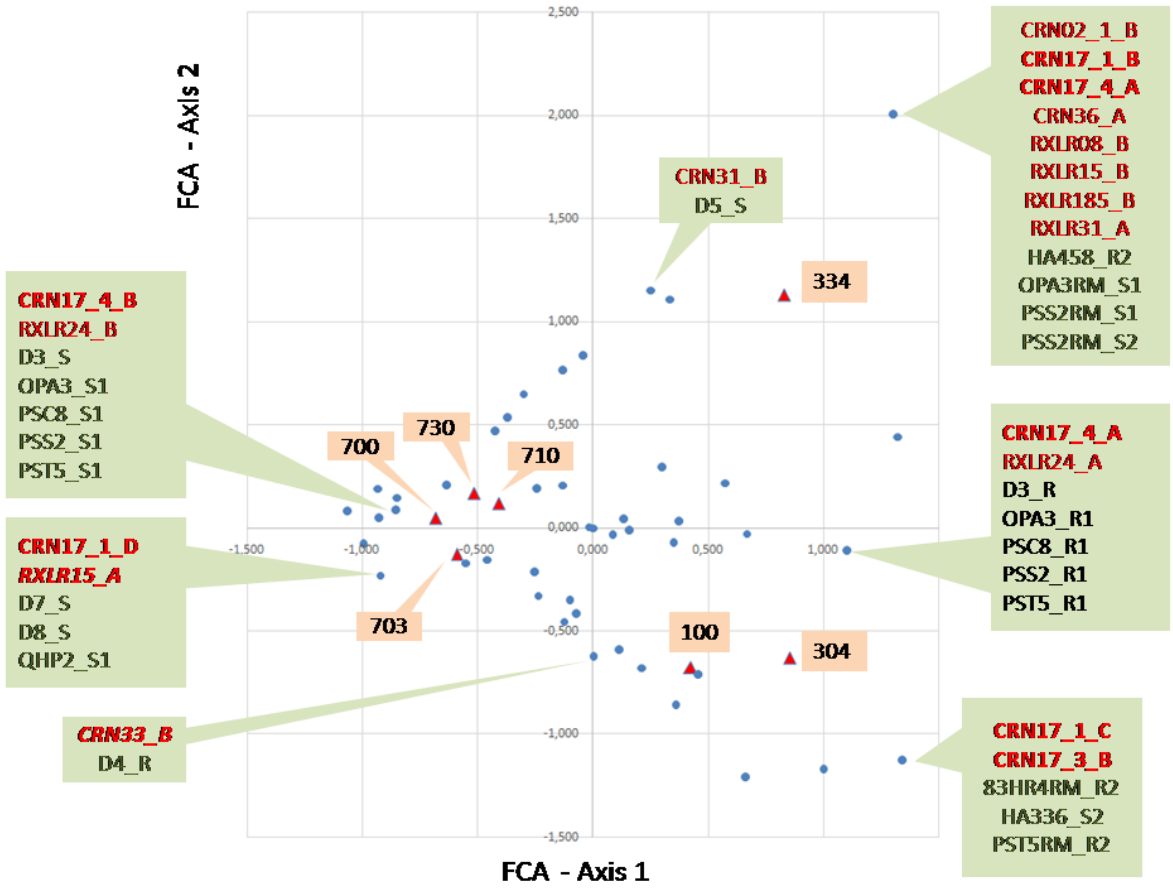

Figure 1: Representation of sunflower genotypic responses *(in green)* , effector polymorphism *(in red)* and reference pathotypes *(pink background)* on the two first axis of the Factorial Correspondence Analysis. See Table 1 for explanations. Detailed results of the FCA are provided here. Only the most significant associations between effector genotypic profile and sunflower specific responses were described in this figure.

It should be however pointed out that the genotypic profile of the members of the CRN17 family is often associated with the host specific response. Of course such associations could be false positive ones for several reasons, including the fact that not all the existing *P. halstedii* pathotypes were characterized or due to the population structure or history, on both sides of the interaction. However, this approach could help to target some effectors before starting functional approaches.

Most of the highly cited works on {plant * oomycete} interactions are dedicated on model species and are involving functional analysis through genetic transformation, which is still considered by the scientific community as the most efficient demonstration at least in terms of mechanistic biology. Only a few studies (as examples Delmotte et al., 2008, Ahmed et al.,2012) have been conducted to analyze the available genetic diversity on the pathogen side and to use it with the purpose, among others, to understand the short term, evolutionary processes taking place during the last century of the agronomic revolution / industrialization. The {*Helianthus annuus* * *Plasmopara halstedii*} interaction seems to be particularly interesting because the expansion of sunflower

cultivation started less than one century ago, and because during this short period of time, the emergence of virulent pathotypes of *P.halstedii* became a real threat for farmers and breeders.

On the pathogen side, such approach benefited and still will benefit from the most recent advances in sequencing technologies. More and more field samples could be genotypically described at an acceptable cost, and it could be predicted that what we call today a "pathotype" – which is based on a phenotypical description of the interaction – will be challenged within the next few years.

Regarding the phenotyping of the {plant * pathotype} interaction, the situation is not so clear. Indeed, in the past, the environmental conditions where the {sunflower * downy mildew} interaction was evaluated was adapted with the aim to (a) provide reproducible evaluations, which is still anyway a scientific requirement, (b) make the scientist able to provide the community with at least a partial explanation of / model for the observed results. Even with a more quantitative scale for the symptoms (S1 to R1, see above), the description of the interaction is probably not multidimensional enough to allow accounting for what is happening inside. Transcriptomic data (As-sadi et al, 2011) might help, as intermediates between genotype and expressed phenotypes, but they are not so easy to produce in the perspective of quantitative and easily reproducible studies. More generally, what is usually called a "pathosystem" does not refer to a particular experimental design. In many cases, such experimental design is not elaborated with the purpose to reflect the environmental conditions where the interaction is taking place in nature or in cultivated fields, but with the purpose to highlight qualitative responses in the frame of mechanistic molecular biology or Mendelian genetics. This trend was prominent during the last 50 years ago, when "strong (hopefully dominant, with Mendelian inheritance) genes are good genes".

The purpose of this obviously limited comment is to show how mobilizing the datasets produced by high throughput (genotyping, phenotyping) technologies could constitute a way to explore such interaction, as a preliminary approach before investing in a costly and risky functional analysis, at least as far as operational, agronomic concerns need to be solved. All together, the joint study of the polymorphism on the plant side and on the pathogen side appears as a way to identify key factors in the present gaming table and thereafter to focus functional studies on the emerging candidates.