EvoEvo Deliverable 1.4

Santiago F. Elena, Guillaume Beslon, Otmane Lamrabet, Dominique Schneider

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

<div style="border:1px solid">

EvoEvo Deliverable 1.4

Analysis of phenotypic innovation (part 1)

</div>

Due date:	M24 (November 2015)
Person in charge:	Santiago Elena
Partner in charge:	CSIC
Workpackage:	WP1 (Experimental observation of EvoEvo in action)
Deliverable description:	Analysis of phenotypic innovation (part 1): Phenotypic innovation: viral host transcriptome and bacterial evolution experiments with changes in regulatory networks.

Revisions:

Revision no.	Revision description	Date	Person in charge
1.0	First version; *E. coli* part	06/10/15	O. Lamrabet and D. Schneider (UJF)
1.1	TEV part	07/10/15	S. Elena (CSIC)
1.2	Formatting and corrections by G. Beslon	13/11/15	G. Beslon (INRIA)
1.3	Reorganisation of tables 1 and 2 for the document to be transferable by mail	09/03/16	G. Beslon (INRIA)
1.4	Adding Intro and conclusions to TEV part	16/03/16	S. Elena (CSIC)
1.5	Complements by UJF	24/03/16	O. Lamrabet (UJF)
1.6	Corrections and validation	25/03/16	G. Beslon (INRIA)

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

Table of Contents

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

1. Introduction

Evolution, the process that created (and still creates) all the diversity of life, is a process by which organisms permanently adapt to their environment. However, the environment of an organism is never stable as it also depends on the evolution of other organisms. While higher eukaryotes have evolved complex sensory-motor systems to adapt their behavior to their changing environment, microorganisms are less sophisticated systems that lack complex sensorimotor abilities. Hence, they efficiently use mutation and selection to dynamically adapt to new conditions. Recent experimental evolution results have shown that they are able to evolve at an amazing speed: in virtually all experimental frameworks that use bacteria or viruses, important phenotypic innovations have emerged in only a few tens of generations (*e.g.*, Rainey & Travisano 1998, Zhang *et al*., 2011), reviewed in Hindré *et al*. (2012). These results show that, more than being adapted to a specific condition, micro-organisms are adapted to evolve: evolution has optimized their own ability to evolve, as a primary mean to react to environmental changes.

The objective of Task 1.3 is to address the mechanisms of phenotypic innovation in microorganisms. In particular, we aim at deciphering these mechanisms at two different levels: population and regulatory network levels. To this aim, we use two different experimental models and framework, a viral strain (TEV) and a bacterial strain (*E. coli*). This choice is based on the assumption that both models use different mechanisms to adapt to new conditions owing to their huge difference of complexity, population sizes and mutation rates. In this deliverable we report the first results gathered on both models by, first, investigating the ability of the TEV experimental model to infect new hosts (Section 2) and the ability of engineered *E. coli* strains to develop phenotypic innovation (Section 3). Note that this deliverable presents the first analysis steps of Task 1.3. As long as the viral model is concerned, the experiments have already been successful and this first part covers the whole story. As explained in Section 3, this has not been the case for the bacterial model that will thus deserves further investigations to be presented in deliverable D1.6.

2. Phenotypic innovation at the population level in the TEV experimental model

The phenotype of a virus is a complex trait that can be defined in several ways. From a viro-centric perspective, the most relevant phenotype is viral fitness. Viral fitness itself is also a quite complex trait, with many components. To better evaluate viral fitness, we have quantified two different fitness components: **infectivity**, which is a proxy to between-host fitness and measures the efficiency of initiating a new infection, and **viral load** which is directly related to within-host viral fitness and measures the efficiency replicating, moving and accumulating.

The general question to be answered in this Section is how likely is a phenotypic innovation to occur in viral populations? Or, in other words, how likely is a viral population to contain a new genotype that successfully infects a new host species and to replicate and accumulate in it? Tackling this question is relevant to shed light on the open-endedness properties of virus evolution. Besides, from the perspective of pathobiology, it is also important because its profound implications for understanding the emergence of new viral diseases.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

We first sought to evaluate a first phenotypic innovation, namely, the ability to infect and replicate in novel hosts. We proposed to evaluate the fraction of all possible mutations that may confer TEV the ability to infect a set of new hosts. We have evaluated the infectivity and viral load of a collection of single-nucleotide substitution mutants of TEV across a panel of eight different hosts. The collection of TEV mutant genotypes was already available in our laboratory (Lalić *et al*. 2011), and was created by single-nucleotide site-directed mutagenesis on an infectious clon of TEV (Carrasco *et al*. 2007). Both the nucleotide residue to be mutagenized and the nucleotide to be inserted were randomly chosen. Infectious capped-RNAs were produced by *in vitro* run-off transcription of the corresponding mutant infectious plasmids, previously linearized by digestion with *Bgl*II, using the SP6 mMESSAGE mMACHINE® kit (Ambion Inc) and following the manufacturer's instructions. These RNAs were used to inoculate at least 10 plants from each species with equal amounts of RNA. The host species selected were *Nicotiana tabacum* (the primary natural host of TEV), *Nicotiana benthamiana*, *Capsicum annuum*, *Datura stramonium*, *Solanum lycopersicum*, *Helianthus annuus*, *Gomphrena globosa*, and *Spinacia oleracea*. The hosts differ in their degree of genetic relatedness to the primary one as shown in Figure 1.
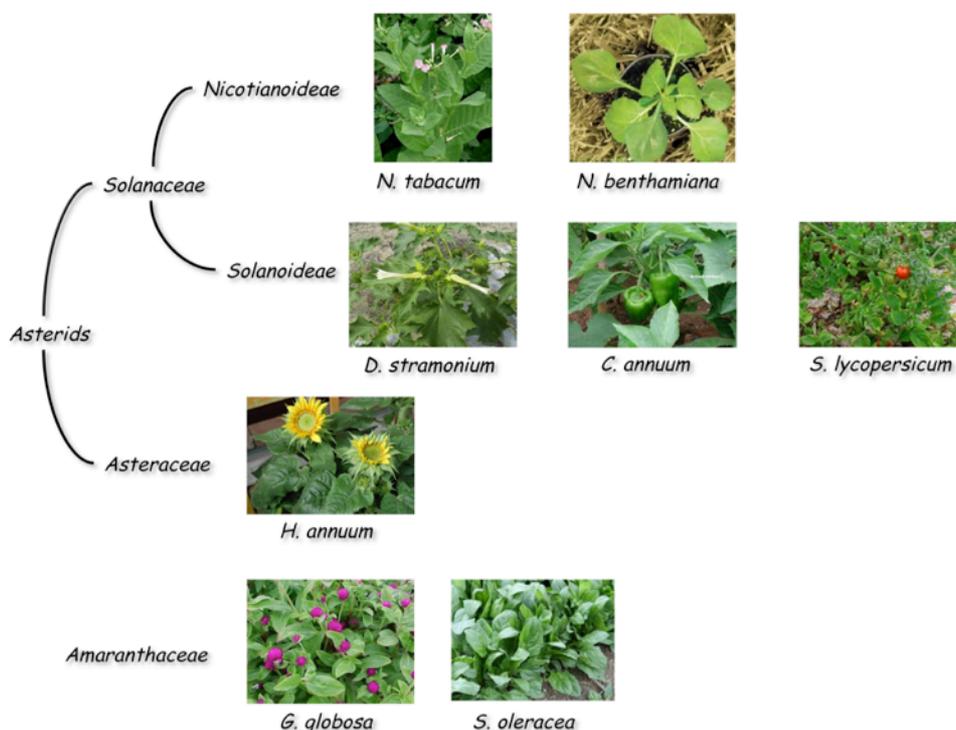

Figure 1. The eight different hosts used to evaluate TEV phenotypic innovation. The hosts were chosen to cover a wide distribution of genetic relatedness with the primary host *N. tabacum*; some belong to the same botanical family (the *Solanaceae*) whereas others pertain to taxonomically divergent families.

The development of symptoms was visually followed, and asymptomatic infections were determined by one-step RT-PCR (TaKaRa Inc). These data allowed us to estimate the infectivity as the frequency of infected plants among the total number of inoculated plants using the LaPlace point-estimator of the Binomial distribution. As a way to assess the extent of phenotypic innovation for the trait infectivity (between host fitness), we have counted the number of genotypes that had infectivity larger than the wildtype (WT) TEV on each host species. If the number of mutant genotypes able of infecting novel hosts increases with the genetic distance from the primary host

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

to the test host, then we will conclude that phenotypic innovation is important for TEV. By contrast, if we found that the number of mutant genotypes able of infecting new hosts remains constant, or even decreases, from the primary host, then we will conclude that phenotypic innovation is not relevant for the infectivity of TEV. Figure 2 shows the frequency of mutant genotypes that had infectivity larger than the wildtype (WT) TEV on each host species. A Spearman's correlation coefficient shows a positive significant association between infectivity and genetic distance between the tested host and the reservoir host (r_S = 0.872, 6 df, P < 0.001). This observation is compatible with the hypothesis that TEV populations show a great potential for phenotypic innovation, at least for between-host traits. This opens the possibility for TEV to initiate infections in widely different host species.
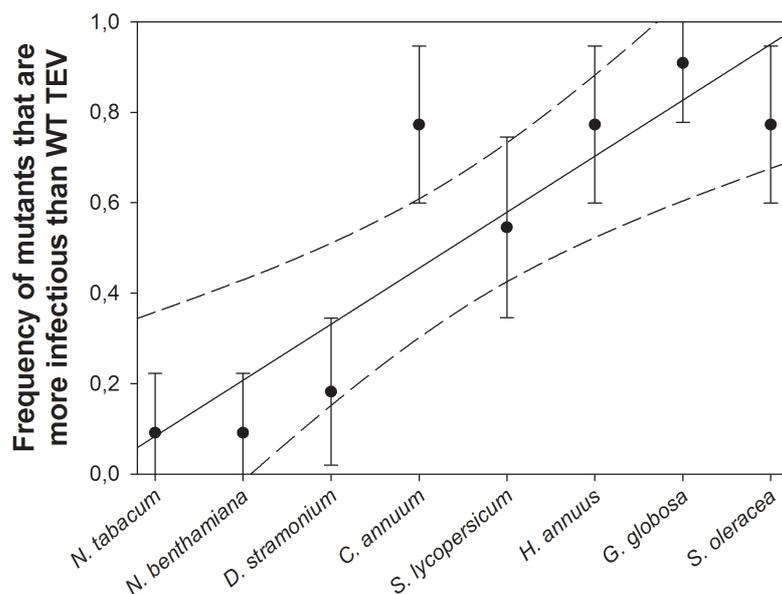

Figure 2. Relationship between the frequencies of mutant genotypes that are more infectious than the WT on each host species. The solid line represents the regression line and the dashed line the 95% confidence interval for the regression line. Error bars represent the margin of error of the LaPlace estimator of the Binomial parameter.

Next, for each infected plant, we evaluated TEV viral load by the RT-qPCR method and using the One Step SYBR Prime Script RT-PCR kit II (TaKaRa Inc) following the instructions provided by the manufacturer and a pair of specific primers that amplify 71-nt fragment within the *CP* cistron of TEV genome. As above, to evaluate the extent of phenotypic innovation for the trait viral load (within host fitness), we counted the number of mutant genotypes with a value of viral load greater or equal than WT TEV. The same rational than above applies: if this value increases with the diversification of hosts, then we will conclude that phenotypic innovation is important; otherwise, we will conclude the opposite. Figure 3 shows the frequency of mutant genotypes that accumulate at least as WT on each tested host as a function of the taxonomic distance from the primary host. Overall, no significant association exists between these two variables (r_S = 0.084, 6 df, P =0.794). However, a more careful observation of the data in Figure 3 suggests that two different allometric relationships exist, one characteristic of the *Solanaceae* hosts and another typical for the non-*Solanaceae* hosts, although the slope of both relationships seems to be the same, the intercept is clearly different. Indeed, a partial correlation coefficient controlling for these two groups turns out to be negative and highly significant (r = −0.944, 5 df, P = 0.001), suggesting that for this within-host fitness trait, phenotypic innovation decreases with the genetic distance between hosts.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

We need to generate and test hypothesis to explain (*i*) why viral load follows two different allometric relationships among host classes for and (*ii*) why within-host fitness and between-host fitness behave in opposite ways in terms of phenotypic innovation.
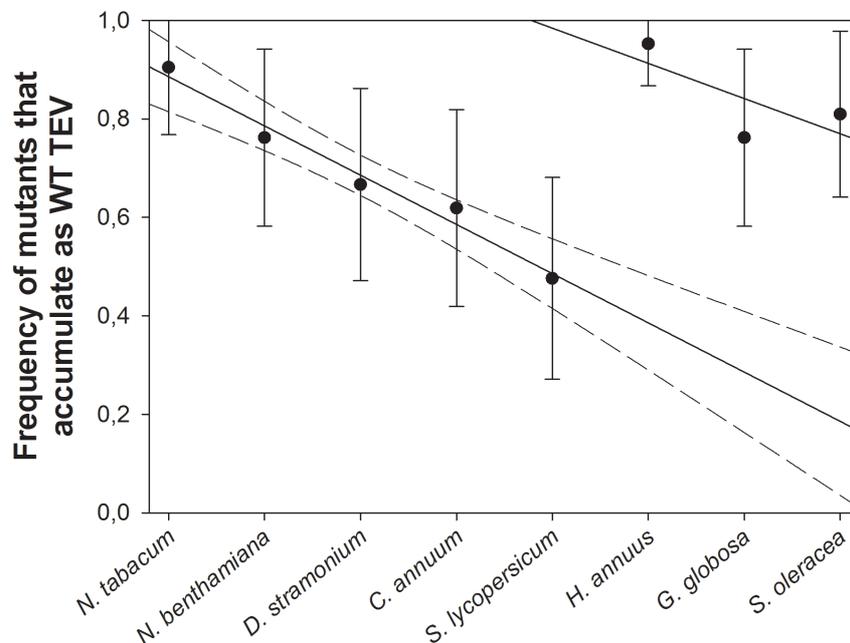

Figure 3. Relationship between the frequencies of mutant genotypes that had a larger viral load than the WT on each host species. The solid lines represent the regression lines and the dashed line the 95% confidence interval for the regression line obtained for the *Solanaceae* hosts. Error bars represent the margin of error of the LaPlace estimator of the Binomial parameter.

So far, we have been concerned with a viro-centric perspective of viral fitness. However, a more biologically meaningful perspective would necessarily involve the interaction between the virus and the host (Elena *et al*. 2011; Rodrigo *et al*. 2012). We proposed to evaluate a second level of phenotypic innovation, namely, the differences in the severity of the perturbation induced by several genotypes of TEV in the coordination of gene expression and the metabolism of the experimental host *Arabidopsis thaliana*. To do so, we selected 15 TEV genotypes evolved by serial passages in five different ecotypes of *A. thaliana* that differ in their susceptibility to infection from the ancestral WT TEV (Lalić *et al*. 2010; Hillung *et al*. 2012, 2014, 2015): Di-2, Ei-2, L*er*-0, St-0, and Wt-1. We have performed a set of transcriptomic experiments using the Agilent *Arabidopsis* (V4) gene expression microarray 4×44K technology. mRNA microarray assays make feasible quantifying the expression for almost all genes of *A. thaliana* during plant response triggered by virus infection. This is a comparative transcriptomics study (Hillung *et al*. submitted), and as such, it provides tremendous amount of information about the changes in expression of a multitude of genes and of derived biological functions. Rather than looking at endless lists of genes and functions, here we have focused our attention to general questions and tried to infer general patterns of viral adaptation. The questions tackled in the followings section are (1) which classes of host genes are the targets for virus adaptation? (2) Do they differ among host ecotypes? (3) Upon adaptation to a given local host ecotype, does the evolved virus interact differently with its ancestral host, L*er*-0? If so, the interaction with which classes of genes has been affected? And finally, (4) are transcriptomic responses induced by specialist and generalist viruses different from

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

each other? Answers to questions (2), (3) and (4) will directly inform about phenotypic innovations at the level of host-virus interactions.

The first question tackled was to which extent viral lineages evolved in the same host ecotype converged in the way they interacted with their plant hosts and diverged from viral lineages evolved in other ecotypes. Here convergence and divergence may be taken as synonymous of lack of phenotypic innovation or presence of phenotypic innovation, respectively. We analyzed the similarity in the profiles of gene expression among the three viral lineages evolved in a common host. In all cases, the correlations were significantly positive, suggesting a certain degree of evolutionary convergence among viral lineages evolved in the same host ecotype. However, the extent of convergence was variable among ecotypes: while for Di-2 and L*er*-0 the average similarity was slightly over 60%, it was around 80% for lineages evolved in St-0, Ei-2 and Wt-1, suggesting that the latter lineages may impose more severe constraints to virus adaptation and thus the number of possible solutions available for the evolving viral populations are limited. Next, we evaluated whether a given viral lineage interacted in a similar way with all host ecotypes. In other words, whether the evolutionary convergence observed in the patterns of gene expression of infected local ecotypes was similar for the alternative hosts. To do so, we computed the correlation coefficients among average patterns of gene expression for each lineage on each ecotype. Overall, two significant groups of virus-ecotype interactions exist. In the first group, lineages evolved in L*er*-0 and Di-2 formed a cluster, which is subsequently divided into two branches, each one grouping lineages of corresponding ecotype. The second cluster incorporated lineages evolved in Ei-2, St-0 and Wt-1, though they segregated in separated ecotype-defined subgroups. Within this second cluster, Wt-1- and Ei-2-evolved lineages were more similar in their interaction with the hosts than St-0-evolved lineages. From these set of analyses, we can conclude that the expression profiles of host's mRNAs and their corresponding functional profiles were heterogeneous among ecotypes (phenotypic innovation at the between-ecotype level), although a significant degree of parallelism exist among lineages evolved in the same ecotype (no phenotypic innovation at the within-ecotype level). These observations suggest that the evolution of new virus-host interactions was restricted by the genetic characteristics of the host. However, clustering based on functional response shows a different topology, suggesting that similar functional categories were probably altered by the effect of viral infection on different genes on each ecotype.

In a second set of analyses we aimed to identify universal host targets of viral adaptation. In other words, we compared the response of each given ecotype to infection with the locally-evolved isolates with the response to the infection with the WT TEV isolate (Hillung *et al*. 2012). Then, the lists of genes altered as a consequence of viral adaptation to each given ecotype were compared in search of commonalities. Tables 1 and 2 show the list for up- and down-regulated functional categories, respectively. Although all ecotypes include a core set of super categories that are specific (*i.e.*, not shared with other ecotypes), some up- and down-regulated functional super clusters were shared among two or more ecotypes. For instance, 98 categories were found for over-expressed genes, 74 of which were ecotype-specific (8 for Di-2, 22 for Ei-2, 12 for L*er*-0, 20 for St-0, and 12 for Wt-1). Although in all ecotypes at least one category was directly related to immune response, this functional group was not as commonly appearing or diverse as other functional groups, indicating that the general resistance mechanisms of the plant were not the main target for viral adaptation. Indeed, not direct modification of any plant immune system pathway could be assessed during the virus adaptation. In terms of changes of virus-host interactions

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

caused by the virus being adapted to each host ecotype, they were frequent for lineages evolved in Ei-2, St-0 y Wt-1, as demonstrated by mayor number of differently affected genes in these ecotypes. Fewer of such evolved transcriptional modifications were found in L*er*-0 and Di-2. This division into two groups reflects the scheme of the similarities in the gene expression of ecotypes infected with ancestral virus and highlighted the importance of the host genotype for the fate evolutionary fate of TEV. The analysis of affected functional categories allows us to fathom what is happening in the hosts while the virus is adapting to them. Specific changes in the interactions, caused by different evolved lineages, had heterogeneous profiles among ecotypes, but also some intersections in affected functional categories were detected, indicating most redundant targets of viral adaptation. There were no intersections between all ecotypes. According to the similarities in functional responses, ecotypes could be classified into disjoint groups. It is worth mentioning that the composition of functional up- and down-regulated categories induced by the lineage Ler-0/2 in his local ecotype showed significant changes compared to the ancestral virus, in spite that no mutations were fixed in this lineage, attributing the transcriptomic changes mostly to low-frequency subpopulations of viruses.
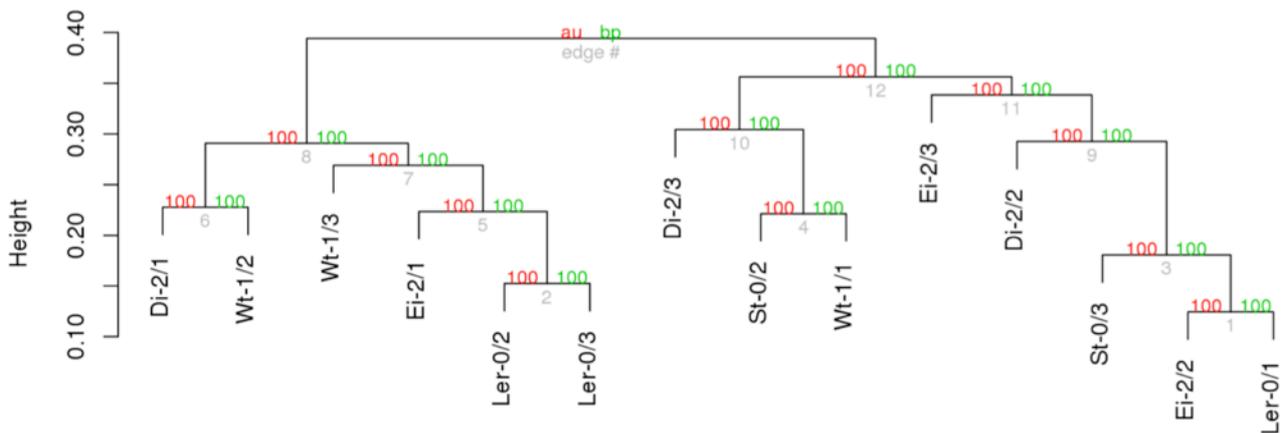

Figure 4. Similarity in transcriptomic profiles between plants of the L*er*-0 ecotype infected with each of the evolved viral lineages. Red numbers represent the approximately unbiased support of each cluster (percentage *P*-value) computed by multiscale bootstrapping. Green numbers represent the support of each cluster based on a standard bootstrapping. Grey numbers indicate the node label.

In a third set of analyses, we evaluated whether the experimental phase of evolution and concomitant adaptation to each new host ecotype was associated with a change in the way the evolved viruses interacted with the original ecotype Ler-0. In other words, which changes in the way evolved viruses interact with the transcriptome of Ler-0 plants may explain the negative pleiotropic fitness effects observed for some of the evolved lineages (Hillung et al. 2014). To do so, the transcriptomic profiles of Ler-0 plants infected with each of the evolved viral lineages were contrasted to the transcriptome of Ler-0 plants infected with the ancestral WT TEV. First, we identified the number of genes with altered expression in Ler-0 plants infected with each of the evolved viruses compared with plants infected with the ancestral virus. A very minor number of significant alterations were found. Next, we sought to evaluate how similar was the response of L*er*-0 to infection with the different evolved lineages. To do so, we computed Person's correlation coefficients among the transcriptomic profiles obtained for each viral lineage when infecting L*er*-0. The matrix of correlation coefficients was used to build a hierarchical cluster in which lineages cluster according to the similarities of the transcriptomic profiles of their L*er*-0 host (Figure 4). The significance of clusters was evaluated using the approximately unbiased *P*-values and bootstrap

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

probability. In this case, all transcriptomes were significantly and positively correlated, with average similarities ranging between 73.6% (St-0 and Wt-1) and 89.9% (Ei-2 and L*er*-0). Clustering pattern shown in Fig. 4 reflects this high similarity in the response of Ler-0 to all the lineages: clustering does not reflect the local host ecotype in which viral lineages evolved.

Finally, we were interested in answering the question of whether generalist and specialist viruses interact in a different manner with the different host ecotypes. Our hypothesis is that a specialist virus (less prone to phenotypic innovation) will show a marked difference between the local and alternative hosts, whereas the generalist virus (more prone to phenotypic innovation) will interact similarly across all hosts. Therefore, we expected (1) that specialist and generalist viruses will alter different sets of host genes and that (2) generalist viruses will alter the expression of similar genes or functions across host ecotypes whereas specialist viruses will show a greater degree of heterogeneity. To test these predictions, we have characterized the transcriptome of plants from all ecotypes infected with the most generalist and the most specialist viruses evolved in our previous experiments (Hillung *et al.* 2014). In our previous evolution experiment lineage L*er*-0/1 was found to be a generalist virus, while lineage St-0/3 was qualified as the most specialist one (Hillung *et al*. 2014). First, we evaluated gene expression differences between both viruses. For the more generalist virus, the expression values for plants from each ecotype infected with L*er*-0/1 were contrasted to the transcriptome of L*er*-0 plants (*i.e.*, the local host) infected with L*er*-0/1. In this way specific response of ecotypes to L*er*-0/1 lineage could be compared. As shown in Fig. 5a, there are only a reduced number of differentially expressed genes when ecotypes are compared (median of 0, interquartile range 8), except in the case of the infection of ecotype Ei-2, where more than 3000 genes showed altered expression in an ecotype-specific manner. Following a similar logic, we contrasted the transcriptomic effect of infecting plants with the most specialized viral lineage St-0/3 to the transcriptomic effect on plants of the local ecotype St-0 infected with the same virus isolate. Fig. 5b shows that in this case the number of differentially expressed genes across ecotypes was more variable (median of 29.50, interquartile range 39) with, again Ei-2 showing the largest number of altered genes (791). Indeed, a signs test showed that both distributions of counts were significantly different (*P* = 0.002), with the generalist virus inducing a very similar perturbation across host ecotypes and the specialist virus inducing different sets of genes on each ecotype. Transcriptomic response of non-local ecotypes to the St-0/3 specialist virus is different from the response of the local ecotype St-0. Ten over- and 6 under-expressed genes are in common for all ecotypes. In summary, these results back up our original hypothesis: generalist virus Ler-0/1 induces very similar perturbations in the transcriptomes of the different ecotypes analyzed. By contrast, the perturbations induced by the specialist virus St-0/3 are divergent among ecotypes.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6
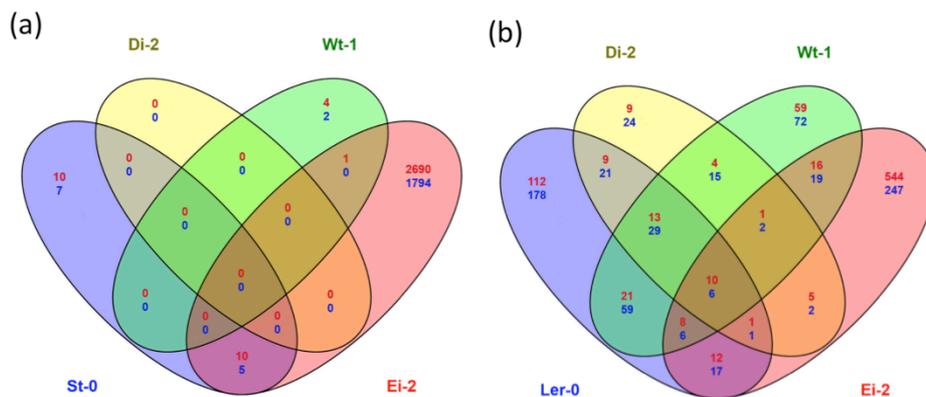
Figure 5. Venn diagrams illustrating the similarities in gene expression patterns across host ecotypes upon infection with (a) the most generalist virus lineage L*er*-0/1 and (b) the most specialist virus lineage St-0/3. In red, number of over-expressed genes; in blue number of under-expressed genes.

Table 1. Up-regulated functional clusters of GO terms for biological processes. Left column represents super categories of GO terms. The rest of columns indicate whether the super category was up-regulated in the corresponding ecotype.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

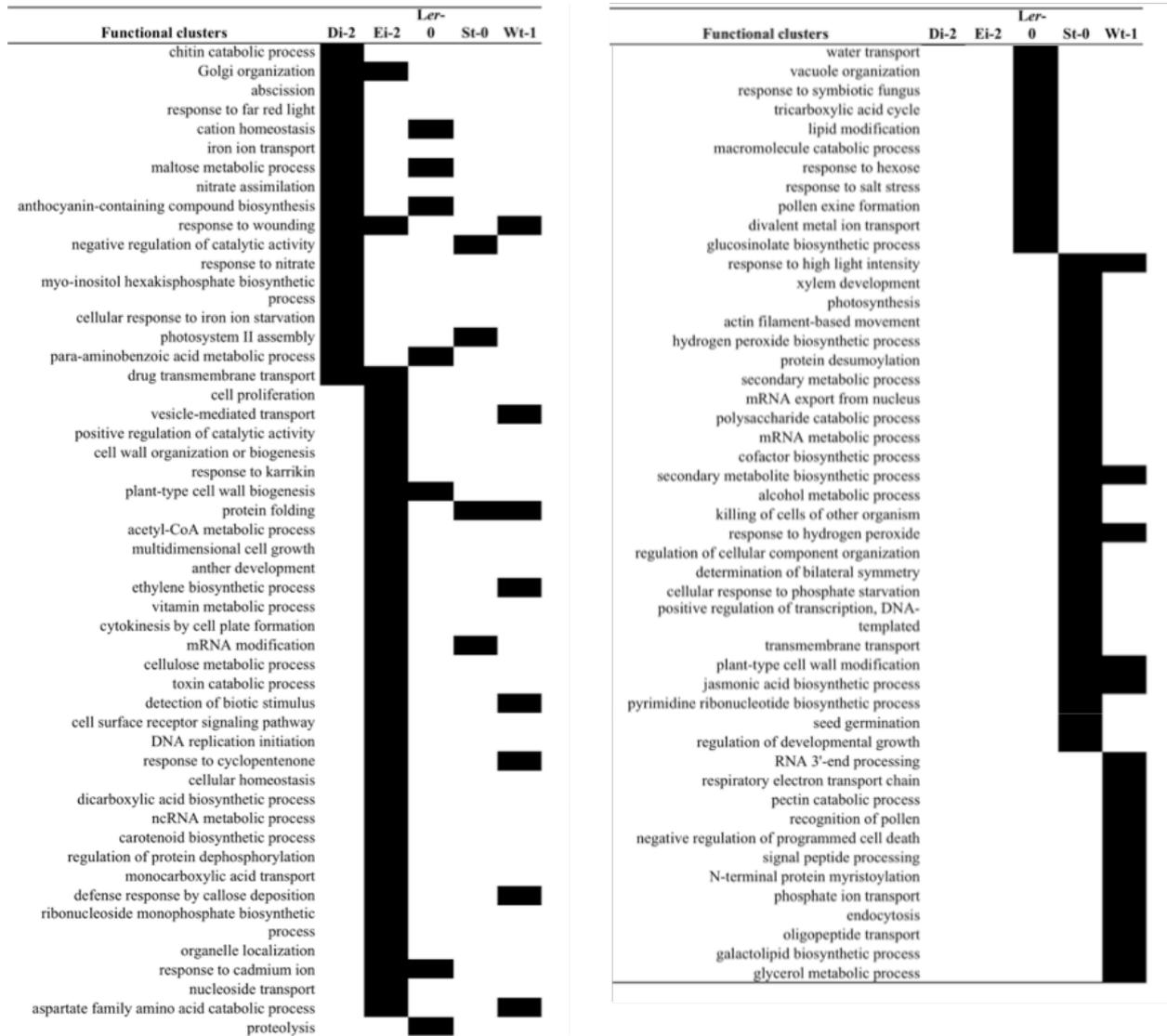| Functional clusters	Di-2	Ei-2	Ler-0	St-0	Wt-0
acetyl-CoA metabolic process	■				
circadian rhythm	■				
cell proliferation			■	■	
tissue development			■		
response to ionizing radiation					
ribosomal small subunit biogenesis					■
positive regulation of hydrolase activity			■		
protein maturation				■	■
photorespiration				■	
transcription factor import into nucleus					
cell-cell signaling					
mitotic recombination			■		
microtubule-based movement	■				
mitotic cell cycle					
dicarboxylic acid biosynthetic process					
response to molecule of bacterial origin					
regulation of cell proliferation					
response to endoplasmic reticulum stress					
cell redox homeostasis					
mRNA metabolic process					
protein folding			■		
nucleus organization					
cullin deneddylation					
pseudouridine synthesis					
DNA-templated transcription, elongation					
regulation of nucleoside metabolic process					
translational elongation					■
proteasomal protein catabolic process	■				
RNA endonucleolytic cleavage and ligation	■				■
posttranscriptional regulation of gene expression	■				
glycerophospholipid biosynthetic process	■				
peptidyl-amino acid modification	■				
N-terminal protein myristoylation			■		
translation			■	■	■
vernalization response					
killing of cells of other organism		■			
protein import into peroxisome matrix		■			
abscission		■			
myo-inositol hexakisphosphate biosynthetic process				■	
chlorophyll catabolic process					
gibberellic acid mediated signaling pathway					
seed dormancy process					
ethylene biosynthetic process			■		
cytoplasmic transport			■		

Functional clusters	Di-2	Ei-2	Ler-0	St-0	Wt-0
response to decreased oxygen levels			■		
chromatin assembly or disassembly					■
anther development			■		
microtubule-based process					■
actin filament-based movement			■		
cytokinesis by cell plate formation			■		
toxin catabolic process			■		
salicylic acid biosynthetic process			■		
response to cyclopentenone			■		
negative regulation of programmed cell death				■	
RNA methylation			■		
ER-nucleus signaling pathway			■		
cell fate specification			■		
response to hydrogen peroxide			■		
pyrimidine ribonucleotide biosynthetic process			■		
heat acclimation			■		
detection of biotic stimulus				■	
proteasome core complex assembly				■	
regulation of G2/M transition of mitotic cell cycle				■	
ER to Golgi vesicle-mediated transport				■	
response to mechanical stimulus				■	
auxin biosynthetic process				■	
response to glucose				■	
respiratory burst involved in defense response				■	
regulation of hydrogen peroxide metabolic process				■	
DNA replication initiation				■	
MAPK cascade				■	
positive regulation of flavonoid biosynthesis				■	
plant-type cell wall cellulose metabolic process				■	
ubiquitin-dependent protein catabolic process				■	
cellular response to hypoxia					■
response to gibberellin					■
vegetative phase change					■
protein targeting to chloroplast					■
DNA recombination					■
maltose metabolic process					■
photosynthetic electron transport in photosystem I					■
cell cycle process					■
response to gamma radiation					■
protein deubiquitination					■
nuclear-transcribed mRNA catabolic process					■
auxin polar transport					■
polarity specification of adaxial/abaxial axis					■
regulation of protein modification process					■
leaf morphogenesis					■
regulation of gene expression, epigenetic					■

Table 2. Down-regulated functional clusters of GO terms for biological processes. Left column represents super categories of GO terms. The rest of columns indicate whether the super category was down-regulated in the corresponding ecotype.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

3. Phenotypic innovation at the regulatory network level in the *E. coli* experimental model

During a long-term evolution experiment (LTEE), twelve populations are propagated by daily transfers from a common *Escherichia coli* ancestor for more than 60,000 generations in a constant glucose-limited environment. Adaptation of *E. coli* during the LTEE has been shown to involve the rewiring of global regulatory networks [Hindré et al., 2012]. In particular, the CRP-controlled regulon has been shown to be increasingly important during evolution [Cooper et al., 2008]. Deletions of the global regulatory *crp* gene have been introduced in both the ancestor of the LTEE and two independently evolved clones, one sampled from each of two of the twelve populations after 20,000 generations [Cooper et al., 2008]. Deleting *crp* had a much more dramatic effect on both the growth in the evolution environment and the global transcription profile of the two evolved clones than of the ancestor. Because the sequence of the *crp* gene was unchanged after 20,000 generations during evolution, these differences indicated epistatic interactions between *crp* and mutations at other loci that accumulated during evolution [Cooper et al., 2008]. Therefore, epistasis has been important in the adaptive evolution of these bacterial populations. However, the nature of these changes in epistatic interactions is unknown. Identification of the types of genetic changes that epistatically interact with the *crp* deletion would give new insights into the mechanisms through which epistasis can evolve. Moreover, phenotypic investigation of these changes in the global regulatory network will allow us to address whether these modifications of the interactions between global regulators within the regulatory network affected the robustness and evolvability of the evolved clones.

Bacterial evolution experiments have shown that adaptive diversification associated with phenotypic innovation may occur in almost all tested environments. This evolutionary outcome is expected and even predictable when environments are heterogeneous (presence of spatial structure, different carbon sources…), owing to the availability of different ecological niches [Kassen & Rainey, 2004]. Moreover, and less expected, adaptive diversification also emerged in more homogeneous environments [Plucain et al., 2014; Le Gac et al., 2012; Rosenzweig et al., 1994], owing to niche construction whereby bacteria generate themselves new ecological opportunities, for instance by secreting metabolites [Laland et al., 1999]. When known, the mutations resulting in such outcomes affect regulatory genes or sequences [Plucain et al., 2014; Bantinaki et al., 2007; Treves et al., 1998].

This deliverable included evolution experiments that used strains constructed during the previous tasks to study phenotypic innovation at the regulatory network level in the 12 populations of the LTEE. The objective will be to investigate the relationships between the structure of global regulatory networks and the bacterial ability to produce phenotypic innovation.

We initiated evolution experiments using clones with different regulatory network structures as ancestors: the ancestor and one evolved clone sampled at 40,000 generations from each of the 12 long-term populations together with each of their *crp*-deleted counterpart obtained from task 1.1 section 2. All strains were propagated by daily transfers for 500 generations (2.5 months) in minimal medium containing glucose and acetate, conditions known to promote adaptive diversification [Le Gac et al., 2008]. We investigated their ability to produce co-existing lineages of bacterial cells with differential phenotypic abilities by plating at regular time points (after 100, 200,

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

300, 400 and 500 generations) on rich medium-containing plates. All *crp+* strains were able to produce small and large colonies, reflecting the presence of two co-existing sub-lineages after diversification in each single population. The sub-lineage of large cells grows on glucose while the small cells switch quickly to acetate consumption. However, after rewiring the regulatory network by deleting *crp*, no mixture of colony size was detected, showing that the ability to diversify into two sub-lineages (at least based on glucose versus acetate consumption and colony size) was lost upon *crp* deletion.

These results precluded further analyses of diversification events after *crp* deletion and showed that deleting *crp* eliminated the ability to produce a polymorphic population during evolution in these conditions.

Therefore, we decided to evaluate the effect of deleting *crp* on an already-established polymorphism. The Ara–2 population is characterized by the presence of two lineages, S and L that diverged at ~ 6500 generations and coexist ever since [Plucain et al. 2014]. Their co-existence involves niche construction through cross-feeding. Hence, the L lineage grows faster on glucose but secretes by-products that S can better exploit, generating negative frequency-dependent selection. We deleted *crp* in one evolved clone sampled from each of the S and L lineages after 40,000 generations of evolution. These strains are currently investigated for the phenotypes that are diagnostic of each S and L lineage.

4. Conclusion

In the experiments presented in this document we investigated to what extent microorganisms (here a virus and a bacteria) are able to cope with changes in their environment, i.e. whether they are able to *innovate* in face of new challenges.

In the TEV model the experiments have been highly successful:

- TEV populations show a great potential for phenotypic innovation: mutations are available within the standing genetic variation present in the reservoir host that allow successfully infecting a novel, distantly related, host species.
- Ability to successfully infect a novel host species does not correlate with within-host fitness (i.e., replication rate), meaning that further fine-tuning adaptation is required to maximize viral fitness.
- The genotype of the host individual drives the adaptation of TEV. Host molecular drivers of virus adaptation have been identified.
- Adaptation to new host genotypes of the same species is, in general, associated to fitness loses in the original host genotype.
- The most generalist virus targets similar molecular factors across different host genotypes. This explains why no-cost generalists have evolved.
- By contrast, the most specialist virus targets different molecular factors on each potential host genotype. This explains why specialists have low within-host fitness in alternative host genotypes.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

On the opposite, on the bacterial model (E. coli) the experiments have been more deceptive:. Indeed, the *crp* gene is involved in carbon source consumption. Therefore, the ability to diversify in an environment containing two carbon sources was lost upon deleting *crp*.

5. References

Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, Rainey PB. 2007. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* **176**: 441-53.

Carrasco, P., de la Iglesia, F., Elena, S.F. 2007. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J. Virol*. **81**: 12979-12984.

Cooper TF, Remold SK, Lenski RE, Schneider D. 2008. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. *PLoS Genet.* **4**: e35.

Elena SF, Carrera J, Rodrigo G. 2011. A systems biology approach to the evolution of plant-virus interactions. *Curr. Opin. Plant Biol*. **14**: 372-377.

Hillung J, Cuevas JM, Elena SF. 2012. Transcript profiling of different *Arabidopsis thaliana* ecotypes in response to tobacco etch potyvirus infection. *Front. Microbiol*. **3**: 229.

Hillung J, Cuevas JM, Elena SF. 2015. Evaluating the within-host fitness effects of mutations fixed during virus adaptation to different ecotypes of a new host. *Phil. Trans. R. Soc. B* **370**: 20140292.

Hillung J, Cuevas JM, Valverde S, Elena SF. 2014. Experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. *Evolution* **68**: 2467-2480.

Hillung J, García-García F, Dopazo J, Cuevas JM, Elena SF. The transcriptomics of an experimentally evolved plant-virus interaction. *Submitted*.

Hindré T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial adaptive abilities by *in vivo* and *in silico* experimental evolution. *Nat. Rev. Microbiol.* **10**: 352-365.

Kassen R, Rainey PB. 2004. The ecology and genetics of microbial diversity. *Annu. Rev. Microbiol*. **58**: 207-231.

Laland KN, Odling-Smee FJ, Feldman MW. 1999. Evolutionary consequences of niche construction and their implications for ecology. *Proc. Natl. Acad. Sci. USA* **96**: 10242-10247.

Lalić J, Agudelo-Romero P, Carrasco P, Elena SF. 2010. Adaptation of tobacco etch potyvirus to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection of resistant ecotypes. *Phil. Trans. R. Soc. B* **65**: 1997-2008.

Lalić J, Cuevas JM, Elena SF. 2011. Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genet*. **7**: e1002378.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 1.4
Analysis of phenotypic innovation (part 1)
Version 1.6

Le Gac M, Plucain J, Hindré T, Lenski RE, Schneider D. 2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **109**:9487-9492.

Plucain J, Hindré T, Le Gac M, Tenaillon O, Cruveiller S, Médigue C, Leiby N, Harcombe WR, Marx CJ, Lenski RE, Schneider D. 2014. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* **343**: 1366-1369.

Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature* **394**:69-72.

Rodrigo G, Carrera J, Ruiz-Ferrer V, del Toro FJ, Llave C, Voinnet O, Elena SF. 2012. A meta-analysis reveals the commonalities and differences in *Arabiodpsis thaliana* response to different viral pathogens. *PLoS ONE* **7**: e40526.

Rosenzweig RF, Sharp RR, Treves DS, Adams J. 1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**: 903-917.

Treves DS, Manning S, Adams J. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol. Biol. Evol*. **15**: 789-797.

Zhang Q, Lambert G, Liao D, Kim H, Robin K, Tung CK, Pourmand N, Austin RH. 2011. Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science* **333**:1764-1767.