Transcriptional plasticity evolution in two strains of Spodoptera frugiperda (Lepidoptera: Noctuidae) feeding on alternative host-plants

Marion Orsucci, Yves Moné, Philippe Audiot, Sylvie Gimenez, Sandra Nhim, Rima Nait-Saidi, Marie Frayssinet, Guillaume Dumont, Alexandra Pommier, Jean Paul Boudon, et al.

Transcriptional plasticity evolution in two strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) feeding on alternative host-plants

Marion Orsucci[1,2], Yves Moné[1], Philippe Audiot[2], Sylvie Gimenez[1], Sandra Nhim[1], Rima Naït-Saïdi[1], Marie Frayssinet[1], Guillaume Dumont[1], Alexandra Pommier[1], Jean-Paul Boudon[3], Marin Vabre[3], Stéphanie Rialle[4], Rachid Koual[4], Gael J. Kergoat[2], Rodney N. Nagoshi[5], Robert L. Meagher[5], Emmanuelle d'Alençon[1] & Nicolas Nègre[1§]

[1] DGIMI, Univ. Montpellier, INRA, Montpellier, France

[2] CBGP, INRA, IRD, CIRAD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[3] UE Diascope, INRA, 34130 Mauguio, France

[4] Montpellier GenomiX, BioCampus Montpellier, CNRS UMS3426, INSERM US009, Univ. Montpellier, Montpellier, France

[5] USDA-ARS, Gainesville, Florida, USA

[§] to whom correspondence should be addressed : nicolas.negre@umontpellier.fr

Abstract

Spodoptera frugiperda, the fall armyworm (FAW), is an important agricultural pest in the Americas and an emerging pest in sub-Saharan Africa, causing damage to major

crops such as corn, sorghum and soybean. While FAW larvae are considered polyphagous, differences in diet preference have been described between two genetic variants: the Corn strain (sf-C) and the Rice strain (sf-R). These two strains are sometimes considered as distinct species, raising the hypothesis that host plant specialization might have driven their divergence. Ecological speciation takes place when adaptations to different ecological niches lead to the reproductive isolation of two populations. Under this hypothesis, we expect that the transcriptional response to the host plants should affect differently the fitness of the two FAW strains. We also expect that these genes should also be linked to a reproductive isolation mechanism between the strains. In this study, we performed controlled reciprocal transplant (RT) experiments to address the impact of plant diet on several traits linked to the fitness of the sf-C and sf-R strains. The phenotypical data suggest that sf-C is specialized to corn. We then used RNA-Seq to analyze the gene expression of FAW larvae from RT experiments. We show that each strain has a different response to the same plant diets. However, we also found constitutive transcriptional differences between strains in laboratory and in natural populations. In particular, we show that mitochondrial transcription is the main difference between strains. A difference in mitochondrial function may be the basis for a shift in host plant and could be involved in hybrid incompatibility, raising the hypothesis that mitochondrial genome is the main target of selection between the two strains.

Introduction

The relatively recent development of agroecosystems modified the ecological niches in many ways (O'Brien and Laland 2012). First and foremost, artificial selection used by early farmers in south-west Asia as of 10,000 years ago to improve their crops, elicited the rapid apparition of new domesticated varieties in the biosphere (Zohary, Hopf, and Weiss 2012). Whilst being selected for human favored traits, cultivated plants concomitantly lost or gained additional properties and thus some animals were prone to exploit these new niches. Some phytophagous insects adapted to cultivated plants and, with the intensification of production based on monoculture activities, these insects eventually become agricultural pests. This adaptation to agricultural plants provides an interesting model system to observe evolution at a relatively small time-scale and assess the genetic changes that may promote speciation in relation to environmental changes (Yoder et al. 2010).

Spodoptera frugiperda (J.E. Smith) (Lepidoptera: Noctuidae: Hadeninae), also known as the fall armyworm (FAW), constitutes a good model to study adaptation of phytophagous insects to agricultural plants. Its native distribution range spans a vast amount of the Americas from Brazil to Canada (Pogue 2002). The FAW has no winter diapause (Sparks 1979) and its wintering range is constrained to warmer regions such as southern Florida and southern Texas in the United States (Nagoshi and Meagher 2004). In 2016 it became invasive on the African continent where massive crop damages have been observed across sub-Saharan Africa in less than a year (Goergen et al. 2016; Jeger et al. 2017).

The FAW is a polyphagous species, being documented on over 100 plants from 27 different families (Pogue 2002). However, using allozymes electrophoresis monitoring, a significant genetic heterogeneity has been observed in FAW

populations that was associated with feeding preferences (Pashley et al. 1985; Pashley 1986). One genetic haplotype was mostly found on corn (*Zea mais*), sorghum (*Sorghum* spp.) and cotton (*Gossypium* spp.) and was named the corn strain (sf-C). Another haplotype was found associated to individuals collected on smaller grasses such as turf, pasture (*Cynodon dactylon*) grasses and rice (*Oryza* spp.), and has been named the rice strain (sf-R) (Pashley 1988). Subsequent studies have confirmed these genetic differences on markers such as the mitochondrial gene cytochrome oxidase c subunit I (COI) (Lu and Adang 1996; Meagher and Gallo-Meagher 2003; Nagoshi et al. 2006; Machado et al. 2008), but also nuclear loci, such as the sex-linked FR1 repeat element (Nagoshi and Meagher 2003a; Nagoshi and Meagher 2003b; Lu et al. 1994) and the Z chromosome-linked *Tpi* gene (Nagoshi 2010). Phylogenetic analyses based on COI only (Dumas et al. 2015a) or on several mitochondrial and nuclear markers (Kergoat et al. 2012) showed that sf-C and sf-R separate in two distinct clades that could represent incipient species. While some degree of hybridization has been reported in field samples (Prowell, McMichael, and Silvain 2004; Nagoshi and Meagher 2003a; Nagoshi et al. 2006; Machado et al. 2008), it has also been shown that pre- and post-zygotic reproductive isolation mechanisms exist between the strains (Groot et al. 2010), with a loss of viability of the hybrids (Dumas et al. 2015b; Kost et al. 2016). Differences in reproductive behavior were also documented, such as the timing of mating being shifted earlier in the night for sf-C compared to sf-R (Schöfl, Heckel, and Groot 2009; Groot et al. 2010; Pashley and Martin 1987; Pashley, Hammond, and Hardy 1992). In order to detect post-zygotic reproductive barriers, many studies tried to quantify the impact of the diet on the general fitness of the FAW larvae (Groot et al. 2010; Roy et al. 2016; Meagher et al. 2004; Silva-Brandão et al. 2017; Pashley 1988; Whitford et al. 1988).

The results of these studies are sometimes contrasted but seem to agree about a better performance of sf-C on corn indicating that sf-C might be specializing to corn (Groot et al. 2010).

We wanted to study the genomic plasticity of sf-C and sf-R strains of FAW when confronted to different plant diets. We first conducted phenotypical experiments in the context of oviposition choice (OV) to different plants and of a reciprocal transplant (RT) during which we surveyed fitness associated traits (also called Life History Traits or LHT; Stearns 2012) to estimate the preference-performance of both strains. Then, in the RT experiment, we compared, at the gene expression level, how both strains adapt to plant feeding during the larval stage. We observed that they don't mobilize the same repertoire of genes when feeding on the same plants. To determine which genes might be linked to adaptive differences, we searched for genes consistently differently transcribed between strains, in laboratory as well as in natural populations. Surprisingly, we identified transcriptional variations in the mitochondrial genome. We make the hypothesis that shifts in host-plant ranges between FAW strains reflect functional differences of mitochondrial haplogroups. We highlight the importance of these mitochondrial differences by the fact that they could be also involved in reproductive isolation in particular by hybrid incompatibilities.

RESULTS AND DISCUSSION

Difference in oviposition choice between sf-C and sf-R

Under the preference-performance hypothesis, the choice of host plants by adult females to lay their eggs should reflect the host plants on which the larval performance is higher (Thompson 1988; Jaenike 1990; Gripenberg et al. 2010; Clark,

Hartley, and Johnson 2011). We conducted an oviposition choice experiment where *S. frugiperda* adult females of each strain (sf-C or sf-R) were set free to lay eggs in a cage containing either their preferred host plant, their alternative host plant ("`no-choice`" trial) or both ("`choice`" trial). We recorded the number of egg masses laid by females in each cage, depending on the substrate (the plant type or the cage net). Analysis by a generalized linear model (see **Methods**) showed that the interaction between the strain and the experimental factors was not significant (LRT, F = 1.29, df = 2, P = 0.1644). Indeed, we found that the number of egg masses laid by females (`Mean fertility`) was similar between trials (LRT, F = 0.29, df = 2, P = 0.75) but significantly different according to the strain (LRT, F = 24.73, df = 1, P < 0.001). Effectively, sf-C laid almost double the number of egg masses than sf-R (`Mean fertility` of 3.89 for sf-C against 2.06 for sf-R across all trials; **Fig. S1A**). When we analyzed the percentage of egg masses hatching within each trial, we observed no significant difference between strains (LRT, χ^2 = 0.17, df = 1, P = 0.68) or laying sites (LRT, χ^2 = 6.39, df = 6, P = 0.38), with 55% to 83% of egg masses in average giving rise to a larva (**Fig. S1B-C**).

By contrast, we observed a striking difference in the distribution of egg masses between the two strains. For each experimental trial ("`choice`", "`corn`" and "`rice`"), sf-C laid between 33% to 52% and sf-R laid almost 85% of their egg masses on the cage net rather than on a plant (**Fig. 1 and Fig. S2**). Neither strain showed a preference for the expected host-plant in female's oviposition choice (*i.e.* corn for sf-C and rice for sf-R). Behavior difference between strains was indicated by the highly significant interaction between strain and laying site in all trials (LRT for maize trial :χ^2 = -68.35, df = 1, P < 0.001; LRT for rice trial :χ^2 = -90.10 ,df = 1, P < 0.001.; LRT for choice trial :χ^2 = -39.53 , df = 2, P < 0.001) . For sf-C, our model shows no difference

in the proportion of egg masses between the net and corn plants in corn trial (LRT, χ^2 = -1.30, df = 1, *P* = 0.25) but did show a significantly (LRT, χ^2 = -20.03, df = 1, P < 0.001) higher number of egg masses on rice plants than on the net in rice trials (**Fig. S2A-B**). For sf-R, in the `no choice` trial, the females laid more eggs on the net than on plants (LRT for maize trial :χ^2 = -83.99, df = 1, *P* < 0.001; LRT for rice trial :χ^2 = -72.95, df = 1, *P* < 0.001; **Fig. S2C-D**). In the `choice` trial, both strains exhibited the same preference pattern. Indeed, the proportion of egg masses for both strains was higher on the net than on corn (sf-C strain :χ^2= -8.2766, df = 1, *P* < 0.01; LRT for sf-R strain :χ^2 = -60.65, df = 1, *P* < 0.001) or on rice (sf-C strain :χ^2= -44.949, df = 1, *P* < 0.001; LRT for sf-R strain :χ^2 = -98.30, df = 1, *P* < 0.001) and lower proportions on rice than on corn (sf-C strain :χ^2= -15.23, df = 1, *P* < 0.001; sf-R strain :χ^2= -7.28, df = 1, *P* < 0.01; **Fig. 1 and S2E-F**).

While these results did not detect a plant host preference for egg laying, behavioral differences between strains were observed, with sf-C laying more egg masses than sf-R, and sf-R placing more egg masses on the cage surface than on plants.

Larval fitness in RT experiment

To test whether different plant diets have an effect on the fitness of *S. frugiperda* larvae, we performed a series of reciprocal transplant (RT) experiments in which larvae freshly hatched of both strains were deposited in cages containing either their current or their alternative host plant. Larvae were allowed to develop on their plants, with the food source being regularly supplied as to avoid deprivation. A control population was reared in parallel on the "Poitout" artificial diet normally used to culture the insects in the laboratory (Poitout and Bues 1974). During the experiment,

we recorded several phenotypic traits: the weight (wt), the developmental stage to measure the time intervals (dt) and the survival (sv).

After hatching, *S. frugiperda* larvae of the first stage (L1) have to undergo five molts to reach their 6th and final stage (L6) prior to metamorphosis. The time intervals between stage (dt) was explained only by the host plant (LRT, χ^2 = -37.41, df = 1, P < 0.001) and there was no strain effect (LRT, χ^2 = -0.93, df = 1, *P* = 0.335; **Fig. S3E-F)**. In sf-C, the larvae took about 11 to 12 days to complete their larval cycle feeding on artificial diet. We obtained the same duration (11 days) with larvae feeding on corn. Remarkably, development of sf-C larvae feeding on rice took 6 to 7 days longer compared to the other diets (**Fig. 2A**). The sf-R larvae took 11 to 13 days after hatching to complete their larval development on corn compared to 17 days for artificial diet and rice (**Fig. 2B**). Finally, both strains exhibited a similar pattern for dt from 1st larval instar to adult emergence, with both strains having a longer dt feeding on rice than on corn (LRT, F = 28.88, df = 1, *P* < 0.0001; **Fig. S3E-F**). Development on corn was similar for both strains (17 days), but sf-R grew faster on rice than sf-C (22 against 24 days, LRT: F = 182.38, df = 1, *P* < 0.0001).

Weight (wt) at the pupal stage was explained by host plant (LRT: χ^2 = -555.25, df = 1, *P* < 0.001), moth strain and sex, with a significant interaction between the last two variables (LRT: χ^2 = -6.61, df = 1, *P* = 0.012). Indeed, we observed, except for sf-C on corn, that males were heavier than females (**Fig. S3A-D**). Both strains had heavier pupae from feeding on corn than feeding on rice (for sf-C: LRT, χ^2 = 67.107, df = 2, *P* < 0.001; for sf-R: LRT, χ^2 =27.18, df = 2, *P* < 0.0001, **Fig. S3A-D**). During the larval cycle, the weight of an individual larva can be measured after five days, when they are approximately at the end of L3 under the temperature conditions of our experiments (**Fig. 2A-B**). For sf-C, larval weight feeding on artificial

diet reached around 500 mg then dropped to 200 mg as pupae and remained unchanged until adult emergence. In comparison, L6 maximum weights feeding on corn and rice were higher, 700 mg and 63 mg, respectively. Pupal weights were higher on corn condition (around 260 mg) than on rice (around 185 mg; **Fig. 2A**). The weight curve of sf-R showed a different pattern. Overall, sf-R larvae and pupae were much lighter than sf-C larvae. In all feeding regimes, the maximum larval weight was between 260 mg and 410 mg, while the pupal weight was between 115 mg and 180 mg. Larvae did best feeding on corn, with higher weight gain than on the artificial diet or on rice (**Fig. 2B**).

The survival (sv) of both strains was linked to the host plant on which the larvae developed. There was a significant interaction to sv between strain and host plant (LRT, χ^2 = -24.22, df = 1, P < 0.0001; **Fig. 2C-D**). The survival of sf-C was significantly greater on corn (about 34%) than on rice (about 7.5%; **Fig. 2C**). However, although sf-R tended to have higher survival on rice (LRT, χ^2 = 2.53, P = 0.11), sv was not significantly different between the two host plants (7.5% on "corn" *vs* 12.5% on "rice"; **Fig. 2D**).

In brief, this analysis indicates that under our laboratory conditions, there is a clear effect of the host plant on the fitness of *S. frugiperda*. Individuals of both strains grew faster and gained more weight feeding on corn than on rice. We observed one major difference between strains, with sf-C surviving better on corn than sf-R, suggesting a specialization of sf-C to corn. However, we didn't find the reciprocal trend for sf-R, which survived equally on both plants.

Gene expression in RT experiment

When confronted with different host plants, polyphagous insects will express a combination of genes that will ensure their optimal fitness. Genes important for

adaptation usually effect chemosensory, digestion, detoxification and immunity processes among others (Simon et al. 2015; Celorio-Mancera et al. 2016). In order to understand how the two *S. frugiperda* strains adapt to different host plants, we extracted RNA from the larvae of the RT experiments. We chose to work on 4th instar larvae, because at this stage larvae survived long enough on their diet and must have expressed the appropriate combination of genes. At this stage they were also still a long way from pupation, which started after the 6th instar; hence it allowed us not to get any interference from the expression of genes involved in the onset of metamorphosis. RNA was used to construct Illumina libraries from the same RT experiment on which LHT were measured, we could perform for each strain two replicates on the corn diet, but only one replicate for the rice diet and one replicate for the artificial diet. We recovered between 30 to 71 million reads per sample (**Table S1**), which we aligned on the OGS2.2 reference transcriptome for sf-C (Gouin et al. 2017) containing 21,778 sequences. The percentages of reads mapped were similar between the two moth strains, with 72.1% to 73.3% of alignments for sf-C under any diet (**Table S1**). For sf-R samples on corn the alignment percentages were similar (71% and 71.2%), and slightly less for the other samples (68.6% on artificial diet and 68.9% on rice; **Table S1**). We used DESeq2 (Love, Anders, and Huber 2014) to identify for each strain, the genes that would change their expression status according to the plant diet type (see below). We extracted the top 50 DE genes in each plant trial and reannotated each one of them carefully using Web-Apollo (Lee et al. 2013) to correct gene models based on RNA-Seq data (**Fig. S4)**, automated blastx homology results and Interpro analysis of the translated peptide sequence (these results are available on the sfrudb platform: http://bipaa.genouest.org/is/lepidodb/spodoptera_frugiperda/). In the cases of some

gene families involved in plant adaptation (*i.e.* Serine Proteases, Odorant receptors and Chemosensory genes, P450s etc…), a manual annotation of each member was previously described (Gouin et al. 2017) and we used it. The top 100 genes responding to corn and responding to rice were treated with BLAST2GO to analyze either the molecular functions (MF) or the biological processes (BP) that were enriched in each trial.

sf-C adaptation to plant diet

For sf-C, we observed among the four samples that the two corn replicates segregated together and opposed to rice and artificial diet replicates (56% of explained variance by PC1; **Fig. 3A**). Less clearly, we observed that the expression state of the artificial diet versus the host plants samples, explains 29% of the variance (PC2 on **Fig. 3A**). At a significant level (*P-adjusted* < 0.05), DESeq2 identifies 1,502/21,778 (6.9%) genes differentially expressed (DE) between the corn and rice conditions, among which 664 are overexpressed on the rice diet and 838 are overexpressed on the corn diet (**Fig. 3B**).

The biological processes (BP) enriched in the rice diet response for sf-C concern mainly metabolic pathways of amino-acids, glycogen and purines (**Fig. 3C**). These genes are likely to be overexpressed when the corresponding nutrients are lacking, as if the rice diet does not provide enough of each nutrient for the larvae to sustain themselves and store resources for their future metamorphosis. However, while overexpressing these metabolic pathways might increase the efficiency of digestion on a poorer diet, it also comes at a cost. Indeed, increasing purine metabolism, for example is necessary to produce purine ribotides when there is a need to increase global transcription levels, but the production from P-ribose is demanding in chemical energy in the form of six ATP (Stincone et al. 2015).

Similarly, the overexpression of glycogen synthase (**Table S2A**) indicates a need for compensation of lower glycogen reserves levels when the larvae are fed on rice diet.

The sf-C response to corn was greatly different. Most of the enriched ontologies in the BP category seem to indicate a better storage efficiency (**Fig. 3D**). Especially, the fatty acid biosynthetic process is used to convert carbohydrates in fatty acids. The "carbohydrate derivative catabolic process" involves the degradation of carbohydrate derivatives produced during the glycolysis. This response indicates that high levels of derivatives are being produced and need to be processed. The "proteolysis" category is represented in the form of Serine Proteases (**Table S2B**) involved in digestion. This enrichment indicates a high digestive activity. This higher digestive and metabolic activity is reflected in the "chitin metabolic pathway" (chitinase and cuticular proteins; **Table S2B**), whose higher expression suggests an increased production of cuticle indicative of a better growth of the larvae. We also noted an increase in detoxification processes in the form of UDP-glycosyl transferases (UGT; Bock 2016) involved in oxidation-reduction of xenobiotics. Finally, we retrieved 8/50 genes belonging to the *osiris* family (**Table S2B**), encoding putative membrane associated proteins, known to play a role in plant adaptation (Whiteman et al. 2012; Hungate et al. 2013; Yassin et al. 2016; Smith et al. 2018).

sf-R adaptation to plant diet

Similarly to sf-C we could perform four RNA-Seq measurements from the RT experiment we described above: two replicates from corn plants, one replicate from the rice plant and one replicate from the artificial diet. After normalization of the four experiments and log transformation of the counts by DESeq2, we also inspected the variance of the data by PCA. We noted that most of the transcriptional variation seems to be associated to the rice diet (82% of the variance is explained by PC1

while 12% by PC2; **Fig. 4A**). After DESeq2 analysis, we identified 1,760/21779 (*P-adj* < 0.05) either 8.1% DE genes in sf-R between the rice diet versus the corn diet, with 1,016 genes overexpressed with rice and 744 genes overexpressed with corn condition (**Fig. 4B**). On the MA-Plot (**Fig. 4B**) it is obvious that the rice diet is attracting most of the transcriptional response in sf-R, conversely to what we observed for sf-C.

In the rice response, we identified 21/50 putative cuticle proteins (**Table S2C**), which are reflected by the enrichment in Gene Ontology (GO) categories for chitin-associated processes (**Fig. 4C**). Again, we identified 7/50 genes of the *osiris* family. Enrichment in 'oxydo-reduction processes' and 'fatty-acid biosynthetic processes' is also observed. This transcriptional response seems to indicate that the sf-R larvae are fabricating cuticle and store fat, which corresponds to a normal growth and development.

In the corn response, we identified a completely different set of genes, with 22/50 most differentially expressed, corresponding to detoxification enzymes (UGT, carboxylesterases, Cytochrome P450 or alcohol dehydrogenase; **Table S2D**). In this trial, there is much less GO enrichment for Biological Processes (**Fig. 4D**), but more categories enriched, mostly involved in catabolism of fatty acids and amino-acids. Taken together, these genes seem to indicate that sf-R is confronted to many more metabolites on the corn diet than on the rice diet, and its response is to overexpress genes to detoxify or assimilate these compounds.

Differences in plant adaptation

The DESeq2 analysis above showed that sf-C and sf-R do not have the same response in the RT experiment, when confronted to the same plant. In fact, there are only 30 and 130 genes in common between sf-R and sf-C that respond to rice and to

corn diet, respectively (**Fig. 5A**). Interestingly sf-R and sf-C share 285 genes in response to their supposedly preferred diet, and 56 in response to their non-preferred diet. When analyzed all the RNA-Seq data from the different RT experiments, we observed a clustering by strain and not by diet (**Fig. 5B**).

In order to better interpret each strain response, we generated for each list of the 50 most DE genes a heatmap of normalized expression across all RT RNA-Seq samples. We could then see that the rice response for sf-C seemed to come from a dual effect, with a clear overexpression of these genes in the rice diet compared to the artificial diet, but also a repression of these genes in the corn diet (**Fig. 5C**). This effect is not seen for sf-R (**Fig. 5C**). However, the sf-C response to corn seems to correspond mostly to a repression of these genes under the rice diet (**Fig. 5D**). Interestingly, these genes seem to have the opposite behavior in sf-R, with an overexpression in the "rice" condition (**Fig. 5D)**. For both strains, this response seems to correspond to a normal digestion coupled with a growth pattern, indicative of a better-adapted state of each strain to its preferred plant.

We retrieved a similar pattern for sf-R but to a lesser extent. The rice diet induces a strong expression of cuticular proteins that we interpreted as reflective of a boosted growth of the larvae (**Fig. 5E**). These genes are also expressed in sf-C on corn compared to rice. So, again, we observed a common response to their preferred diet (**Fig. 5E)**. On the opposite, the sf-R response to corn, can be interpreted mostly as a repression of these genes in the rice diet for both strains (**Fig. 5F**). Since most of these genes correspond to detoxification, we can conclude that both variants respond to corn or to artificial diet (made of corn flour) by expressing their efficient detoxification genes repertoire (Gouin et al. 2017).

So, while the LHT analysis of the RT experiment allowed us to show a slight adaptation of sf-C to corn but no obvious adaptation of sf-R to rice, the transcriptional analyses reveal that, at least at larval stage L4, both strains have the same transcriptional response to their preferred plant: corn for sf-C and rice for sf-R.

Constitutive transcriptional differences between sf-C and sf-R

When we combine all the RNA-Seq data from the different RT experiments, we observed that the samples cluster by strain (29% of explained variance on PC2; **Fig. 6A**), suggesting there may be fundamental differences between sf-C and sf-R that could explain their plant preferences. However, this observation was contrasted by PC1, which explained 53% of the variance and revealed a pattern of separation by preferred diets. Indeed, an important part of the variance was explained by the sample sf-R on rice, clustering with sf-C on corn (**Fig. 6A**). We used DESeq2 to identify consistent differences between the two strains regardless of the diet trial. We identified 1,697 (7.8%; *p.adj* < 0.05) genes overexpressed in sf-R compared to sf-C and 2,016 (9.3%; *p.adj* < 0.05) genes overexpressed in sf-C compared to sf-R (**Fig. 6B**). We verified by q-PCR on independent samples raised on artificial diet that this strain-specific difference of expression is stable. Out of 50 genes selected to be overexpressed in sf-R compared to sf-C in our RNA-Seq experiments (**Fig. S5**), all except one (peroxidase), were systematically overexpressed in sf-R when measured by qPCR (**Table S3**).

The GO enrichment analysis did not detect any significant enrichment of either BP or MF terms in both gene lists. sf-R expresses some enzymes involved in digestion, metabolism and detoxification as well as, intriguingly, ribonucleoproteins involved in mRNA splicing (**Table S2E**) but no coherent pattern emerges. While no GO enrichment has been observed for sf-C, manual re-annotation of the 50 most

expressed genes showed that at least 13/50 genes correspond to transposable elements (TE) (**Table S2F**). Other genes encode putative endonucleases that could also be of TE origin, such as the Harbinger transposase-derived nuclease, HARBI. In addition, we could not find evidence for gene annotation by homology or protein domain analysis for 16/50 genes. Other genes encode proteins that could be linked to plant adaptation. For example, sf-C shows a strong expression of fatty acid synthase, suggesting that sf-C is constitutively more efficient at energy production and storage. We also found two peptidases, and the cytochrome P450: CYP9A31 indicating inherent digestive and detoxification potential for sf-C. While we have detected no transcriptional regulators in our plant adaptation datasets, we could at this time detect one important transcription factor (TF), expressed only in sf-C: *apterous-1*. This homeodomain (HD)-containing TF is known in Drosophila to be involved in wing development (http://flybase.org/reports/FBgn0267978.html). Annotation of HD genes in *Spodoptera* (Gouin et al. 2017) showed that *apterous* has two paralogs, suggesting a yet-to-be-determined potential shift in function for this TF. Finally, we detected overexpression of a small genomic sequence corresponding to a fragment of the mitochondrial gene cytochrome oxidase c subunit III (COIII). Genomes often contain insertions of mitochondrial sequences (Hazkani-Covo, Zeller, and Martin 2010). Such insertions are termed *numts*. Around 95 *numts* can be identified in the *Spodoptera frugiperda* genomes. They sometimes confound gene prediction because they contain the open reading frame (ORF) sequence of the original mitochondrial gene. However, *numts* are usually not transcribed, lacking the promoter region sequence. In the case of the COIII-*numt*, the differential expression we measured comes from polyadenylated RNAs of mitochondrial origin, whose reads also align on the *numt* region. In practice, *numts* show differences of expression at

the level of mitochondria. Here, it would suggest that sf-C has a major difference in energy production at the mitochondrial level compared to sf-R.

Exploration of strain transcriptional differences in natural populations

We wanted to know if the transcriptional differences between *S. frugiperda* strains measured in the laboratory conditions can also be observed in the wild. We performed a field collection of FAW larvae in a sweet corn field (Citra, FL), in a volunteer corn field (Tifton, GA) and in a pasture grass field (Jacksonville, FL). We performed both DNA and RNA extractions from individual L4 larvae. DNA was used to genotype the individuals (see **Methods**). Based on the detection of mitochondrial Cytochrome Oxidase I polymorphism (Nagoshi et al. 2006), the Citra corn field contained 32/33 sf-C associated genotypes, the Tifton corn field contained 14/18 sf-C strains and the Jacksonville field contained 6/6 sf-R strains (**Fig. S8)**. We selected some sf-R and sf-C individuals from each field to genotype according to one SNP on the *Tpi* gene located on the Z chromosome (Nagoshi 2010) and presence of the FR1 repeat (Lu et al. 1994; Nagoshi and Meagher 2003a). Interestingly, most sf-R haplotypes recovered from corn fields seem to be hybrids from a sf-R mother. We didn't detect any potential hybrids in the pasture grass field (**Fig. S9-S10**).

From the 20 most differentially expressed genes between sf-C and sf-R on corn, we selected 15 genes to perform qPCR measurements of their expression in individual L4 larvae from the laboratory strains raised the artificial diet as well as in individual L4 larvae from the Tifton field where we recovered both sf-C and sf-R mitochondrial haplotypes. The qPCR analysis showed that the genes we selected from RNA-Seq studies are concordantly differentially expressed between laboratory strains. However, for the genes we selected, we detected no difference in expression between natural populations of sf-C and sf-R (**Fig. S11**). This result seems to

indicate that studies of plant adaptation in laboratory conditions might not be directly applicable to natural conditions. Indeed, while in laboratory conditions, we can control the genetic background of insects, the environmental conditions, the plant types and supply, natural populations experience many more variables. Their genetic background might be different from one another, they may be infected or parasitized, they may be stressed by climate conditions, predators, competitors or parasites. In these conditions, to identify transcriptional differences between strains, one might want to turn to RNA-Seq experiments, which allow interrogating all genes at once.

Transcriptomic studies of natural *S. frugiperda* populations

We thus decided to perform RNA-Seq experiments on 3 sf-C individuals from Tifton, 3 sf-R individuals from Tifton and 3 sf-R individuals from Jacksonville (**Fig. 7A**). We recovered between 23 to 74 million reads per sample (**Table S1**) with alignment percentages ranging from 45.32% to 58.40%, slightly less than in laboratory experiments. On a PCA analysis of FL15 only, replicates of the same "`trial + strain`" individuals group well together with the FL15_B1J being slightly outlier (**Fig. S12A-B**). When integrating all FL15, and RT experiments, it becomes impossible to group together all Sf-C genotypes independently of trials (**Fig. S12C**). Moreover, when we looked at the expression of the 50 most differentially expressed genes in sf-R versus sf-C in RT2 experiments (**Table S2E,F**), and observed the expression of these genes in two independent RT experiments RNA-Seq from our laboratory (RT1), a previously published study on the midgut Roy-RT (Roy et al. 2016) and the FL15 natural populations, we observed that most transcriptional response detected in RT2 was not recapitulated in the other experiments (**Figs. S13-S14**).

Conserved expression differences between strains in laboratory and wild populations.

To see to what extant the response to diet in the lab and in the fields are concordant, we analyzed the transcriptional response of candidate gene families. Indeed, polyphagous noctuids seem to have evolved large repertoires of several gene families such as gustatory receptors and P450s (Pearce et al. 2017; Xu et al. 2016; Gouin et al. 2017; Cheng et al. 2017). We focused on one family that seemed to respond to the plant diet differently according to the strain: the *osiris* gene family, a cluster of secreted proteins found conserved and syntenic in many insect genomes (Shah et al. 2012) which are known to be involved in plant response in herbivorous insects (Hungate et al. 2013; Whiteman et al. 2012; Yassin et al. 2016). In our laboratory experiments, the *osiris* genes are overexpressed in larvae fed on their preferred plant (**Fig. 7B**). After a careful manual reannotation of the *osiris* genes in the *S. frugiperda* genome (**Fig. S15**), we looked at the expression of this gene family in the other RNA-Seq datasets at our disposal. In RT1, we confirm the overexpression of many *osiris* genes in sf-R larvae reared on rice compared to sf-R larvae reared on corn. We observed the `strain + diet` interaction in datasets from RoyRT, where *Osi9* paralogs are overexpressed in the sf-R strain reared on corn only. This inverted response is also observed in individuals from the wild where the same set of *osiris* genes is overexpressed in sf-R individuals found in the Tifton corn field. No expression is observed in sf-C individuals from the same field nor from sf-R individuals from the pasture grass field in Jacksonville (**Fig. 7B**). The fact that in four different RT experiments we detected the *osiris* genes reacting differently in function to the strain and the diet make them important candidates to plant adaptation in polyphagous animals. However the precise mechanism of this *osiris* response is still puzzling since it reacts in opposite ways: either to the preferred

plants in our laboratory experiments, or to the non-preferred plant in Florida fields and in a Swedish laboratory.

These discrepancies are not observed for every gene family. For example, we also reannotated the *S. frugiperda* Serine Proteases (Gouin et al. 2017) and surveyed their expression across all the above mentioned RT-RNA-Seq experiments. We can observe (**Fig. 7C**) that the expression of Serine proteases is concordant across all experiments. Some genes of this repertoire are constitutively highly expressed or constitutively not expressed. Only a few of them seem to be regulated in function to the diet or the strain. Based on their differential expression, we can easily group together experiments performed on the midgut (RoyRT) apart from the ones performed on whole larvae (**Fig. 7C**). We could also group together experiments performed on artificial diet. The sf-R strain specific response also tends to cluster away from the sf-C response in all cases.

Altogether these data suggest that laboratory controlled RT experiments are not sufficient to decouple the parameters triggering the adaptation to plant. Comparing gene families response across independent experiments, including natural populations and applying clustering methods seem to be a more efficient method in order to associate gene response to the different interacting factors: genetic background, type of diet, tissue and experimental conditions.

Strain specific expression in laboratory and in field collections

Having large datasets to explore allowed us to ask a simple question: what are the genes whose expression is specifically restricted to one strain compared to the other? We performed a differential expression analysis across our laboratory RT experiment and our FL15 collection to identify these genes. We found 76 genes

consistently overexpressed in sf-R compared to sf-C and 73 genes overexpressed in sf-C compared to sf-R (**Fig. 7D**). To verify the validity of these genes we again surveyed their expression across all the RT-RNA-Seq data at our disposal. We could see that for the majority of these genes their strain specific overexpression is confirmed in the different laboratory populations as well as in natural populations (**Fig. 7E** and **Supplementary Fig. 16**). While many genes in this list have functions of potential interest to study the molecular basis of ecological speciation, we noticed a peculiar outlier corresponding to the previously mentioned *numts* (**Fig 7E**). What these numts reveal are parts of the mitochondrial genome that are differentially expressed according to the strain. Two *numts* in particular, corresponding to fragments in the mitochondrial genes COI and COIII are clearly differentially expressed in sf-C compared to sf-R in all the RNA-Seq datasets we analyzed (**Supplementary Fig 17**). To rule out any effect of genome misassembly, we amplified both *numts* and mitochondrial sequence for COI and COIII and sequence them. We could confirm the presence of these *numts* within the genome of sf-C and sf-R strains with a sequence slightly different than the one from mitochondria. To rule out any sequence specific alignment bias, we retrieve from NCBI the reference genome sequence from *S. frugiperda* mitochondrion (accession KM362176.1) and realigned our RNA-Seq data on it. It was obvious that, in the regions corresponding to *numts*, there was a clear underexpression in the sf-C strain (**Fig. 7F**). The implication of this result on the metabolism of the larvae remains to be established, but nevertheless, it explains why the mitochondrial haplotypes in the COI gene are the principal marker for strain discrimination. It may very well be that a difference in energy production between these two strains was linked at some point of their evolutionary history to a shift in host plant preference.

CONCLUSION

In this study, we wanted to determine if the adaptation to host plant diet was the basis for the differentiation of *S. frugiperda* in two strains: sf-C and sf-R. First, we measured a combination of Life History Traits in the context of an oviposition preference experiment (OV) and of a reciprocal transplant (RT) experiment in controlled environments to characterize the specialization to host plants. Then we coupled our LHT measurements to gene expression variations of L4 larvae during the RT experiments. Finally, we compared our expression data to gene expression of sf-C and sf-R strains larvae in natural populations on different fields.

From this set of experiments, we concluded that the LHT of our laboratory colonies are consistent with a specialization of sf-C to corn, but does not provide evidence that rice is the preferred plant for sf-R, which showed only a slight trend to survive better on this plant than on corn. Interestingly, however, RNA-Seq experiments show that both strains express a similar set of genes, involved in growth and nutriment storage, when confronted to their main host-plant (corn for sf-C and rice for sf-R). This similarity in the transcriptional responses suggests that rice is indeed recognized as a suitable host for sf-R but maybe not its most preferred one.

In contrast, the molecular response of the two strains to their non-preferred host-plant differs. While sf-C seems to compensate a potential lack of nutriment in the rice plant by overexpressing secondary metabolites pathways, sf-R mainly responds to corn by overexpressing genes involved in detoxification, potentially of corn xenobiotic compounds. At the gene expression level, the difference between strains seems to be associated to the plant response and thus could support the ecological speciation model hypothesis if there was a link between traits under divergent selection and reproductive isolation. This led us to investigate whether

there were constitutive gene expression differences between the two strains that could explain this link.

We found several candidate genes that are differentially expressed between the strains regardless of the diet. However, when we looked at natural populations, almost none of these genes were differentially expressed between strains. But by combining the analysis of RNA-Seq data from laboratory populations as well as from natural populations, we detected a narrower set of genes constitutively differentially expressed between strains. Among those, one candidate stood out and turned out to be the mitochondrial gene COI. This gene has been used as a genetic marker for strain identification. The fact that it is also consistently differentially expressed may indicate that the COI gene, and potentially other mitochondrial genes, may be the original target of selection between the strains (Meiklejohn, Montooth, and Rand 2007). Changes in mitochondrial functions are associated to changes in energy demand or supply (Jose et al. 2013). In addition, variations in mitochondrial sequences can be the cause of mitonuclear incompatibilities between species (Hill 2015). The evolution of mitonuclear interactions can maintain the segregation of various mitochondrial haplotypes in the context of ecological speciation (Morales et al. 2016). These features are consistent with a model of ecological speciation for *S. frugiperda*, in which divergence in mitochondrial functions have been selected on plants with different nutritive values. For example, the sf-C haplotype, which has a lesser expression of mitochondrial genes might have a reduced energy production efficiency compared to sf-R. This reduced efficiency may be compensated by the higher nutritive value of the corn plant. Consistent with this explanation, we found sf-R haplotype in corn fields but almost no sf-C haplotype on pasture grass fields. Alternative explanations might involve adaptation to the redox state imposed by the

host-plant xenobiotic compounds. Several insect proteins such as UGTs and P450s catalyze oxidation-reduction reactions to resist against these natural pesticides. Consistent with this second hypothesis, we also detected plastic and evolved differential expression of several P450 proteins (**Table S2**). Finally, it is possible that variations in mitochondrial function reflect variations in energy demand associated with the different field environments. Indeed, corn plants, especially the hideouts within the whorl or the ear, may also provide more protection against competitors, predators and parasites than grass lands, which are more open spaces. Thus sf-R strain, that has a higher level of expression in mitochondrial genes might require more energy to move around. Consistent with this explanation, sf-R larvae are consistently smaller than sf-C larvae (**Supplementary Fig. 3A-D**). Energy consumptions at adult stage, especially regarding migratory capacities should also be considered.

Compared to other studies using a similar RT experimental design to identify adaptation genes or evolved genes in *Spodoptera frugiperda*, our study highlighted one important point that could explain the inconsistencies observed over the years in the determination of the plant adaptation process in *S. frugiperda*. Traditionnally, two different RT strategies were used, either by using colonies from natural populations or long maintained laboratory colonies and each approach has its pros and cons. Working with laboratory colonies allows one to control for genetic background variations as well as environmental conditions. But in turn, they might be subject to genetic drift or adaptation to the artificial diet used to maintain them. Here, we show that by combining the two approaches, we revealed a smaller set of genetic events that could explain the differentiation of the two strains. In particular, we identified COI as both a genetic marker and a selected locus between the two strains. The

consequences of functional variations in the mitochondrial genome on the shift of host-plant range range in *S. frugiperda* remains to be elucidated.

Material and Methods

Biological material: Moths and Plants

We used individuals from the two strains of *S. frugiperda*: corn (sf-C) and rice strain (sf-R). Those strains were seeded with 30 to 50 pupae ten and four years ago for the corn and rice strain, respectively. From the time of their collection they have been reared under laboratory conditions on artificial diet (from Poitout et al. 1972, principal components: 77% H_2O, 2% Agar-agar, 13% maize flour, 6% other nutrients, 1% vitamins; 1% antibiotics), at 24°C with a 16h:8h Light:Dark photoperiod (L:D) and 70 % Relative Humidity (R:H). The individuals that seeded the corn strain came from French Guadeloupe whereas those that founded the rice strain came from Florida (U.S.A.).

Corn (Corn line B73) and rice (Arelate variety from CFR, Centre Français du Riz) were produced from organic seed at the DIASCOPE experimental research station (INRA, Mauguio, France, 43°36'37"N, 3°58'35"E) in plastic pots (7 x 8cm for both plants in RT and 6L plastic pots for maize in OV) filled with conventional substrate. Corn and rice cultivation was carried out in a warm chamber at 25°C 2, 60% RH and 16:8 h (L:D) under organic conditions. Corn and rice plants were used 15 days or a month after seeding, respectively, to have an equivalent of two biomass plants.

Experimentation

Experimental trials

Spodoptera frugiperda is not present in France and considered as a quarantine pest. Consequently experiments on this study model are regulated. Our experiment described hereafter was conducted in confined environment on insect quarantine platform (PIQ, University of Montpellier, DGIMI laboratory).

Oviposition experiment

The oviposition (OV) experiment consisted in release of 12 to 20 virgin females and males of the same strain per cage, and for three nights (72 hours) in three different set-ups: *choice*, *corn-only* and *rice-only*. For the choice modality, each cage contained five maize plants and 15 rice plants (the number of maize and rice were adjusted to provide an equivalent biomass) arranged in two patches in two opposite corners of the cage. For the rice- and corn-only modalities, we used either 10 maize or 30 rice plants. Plants were arranged in two equal patches (2 x 5 maize or 2 x 15 rice) located in two opposite corners of each cage. The experiment was conducted in insect rearing cages covered by an insect-proof net (175 x 175 cm) and 4 replicates of each set-up were done under the same climatic conditions.

In each cage, at the end of the third night, all egg masses were counted and immediately individualized. We measured three variables for each cages:

(1) The number of egg masses laid by females in a given cage (on plants and on the net) to measure the fecundity. As the adult number was not similar in cages, it was important to balance the number of egg masses per the number of females in the cage. Indeed, the number of adults had a significant effect on the egg masses number ($P < 0.01$), so we decide to create a variable, *Mean Fecundity*, which take account the egg masses number divided by the

number of females in the replicate. The following variables were the strain (sf-C and sf-R) and the trials (choice, rice-only, corn-only).

(2) The proportion of egg masses laid by females on one particular site (one given plant species or the net). This percentage was calculated in three set-ups to estimate the preference of each moth species according to present substrates in the cage. We performed the analysis on each set-up independently with two following factors, the strain and the oviposition site.

(3) The hatching proportion is the number of egg masses hatching on one particular site (one given plant species or the net) whatever the set-up. This percentage provides an estimate of the fertility of both strains according to the choice of oviposition site by the females. The following factors are the strain and the oviposition site (nested in set-up).

Reciprocal transplant experiment

The reciprocal transplant (RT) experiment consisted in controlled infestations of corn and rice plants with first instar larvae in 8 insect rearing cages (32.5 x 32.5 cm) covered by an insect-proof veil to prevent contaminations and escapes in the incubator (24°C, 16h:8h L:D cycle and 70% R. H.). The RT experiment was conducted in the same incubator for four modalities: 1) corn plants infested by sf-C (native condition); 2) rice plants infested by sf-C (alternative condition); 3) corn plants infested by sf-R (alternative condition); 4) rice plants infested by sf-R (native condition). We realized two replicates by modality. Each cage contained four corn or rice pots, which were changed before the 4^{th} larval instar and each day after this instar until the pupation. We deposited in each cage 80 larvae hatched the morning of the experiment initiation (larvae obtained from several egg-masses provided from breeding). Two generations have been conducted on plants; during the first

generation we measured life history traits (LHT) for each strain in native and alternative conditions and during the second generation, the larvae had been sampled at 4th larval instar for a RNA-Seq analyzes.

As of the 2nd larval instar, we measured several LHT every other day until pupartion, during which we determined the sex of each individual. In addition, at each counting, we determined the larval stage by the width of the head capsule. To limit the possible contamination between strains, we isolated two floors of the incubator with an insect proof net (150 µm) and to avoid a floor and edge effect, rotations between floor were conducted and cages were randomly deposed after counting. We measured three variables:

- Survival (`sv`) is the number of emerging adults counted over the initial number of larvae;
- Developmental time (`dt`) is the number of days between the beginning of the experimental start until adult emergence (mean on all emerging adult in same cage);
- The weight (`wt`) of individual larvae and of individual pupae of each sex in mg.

For all variables from RT, we analyzed by following factors: the strain (sf-C or sf-R) and the host plant (corn or rice). Replicate effect was negligible.

Statistical analysis of LHT

All computations were performed using “lme4” package (Bates et al. 2015) of the R software version 3.0.3. We used different generalized linear models depending on the distribution of the residuals. For all the variables, we analyzed by following factors and we also included the interaction between the following factors. If the replicates had a negligible influence on model outcome, they were not included in the models

(using “glm” function), or if the replicates had a significant effect, they were added as a random factor (using “glmer” with replicate factor in random effect). Model selection was performed as follows: the significance of the different terms was tested starting from the higher-order terms using likelihood-ratio-tests (LRT). Non-significant terms ($P > 0.05$) were removed and factor levels of qualitative variables that were not significantly different were grouped (LRT; Crawley 2007).

Genomic

Sample preparation and sequencing

We collected 4th instar larvae of the second generation on native and alternative plants. The larvae number was variable between experimental set-ups (n = 3 to 12 larvae). Larval instar was determined by the width of the head capsule (**Supplementary Figure 18**), if the larvae were considered like 4th instar, three larvae of the same experimental set-up were pooled. We weighed the pools and crushed them in liquid nitrogen to obtain a fine powder, which was placed in TRIzol® Reagent (Invitrogen) and stored at -80°C. After collection of samples in all experimental set-ups, total RNA was extracted using a TRIzol® Reagent, according to the manufacturer’s RNA protocol. To remove contaminating DNA from RNA preparations, we used DNase from TURBO DNA-free™ Kit (Ambion). qRNA quantity and quality were assessed using a Bioanalyzer and 1 µl of total RNA from each sample of three larvae. The sub-pools of three larvae, having a good quality and quantity, were pooled again to obtain samples corresponding to the 2 biologic replicates of the 4 experimental set-ups.

High throughput sequencing was performed for the pool samples using Illumina technologies to obtain single-end 50-bp reads. Library construction and

sequencing were performed by MGX-Montpellier GenomiX (Montpellier, France) on a HiSeq 2000 (Illumina). For each pool, tagged cDNA libraries were generated using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina) following manufacturer's protocol.

Reference and annotation

All RNAs-Seq experiments were aligned against a common reference. This reference is OGS2.2 (Gouin et al. 2017), generated from the sequencing and annotation of the C-strain genome. Gene models result from direct ORF prediction, guided by expression data published earlier (Legeai et al. 2014) and the mapping of RNA-Seq reads. Gene models for selected gene families also underwent an expert annotation by manual curators.

Differential expression analysis

To identify differentially expressed genes, we first mapped reads on gene prediction using Bowtie2 (Langmead and Salzberg 2012). We chose to use the same reference for both the sf-C and the sf-R strain samples. For read mapping we used "very sensitive" parameter setting in Bowtie2, which allowed searching extensively for the best alignment for each read. Counting of aligned reads number to each gene is produced by SAMtools program (Li et al. 2009). Then to detect the genes differentially expressed we used DESeq2 (R package; Love, Anders, and Huber 2014). To measure gene expression variations between conditions, DESeq2 uses a negative binomial generalized mixed model. The estimates of dispersion and the logarithmic fold-changes incorporate data-driven prior distributions. Genes were considered differentially expressed if they satisfy a false discovery rate lesser than 1%.

Characterizing gene function and comparison between two strains

After identifying differentially expressed genes between two strains for the same food resource, we used the Fisher's exact test (cut-off of FDR < 0.10) to identify GO categories possibly involved in corn specialization. The resulting list of GO-terms may contain redundant categories (*i.e.* there was a parent-child relationship in enriched function or process). We used REVIGO (http://revigo.irb.hr/) that summarizes and regrouped terms by a clustering algorithm based on semantic similarities (Supek et al. 2011). We used the default parameter ("medium").

Natural Populations collections

Spodoptera frugiperda wild larvae were collected in Florida and Georgia between September, 18th and September, 25th 2015 in three different field locations. One sweet corn field in Citra (Marion County, Florida), one volunteer corn in Tifton, (Tift County, Georgia) and one pasture grass field in Jacksonville (Duval County, Florida). In corn fields, plants were cut and larvae collected in situ. In the pasture grass field, collections were made using a sweeping net. After confirming their identification as *Spodoptera frugiperda* according to LepIntercept (http://idtools.org/id/leps/lepintercept/frugiperda.html), larvae were placed in individual plastic cups with cut leaves (either corn or grass) as a food source and brought back in a cooler to the laboratory after a few hours of collection. Once in the laboratory, larvae were sorted according to stage. Stages were measured according to the chart in **Supplementary Figure 18**, where the width of the cephalic capsule should match the width of the line for each stage. This chart has been determined based on rearing conditions of lab strains in Montpellier and confirmed with a similar chart based on the rearing of lab strains in Gainesville, Florida. L4 larvae were sacrificed with a razor blade and immediately placed in a screw-cap 2ml tube

containing 1ml of RNAlater (Sigma; R0901). Younger larvae were placed again in individual cups with fresh cut leaves and were grown in a 25°C incubator until they reached the L4 stage.

DNA/RNA extractions

Larvae from field collections were placed in a 1.5ml Eppendorf tube with RLT buffer from Qiagen. Larvae were ground using a TissueLyser II from Qiagen (Cat No./ID: 85300) using one bead (size 5mm) by tube and processed for dual DNA and RNA extraction using an AllPrep DNA/RNA Mini Kit (50) (Qiagen Cat. 80204).

Genotyping

We used the COI genotype described in (Meagher Jr. and Gallo-Meagher 2003) to discriminate between the sf-C and the sf-R strains. A PCR on genomic DNA was performed using the following primer sequences (F : ATC ACC TCC ACC TGC AGG ATC and R : GAG CTG AAT TAG GGA CTC CAG G) to amplify a DNA fragment of 550bp corresponding to the mitochondrial cytochrome oxidase c subunit I. The MspI enzyme is used to reveal a polymorphism between the 2 strains. The COI fragment of the C-strain is digested by MspI to produce a 500bp and a 50bp fragment (**Supplementary Figure 8A**).

Quantitative PCR

qPCR have been performed on a LightCycler 480 (Roche) with SYBR green. Program used was 95°C for 10min and then 40 cycles of 94°C 10s, 60°C 10s, 72°C 10s. Relative expression was calculated using the $\Delta\Delta$Ct method with the laboratory sf-C strain as a reference point for each gene.

Data availability

Spodoptera frugiperda reference genome and reference transcriptome can be publicly accessed *via* the BIPAA (BioInformatics Platform for Agroecosystem Arthropods) interface (http://bipaa.genouest.org/is/lepidodb/spodoptera_frugiperda/). fastq files and RNAseq counts from this study are accessible in ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) with the following accession number : E-MTAB-6540.

Acknowledgments

This work was partially supported by a grant from Institut Universitaire de France for N.N. and by a grant from the French National Research Agency (ANR-12-BSV7-0004-01; http://www.agence-nationale-recherche.fr/) for E. A. including a post-doctoral fellowship for Y.M. We thank the quarantine insect platform (PIQ), member of the Vectopole Sud network, for providing the infrastructure needed for pest insect experimentations. We are also grateful to Clotilde Gibard and Gaëtan Clabots for maintaining the insect collections of the DGIMI laboratory in Montpellier.

Authors Contributions

NN, EA and MO designed the project. JPB and MV produced the corn and rice plants used in the RT experiments. MO, PA, AP, NN and performed the RT and OV experiments. MO, GD performed the statistical analyses of LHT in the RT experiments. MO, GD, RNS and NN performed the RT-qPCR experiments. MO performed the RNA extractions for the RNA-Seq experiments. RK and SR produced the Illumina libraries, performed the Illumina sequencing and realized the computational analyses and quality control necessary to produce .fastq files of sequences. MO, YM, SN and NN performed the RNA-Seq analyses. MF, GJK, RNN,

RLM and NN performed the field collections. RNS and NN performed the genotyping and RNA extractions of field samples. MO and NN wrote the manuscript and produced the figures. YM, SR, GJK, RNN, RLM and EA edited the current manuscript. All authors approved the present manuscript submission.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 51. doi:10.18637/jss.v067.i01.

Bock, Karl Walter. 2016. “The UDP-Glycosyltransferase (UGT) Superfamily Expressed in Humans, Insects and Plants: Animal-Plant Arms-Race and Co-Evolution.” *Biochemical Pharmacology*. doi:10.1016/j.bcp.2015.10.001.

Celorio-Mancera, Maria De La Paz, Christopher W. Wheat, Mikael Huss, Francesco Vezzi, Ramprasad Neethiraj, Johan Reimegård, Sören Nylin, and Niklas Janz. 2016. “Evolutionary History of Host Use, rather than Plant Phylogeny, Determines Gene Expression in a Generalist Butterfly.” *BMC Evolutionary Biology* 16 (1). BMC Evolutionary Biology. doi:10.1186/s12862-016-0627-y.

Cheng, Tingcai, Jiaqi Wu, Yuqian Wu, Rajendra V. Chilukuri, Lihua Huang, Kohji Yamamoto, Li Feng, et al. 2017. “Genomic Adaptation to Polyphagy and Insecticides in a Major East Asian Noctuid Pest.” *Nature Ecology & Evolution*. doi:10.1038/s41559-017-0314-4.

Clark, Katherine E., Susan E. Hartley, and Scott N. Johnson. 2011. “Does Mother Know Best? The Preference-Performance Hypothesis and Parent-Offspring Conflict in Aboveground-Belowground Herbivore Life Cycles.” *Ecological Entomology* 36 (2): 117–24. doi:10.1111/j.1365-2311.2010.01248.x.

Crawley, Michael J. 2007. *The R Book*. *The R Book*. doi:10.1002/9780470515075.

Dumas, Pascaline, Jérôme Barbut, Bruno Le Ru, Jean François Silvain, Anne Laure Clamens, Emmanuelle D’Alençon, and Gael J. Kergoat. 2015. “Phylogenetic Molecular Species Delimitations Unravel Potential New Species in the Pest Genus Spodoptera Guenée, 1852 (Lepidoptera, Noctuidae).” *PLoS ONE* 10 (4). doi:10.1371/journal.pone.0122407.

Dumas, Pascaline, Fabrice Legeai, Claire Lemaitre, Erwan Scaon, Marion Orsucci, Karine Labadie, Sylvie Gimenez, et al. 2015. “Spodoptera Frugiperda (Lepidoptera: Noctuidae) Host-Plant Variants: Two Host Strains or Two Distinct Species?” *Genetica* 143 (3): 305–16. doi:10.1007/s10709-015-9829-2.

Goergen, Georg, P. Lava Kumar, Sagnia B. Sankung, Abou Togola, and Manuele Tamo 2016. “First Report of Outbreaks of the Fall Armyworm Spodoptera Frugiperda (J E Smith) (Lepidoptera, Noctuidae), a New Alien Invasive Pest in West and Central Africa.” *PLoS ONE* 11 (10). doi:10.1371/journal.pone.0165632.

Gouin, A., A. Bretaudeau, K. Nam, S. Gimenez, J.-M. Aury, B. Duvic, F. Hilliou, et al. 2017. “Two Genomes of Highly Polyphagous Lepidopteran Pests (Spodoptera Frugiperda, Noctuidae) with Different Host-Plant Ranges.” *Scientific Reports* 7 (1). doi:10.1038/s41598-017-10461-4.

Gripenberg, Sofia, Peter J. Mayhew, Mark Parnell, and Tomas Roslin. 2010. “A Meta-Analysis of Preference-Performance Relationships in Phytophagous Insects.” *Ecology Letters*. doi:10.1111/j.1461-0248.2009.01433.x.

Groot, Astrid T., Melanie Marr, David G. Heckel, and Gerhard SchÖfl. 2010. “The Roles and Interactions of Reproductive Isolation Mechanisms in Fall Armyworm (Lepidoptera: Noctuidae) Host Strains.” *Ecological Entomology* 35 (SUPPL. 1): 105–18. doi:10.1111/j.1365-2311.2009.01138.x.

Hazkani-Covo, Einat, Raymond M. Zeller, and William Martin. 2010. “Molecular Poltergeists: Mitochondrial DNA Copies (Numts) in Sequenced Nuclear Genomes.” *PLoS Genetics*. doi:10.1371/journal.pgen.1000834.

Hill, Geoffrey E. 2015. “Mitonuclear Ecology.” *Molecular Biology and Evolution* 32 (8): 1917–27. doi:10.1093/molbev/msv104.

Hungate, Eric A., Eric J. Earley, Ian A. Boussy, David A. Turissini, Chau Ti Ting, Jennifer R. Moran, Mao Lien Wu, Chung I. Wu, and Corbin D. Jones. 2013. “A Locus in Drosophila Sechellia Affecting Tolerance of a Host Plant Toxin.” *Genetics* 195 (3): 1063–75. doi:10.1534/genetics.113.154773.

Jaenike, J. 1990. “Host Specialization in Phytophagous Insects.” *Annual Review of Ecology and Systematics*. doi:10.1146/annurev.es.21.110190.001331.

Jeger, Michael, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, et al. 2017. “Pest Categorisation of Spodoptera Frugiperda.” *EFSA Journal* 15 (7). doi:10.2903/j.efsa.2017.4927.

Jose, Caroline, Su Melser, Giovanni Benard, and Rodrigue Rossignol. 2013. “Mitoplasticity: Adaptation Biology of the Mitochondrion to the Cellular Redox State in Physiology and Carcinogenesis.” *Antioxidants & Redox Signaling* 18 (7): 808–49. doi:10.1089/ars.2011.4357.

Kergoat, G J, D P Prowell, B P Le Ru, A Mitchell, P Dumas, A L Clamens, F L Condamine, and J F Silvain. 2012. “Disentangling Dispersal, Vicariance and Adaptive Radiation Patterns: A Case Study Using Armyworms in the Pest Genus Spodoptera (Lepidoptera: Noctuidae).” *Molecular Phylogenetics and Evolution* 65 (3): 855–70. doi:10.1016/j.ympev.2012.08.006.

Kost, Silvia, David G. Heckel, Atsuo Yoshido, František Marec, and Astrid T. Groot. 2016. “A Z-Linked Sterility Locus Causes Sexual Abstinence in Hybrid Females and Facilitates Speciation in Spodoptera Frugiperda.” *Evolution* 70 (6): 1418–27. doi:10.1111/evo.12940.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. doi:10.1038/nmeth.1923.

Lee, Eduardo, Gregg a Helt, Justin T Reese, Monica C Munoz-Torres, Chris P Childers, Robert M Buels, Lincoln Stein, Ian H Holmes, Christine G Elsik, and Suzanna E Lewis. 2013. “Web Apollo: A Web-Based Genomic Annotation Editing Platform.” *Genome Biology* 14 (8): R93. doi:10.1186/gb-2013-14-8-r93.

Legeai, Fabrice, Sylvie Gimenez, Bernard Duvic, Jean-Michel Escoubas, Anne-Sophie Gosselin Grenet, Florence Blanc, François Cousserans, et al. 2014. “Establishment and Analysis of a Reference Transcriptome for Spodoptera Frugiperda.” *BMC Genomics* 15 (1): 704. doi:10.1186/1471-2164-15-704.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79. doi:10.1093/bioinformatics/btp352.

Love, M. I., Simon Anders, and Wolfgang Huber. 2014. *Differential Analysis of Count Data - the DESeq2 Package*. *Genome Biology*. Vol. 15. doi:110.1186/s13059-014-0550-8.

Lu, Y.-J., G. D. Kochert, D. J. Isenhour, and M. J. Adang. 1994. “Molecular Characterization of a Strain-Specific Repeated DNA Sequence in the Fall Armyworm *Spodoptera Frugiperda* (Lepidoptera: Noctuidae).” *Insect Molecular Biology* 3 (2): 123–30. doi:10.1111/j.1365-2583.1994.tb00159.x.

Lu, Y., and M.J. Adang. 1996. “Distinguishing Fall Armyworm (Lepidoptera: Noctuidae) Strains Using a Diagnostic Mitochondrial DNA Marker.” *Florida Entomologist* 79 (1): 48–55. http://www.fcla.edu/FlaEnt/fe79p48.pdf.

Machado, Vilmar, Milena Wunder, Vanessa D Baldissera, Jaime V Oliveira, and Lidia

M Fiu. 2008. “Molecular Characterization of Host Strains of Spodoptera Frugiperda (Lepidoptera : Noctuidae) in Southern Brazil.” *Annals of the Entomological Society of America* 101: 619–26. doi:10.1016/j.bbmt.2008.09.012.

Mauchamp, Bernard, Corinne Royer, Annie Garel, Audrey Jalabert, Martine Da Rocha, Anne Marie Grenier, Valérie Labas, et al. 2006. “Polycalin (Chlorophyllid A Binding Protein): A Novel, Very Large Fluorescent Lipocalin from the Midgut of the Domestic Silkworm Bombyx Mori L.” *Insect Biochemistry and Molecular Biology* 36 (8): 623–33. doi:10.1016/j.ibmb.2006.05.006.

Meagher, R L, R N Nagoshi, C Stuhl, and E R Mitchell. 2004. “Larval Development of Fall Armyworm (Lepidoptera : Noctuidae) on Different Cover Crop Plants.” *Florida Entomologist* 87 (December): 454–60. doi:10.1653/0015-4040(2004)087[0454:LDOFAL]2.0.CO;2.

Meagher Jr., R L, and M Gallo-Meagher. 2003. “Identifying Host Strains of Fall Armyworm (Lepidoptera: Noctuidae) in Florida Using Mitochondrial Markers.” *Florida Entomologist* 86 (4): 450–55. doi:10.1653/0015-4040(2003)086.

Meiklejohn, Colin D, Kristi L Montooth, and David M Rand. 2007. “Positive and Negative Selection on the Mitochondrial Genome.” *Trends Genet* 23 (6): 259–63. doi:10.1016/j.tig.2007.03.008.

Morales, Hernan E, Alexandra Pavlova, Nevil Amos, Richard Major, Andrzej Kilian, Chris Greening, Paul Sunnucks, and Paul Sunnucks. 2016. “Mitochondrial-Nuclear Interactions Maintain Geographic Separation of Deeply Diverged Mitochondrial Lineages in the Face of Nuclear Gene Flow.” *Doi.Org*, 95596. doi:10.1101/095596.

Nagoshi, R. N., and R. L. Meagher. 2016. “Using Intron Sequence Comparisons in the Triose-Phosphate Isomerase Gene to Study the Divergence of the Fall Armyworm Host Strains.” *Insect Molecular Biology* 25 (3): 324–37. doi:10.1111/imb.12223.

Nagoshi, Rod N., and R. Meagher. 2003a. “Fall Armyworm FR Sequences Map to Sex Chromosomes and Their Distribution in the Wild Indicate Limitations in Interstrain Mating.” *Insect Molecular Biology* 12 (5): 453–58. doi:10.1046/j.1365-2583.2003.00429.x.

Nagoshi, Rod N., and Robert L Meagher. 2004. “Seasonal Distribution of Fall Armyworm (Lepidoptera: Noctuidae) Host Strains in Agricultural and Turf Grass Habitats.” *Environmental Entomology*. doi:10.1603/0046-225X-33.4.881.

Nagoshi, Rod N., and Robert L. Meagher. 2003b. “Tandem-Repeat Sequence in Fall Armyworm (Lepidoptera: Noctuidae) Host Strains.” *Annals of the Entomological Society of America* 96 (3): 329–35. doi:10.1603/0013-8746(2003)096[0329:FTSIFA]2.0.CO;2.

Nagoshi, Rod N., Robert L Meagher, Gregg Nuessly, and David G Hall. 2006. “Effects of Fall Armyworm (Lepidoptera: Noctuidae) Interstrain Mating in Wild Populations.” *Environmental Entomology*. doi:10.1603/0046-225X-35.2.561.

Nagoshi, Rod N, Robert L Meagher, John J Adamczyk, S Kristine Braman, Rick L Brandenburg, and Gregg Nuessly. 2006. “New Restriction Fragment Length Polymorphisms in the Cytochrome Oxidase I Gene Facilitate Host Strain Identification of Fall Armyworm (Lepidoptera: Noctuidae) Populations in the Southeastern United States.” *Journal of Economic Entomology* 99 (3): 671–77. doi:10.1603/0022-0493-99.3.671.

Nagoshi, Rodney N. 2010. “) Gene as a Marker of Strain Identity and Interstrain Mating.” *Annals of the Entomological Society of America* 103 (2): 283–92. doi:10.1603/AN09046.

O'Brien, Michael J, and Kevin N Laland. 2012. "Genes, Cuture and Agriculture: An Example of Human Niche Construction." *Current Anthropology* 53 (4): 434–70. doi:10.1086/666585.

Pashley, D P. 1986. "Host-Associated Genetic Differentiation in Fall Armyworm (Spodoptera Frugiperda) (Lepidoptera: Noctuidae): A Sibling Species Complex?" *ANNALS OF THE ENTOMOLOGICAL SOCIETY OF AMERICA 79(6): 898-904*.

Pashley, Dorothy P. 1988. "Current Status of Fall Armyworm Host Strains." *Florida Entomologist* 71 (3): 227–34. doi:10.2307/3495425.

Pashley, Dorothy P., Abner M. Hammond, and Tad N. Hardy. 1992. "Reproductive Isolating Mechanisms in Fall Armyworm Host Strains (Lepidoptera: Noctuidae)." *Annals of the Entomological Society of America* 85 (4): 400–405. doi:10.1093/aesa/85.4.400.

Pashley, Dorothy P., Seth J. Johnson, and Alton N. Sparks. 1985. "Genetic Population Structure of Migratory Moths : The Fall Armyworm (Lepidoptera : Noctuidae)." *Annals of the Entomological Society of America* 78(6) (November): 756–62. doi:10.1093/aesa/78.6.756.

Pashley, Dorothy P, and Julie a Martin. 1987. "Reproductive Incompatibility Between Host Strains of the Fall Armyworm (Lepidoptera: Noctuidae)." *Annals of the Entomological Society of America* 80 (6): 731–33. doi:10.1093/aesa/80.6.731.

Pearce, S. L., D. F. Clarke, P. D. East, S. Elfekih, K. H. J. Gordon, L. S. Jermiin, A. McGaughran, et al. 2017. "Genomic Innovations, Transcriptional Plasticity and Gene Loss Underlying the Evolution and Divergence of Two Highly Polyphagous and Invasive Helicoverpa Pest Species." *BMC Biology* 15 (1): 63. doi:10.1186/s12915-017-0402-6.

Pogue, Michael. 2002. *A World Revision of the Genus Spodoptera Guenée (Lepidoptera: Noctuidae)*. *American Entomological Society*.

Poitout, S, and R Bues. 1974. "Elevage de Chenilles de Vingt-Huit Espèces de Lépidopteres Noctuidae et de Deux Especes d'Arctiidae Sur Milieu Artificiel Simple. Particularités de L'élevage Selon Les Espèces ." *Annales de Zoologie, Ecologie Animale* 6 (3): 431–41.

Prowell, Dorothy Pashley, Margaret McMichael, and Jean-François Silvain. 2004. "Multilocus Genetic Analysis of Host Use, Introgression, and Speciation in Host Strains of Fall Armyworm (Lepidoptera: Noctuidae)." *Annals of the Entomological Society of America* 97 (5): 1034–44. doi:10.1603/0013-8746(2004)097[1034:MGAOHU]2.0.CO;2.

Roy, A., W. B. Walker, H. Vogel, S. Chattington, M. C. Larsson, P. Anderson, D. G. Heckel, and F. Schlyter. 2016. "Diet Dependent Metabolic Responses in Three Generalist Insect Herbivores Spodoptera Spp." *Insect Biochemistry and Molecular Biology* 71: 91–105. doi:10.1016/j.ibmb.2016.02.006.

Schöfl, G., D. G. Heckel, and A. T. Groot. 2009. "Time-Shifted Reproductive Behaviours among Fall Armyworm (Noctuidae: Spodoptera Frugiperda) Host Strains: Evidence for Differing Modes of Inheritance." *Journal of Evolutionary Biology* 22 (7): 1447–59. doi:10.1111/j.1420-9101.2009.01759.x.

Shah, Neethu, Douglas R. Dorer, Etsuko N. Moriyama, and Alan C. Christensen. 2012. "Evolution of a Large, Conserved, and Syntenic Gene Family in Insects." *G3: Genes|Genomes|Genetics* 2 (2): 313–19. doi:10.1534/g3.111.001412.

Silva-Brandão, Karina Lucas, Renato Jun Horikoshi, Daniel Bernardi, Celso Omoto, Antonio Figueira, and Marcelo Mendes Brandão. 2017. "Transcript Expression Plasticity as a Response to Alternative Larval Host Plants in the Speciation

Process of Corn and Rice Strains of Spodoptera Frugiperda." *BMC Genomics* 18 (1). BMC Genomics: 1–15. doi:10.1186/s12864-017-4170-z.

Simon, Jean Christophe, Emmanuelle D'alen??on, Endrick Guy, Emmanuelle Jacquin-Joly, Julie Jaqui??ry, Pierre Nouhaud, Jean Peccoud, Akiko Sugio, and R??jane Streiff. 2015. "Genomics of Adaptation to Host-Plants in Herbivorous Insects." *Briefings in Functional Genomics*. doi:10.1093/bfgp/elv015.

Smith, C. R., C. Morandin, M. Noureddine, and S. Pant. 2018. "Conserved Roles of Osiris Genes in Insect Development, Polymorphism and Protection." *Journal of Evolutionary Biology*, February. doi:10.1111/jeb.13238.

Sparks, Alton N. 1979. "A Review of the Biology of the Fall Armyworm." *The Florida Entomologist*. doi:10.2307/3494083.

Stearns, Stephen C. 2012. "The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data." *Annual Review of Ecology and Systematics* 8 (1977): 145–71. doi:10.1146/annurev.es.08.110177.001045.

Stincone, Anna, Alessandro Prigione, Thorsten Cramer, Mirjam M.C. Wamelink, Kate Campbell, Eric Cheung, Viridiana Olin-Sandoval, et al. 2015. "The Return of Metabolism: Biochemistry and Physiology of the Pentose Phosphate Pathway." *Biological Reviews* 90 (3): 927–63. doi:10.1111/brv.12140.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PLoS ONE* 6 (7). doi:10.1371/journal.pone.0021800.

Thompson, John N. 1988. "Evolutionary Ecology of the Relationship between Oviposition Preference and Performance of Offspring in Phytophagous Insects." *Entomologia Experimentalis et Applicata*. doi:10.1111/j.1570-7458.1988.tb02275.x.

Whiteman, Noah K., Andrew D. Gloss, Timothy B. Sackton, Simon C. Groen, Parris T. Humphrey, Richard T. Lapoint, Ida E. Sønderby, et al. 2012. "Genes Involved in the Evolution of Herbivory by a Leaf-Mining, Drosophilid Fly." *Genome Biology and Evolution* 4 (9): 900–916. doi:10.1093/gbe/evs063.

Whitford, F., S. S. Quisenberry, T. J. Riley, and J. W. Lee. 1988. "Oviposition Preference, Mating Compatibility, and Development of Two Fall Armyworm Strains." *The Florida Entomologist* 71 (3): 234–43. doi:10.2307/3495426.

Xu, Wei, Alexie Papanicolaou, Hui-Jie Zhang, and Alisha Anderson. 2016. "Expansion of a Bitter Taste Receptor Family in a Polyphagous Insect Herbivore." *Scientific Reports* 6 (1): 23666. doi:10.1038/srep23666.

Yassin, Amir, Vincent Debat, Héloïse Bastide, Nelly Gidaszewski, Jean R David, and John E Pool. 2016. "Recurrent Specialization on a Toxic Fruit in an Island Drosophila Population." *Proceedings of the National Academy of Sciences* 113: 4771–76. doi:10.1073/pnas.1522559113.

Yoder, J. B., E. Clancey, S. Des Roches, J. M. Eastman, L. Gentry, W. Godsoe, T. J. Hagey, et al. 2010. "Ecological Opportunity and the Origin of Adaptive Radiations." *Journal of Evolutionary Biology*. doi:10.1111/j.1420-9101.2010.02029.x.

Zohary, Daniel, Maria Hopf, and Ehud Weiss. 2012. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. doi:10.1093/acprof:osobl/9780199549061.001.0001.

Figure Legends

Figure 1: *oviposition choice of sf-C and sf-R.*

Proportion of egg masses laid in the three experimental trials (corn-only, rice-only and choice) by sf-C (**A**) and sf-R (**B**) according to the site of oviposition. There are three oviposition sites available: the net (light gray), the corn plant (yellow) and the rice plant (green). Here, the relative proportions on each laying site represented the mean of proportions obtained about the four replicates. Replicate measurements and standard deviations are shown in **Supplementary Figure 2**.

Figure 2: *Larval traits of sf-C and sf-R according to the diet.*

(**A**,**B**) Weight (`wt`) evolution according to time is measured in duplicate for sf-C (**A**) and sf-R (**B**) on three diet types: corn plant (yellow), rice plant (green) and artificial diet (brown). The weight increases exponentially until the 6th larval instar, then declines sharply during metamorphosis until the final pupal weight, which will remain constant. (**B**,**C**) The survival (`sv`) rate is measured from the 1st larval instar to adult emergence for sf-C (**C**) and sf-R (**D**) according to plant diet. Bars represent the mean of survival rate of the two experimental replicates with the standard error. Different letters above bars indicate significant differences of survival between plant diets for each strain ($P < 0.05$).

Figure 3: *Transcriptional response of sf-C according to the diet.*

A. Principal component analysis on normalized RNA-seq reads of four samples of sf-C when the larvae feed on corn (red), on rice (blue) or on artificial diet (green). **B.** Multidimensional scaling plot (MA-plot) reporting the log2 fold changes between the plant diet (rice *vs* corn) over the mean of normalized counts. Each point represents a

gene either with a non-significant differential expression between trials (gray point) or with a significant differential expression (red point). **C-D.** Blast2GO analysis of biological processes over-represented (red), compared to the reference (blue), for the 100 most differentially expressed genes in rice (**C**) versus corn trial (**D**).

Figure 4: *Transcriptional response of sf-R according to the diet.*

A. Principal component analysis on normalized RNA-seq reads of four samples of sf-R when the larvae feed on corn (red), on rice (blue) or on artificial diet (green). **B.** Multidimensional scaling plot (MA-plot) reporting the log2 fold changes between the plant diet (rice vs corn) over the mean of normalized counts. Each point represents a gene either with a non significant differential expression between trials (gray point) or with a significant differential of expression (red point). **C-D.** Blast2GO analysis of biological processes over-represented (red), compared to the reference (blue), for the 100 most differentially expressed genes in rice (**C**) versus corn trials (**D**).

Figure 5*: Sf-C and Sf-R have different responses to plants.*

A. Venn Diagrams illustrating shared and unshared sets of genes whose expression was correlated to strains and plant diets. **B.** Correlogram showing the Euclidean distance of normalized log transformed read counts between all RT samples. Samples are ordered according to hierarchical clustering. All sf-C samples cluster together, and all sf-R samples as well. There is no grouping according to diet. **C-F.** Heatmaps of 50 most differentially expressed genes split. Each raw represents z-score normalized expression for one gene across all RT samples. In each heatmap, genes are ordered from top to bottom, from the most significant to the 50th most significant and the blue-white-red color scale indicates lower, no and higher variation

of gene expression for each gene across RT samples. The heatmaps correspond to the genes overexpressed either on rice plant for sf-C (**C**) and for sf-R (**E**) or on corn plant for sf-C (**D**) and for sf-R (**F**).

Figure 6: *Transcriptional response of sf-R versus sf-C regardless of the diet.*

A. Principal component analysis on normalized RNA-seq reads for all RT samples of sf-R and sf-C when the larvae feed on corn (red), on rice (blue) or on artificial diet (green). **B.** Multidimensional scaling plot (MA-plot) reporting the log2 fold changes between the strains (sf-R vs sf-C) over the mean of normalized counts. Each dot represents a gene either with a non significant differential expression between trials (gray dots) or with a significant differential of expression (red dots).

Figure 7: *RNA-Seq of individual larvae from the fields*

A. Genotypes of the individual L4 larvae from natural populations used for RNAseq studies. CoI, tpi and FR1 repeat genotyping has been done by PCR-RFLP (**Supplementary Figures 8-10**). Color code is dark green for presumptive C-strain genotype according to the literature while purple is for presumptive R-strain genotypes. Sex has been determined *post-facto* by examining the alignments of reads on the Z-associated *tpi* locus. If all SNP positions within the scaffold are homozygous, we assumed the individual was female. Heterozygocity indicates a male. **B**. Heatmap of normalized expression levels for genes of the *Osiris* (*osi*) family across all RNAseq experiments: FL15 (col.1-9); RT1 (col. 10-13); RT2 (col. 14-21) and RoyRT (col. 22-25) with dark red as the highest expression and white as no expression. **C**. Heatmap of normalized expression levels for genes of the Serine Proteases (SP) family across all RNAseq experiments: FL15 (col.1-9); RT1 (col. 10-

13); RT2 (col. 14-21) and RoyRT (col. 22-25) with dark red as the highest expression and white as no expression. **D**. Multidimensional scaling plot (MA-plot) reporting the log2 fold changes between the strains (sf-R vs sf-C) over the mean of normalized counts when combining FL15 and MORT2 experiments. Each dot represents a gene either with a non significant differential expression between conditions (gray dots) or with a significant differential of expression (red dots). 76 genes are overexpressed in sf-R and 73 in sf-C. **E**. Heatmap of expression variations (expressed as z-scores) of the 76 sf-R expressed genes across all RNAseq experiments (**TableS2H**). For each gene, red indicates a higher expression and blue a lesser expression across the experimental dataset. Genes have been hierarchically clustered as indicated by the dendrogram on the left by similarity of expression variation. The red asterisk identifies the COI-*numt* expression. **F**. View of the mitochondrial genome corresponding to the COI-numt sequence and alignment coverage of reads corresponding to sf-R (red) or sf-C (green) samples of the MORT2 experiment. We can observe a trough of expression in this region associated with sf-C strain.

Fig. S1 - A. Fertility represented by the number of egg-masses divided by the number of females present in mating cages. Values represent the mean of fertility with the standard error for sf-C (green) and sf-R (red) according to different experimental trials: choice (in presence of corn and rice plants), no-choice (either in presence of corn only or in presence of rice only). The letters above the bars means indicated the significant differences in the mean fertility ($P < 0.05$). For sf-C (**B**) and sf-R (**C**), we counted the percentage of eggs (y-axis) that gave rise to a live larva for sf-C and sf-R in each trial. Error bars represent the variations between egg-masses. No statistical differences were observed between trials.

Fig. S2 - Proportion of egg masses laid on corn (yellow), rice (green) or net (light gray) by sf-C (**A**, **C** and **E**) and sf-R (**B**, **D** and **F**) in three experimental conditions: cages with rice only (**A** and **B**), cages with corn only (**C** and **D**), and with both plants corn and rice (**E** and **F**). Bars represent the mean of proportion of four experimental replicates with the standard error. Different letters above bars indicate significant differences between the egg masses proportions of different laying sites ($P < 0.05$).

Fig. S3 - Larval traits in sf-C and sf-R according to the diet: corn plant (yellow), rice plant (green). Bars represent the mean of female weight for sf-C (**A**) and sf-R (**B**), the mean of male weight for sf-C (**C**) and sf-R (**D**) and the developmental time until adult emergence for sf-C (**E**) and sf-R (**F**). The variation between replicates is represented by the standard error (except for the developmental time which are exactly the same for both strain on corn plant) and the different letters above bars indicate significant differences between plant diets for each strains ($P < 0.05$).

Fig. S4 : Example of manual gene annotation

A. In the *S. frugiperda* genome (Gouin et al., 2017) the gene GSSPFG00032711001 is differentially expressed between sf-C and sf-R, however its function is unknown. In this WebApollo browser screenshot, the predicted gene of the official gene set (OGS2.0) is shown in green. The alignment of RNAseq reads in this region, shown in gray, reveals an intron darker gray. We used this support to correct the structure of this gene in the yellow track. **B**. The corrected sequence is now used to perform blastp annotations and reveal that this gene has in fact been identified as *polycalin* in other Lepidoptera (Mauchamp et al. 2006).

Fig. S5 - 50 most expressed genes in sf-R

This heatmap displays the relative gene expression of the top 50 most differentially expressed gene in sf-R across the MORT2 experimental datasets, where red is overexpressed and blue underexpressed (z-scores). The columns on the right indicate the gene identification name and its manual reannotation. Genes are ordered from most overexpressed (top) to less.

Fig. S6 - 50 most expressed genes in sf-C

This heatmap displays the relative gene expression of the top 50 most differentially expressed gene in sf-C across the MORT2 experimental datasets, where red is overexpressed and blue underexpressed (z-scores). The columns on the right indicate the gene identification name and its manual reannotation. Genes are ordered from most overexpressed (top) to less.

Fig. S7 - qPCR validation of RT RNAseq experiments

This figures shows two examples of strain associated gene expressions. The first one (top left: slack-LINE1) is a series of 3 LINE-type transposable elements expressed in sf-R. The IGV browser screenshot shows the RNA-Seq coverage across this region. On the right are the qPCR measurements ($\Delta\Delta$Ct values on the y-axis) of expression associated to slack-LINE1 in three independent individual larvae of each strain, confirming its overexpression in sf-R.

At the bottom, another example is shown for the Fatty Acid Binding protein 10 (FABP-10), a member of a cluster of similar genes involved in fatty acid transport in the midgut, whose expression is associated to sf-R.

Fig. S8 - Genotyping of individual larvae using the COI diagnostic gene

A diagnostic locus of 550 bp in the mitochondrial gene Cytochrome Oxidase I (COI) (Meagher Jr. and Gallo-Meagher 2003) has been amplified by PCR. **A.** Digestion by the MspI restriction enzyme is possible only in the sf-C strain and liberates one 500 bp fragment and a 50bp fragment. This PCR_RFLP is tested on individual L4 larvae from our laboratory colonies. All sf-C are digested, none of the sf-R. **B.** Test on 32 L4 individual larvae from the Citra sweet corn field. **C.** Test on 18 larvae from the Tifton corn field and 6 larvae from the Jacksonville pasture grass field. **D.** Proportion of diagnosed sf-C and sf-R individuals in each field.

Fig. S9 - Genotyping of individual larvae using the *tpi* gene SNP

A diagnostic locus of 800 bp in the Z-linked gene *Triose Phosphate Isomerase* (*Tpi*) (Nagoshi 2010) has been amplified by PCR. The PCR fragment encompasses introns 2 and 3 of the Tpi gene. **A.** Digestion by the AvaII restriction enzyme is possible only in the sf-R strain and liberates one 500 bp fragment and one 300bp fragment. This PCR-RFLP method is tested on individual L4 larvae from our laboratory colonies. All sf-R are digested, none of the sf-C. **B.** Test of the marker in select individuals from each field. The names in red indicate the putative sf-R larvae according to COI genotype. An R is noted when individuals show a proper restriction. Only one individual from Tifton (B25) is tested as sf-R with this marker. Individuals A11 and B20 show two amplified bands, indicating that they may be heterozygous for the intron length. It has been shown that intron length polymorphism exists at this gene (Nagoshi and Meagher 2016). All tested larvae from Jacksonville show the expected sf-R digestion pattern.

Fig. S10 - Genotyping of individual larvae using the FR1 repeat

The FR1 repeat is a sex-linked repeat element associated with the sf-R strain. It is present in sf-C but with less copies (Nagoshi and Meagher 2003b; Nagoshi and Meagher 2003a). **A.** In the laboratory population, some sf-R individuals show a strong multiband amplification, indicative of the presence of this repeat. These copies are supposedly on the W chromosome and as such can only be detected in males. In natural populations, only two individuals from the Tifton field show this amplification. The B25 individual, that was genotyped as sf-R with COI and *Tpi* markers, doesn't show the FR1 amplification, probably because it is a male. **B.** Low copy numbers are detected in the Jacksonville individuals, except for the B5 individual, which might be the only female.

Fig. S11 - qPCR measurement of DE genes in natural populations

Examples shown here are qPCR expression measurements ($\Delta\Delta$Ct values on the y-axis) examples for two sf-R strain associated gene expressions: slack-LINE1 and ngf1a, a nervous system associated transcription factor. We tested the expression of these 2 genes in individual larvae from our laboratory colonies (Lab) and from the Florida collections of sf-C or sf-R genotypes. The overexpression is observed only in laboratory sf-R larvae.

Fig. S12 - **A**. Principal component analysis (PCA) of normalized RNA-seq reads of sf-R and sf-C individual larvae sampled in Tifton (blue) or Jacksonville fields (red). The samples cluster by collection groups. **B**. Correlogram of the FL15 RNAseq experiments showing no clear overall correlation per genotype. **C**. PCA of all RNA-

seq samples from the laboratory and field conditions. The laboratory sf-R experiments cluster with field individuals while laboratory sf-C samples cluster away.

Fig. S13 - Heatmap of 50 most DE genes overexpressed on corn plant for sf-R (same as Fig. S5). Each raw represents z-score normalized expression for one gene across all RT and field samples. Genes are ordered from top to bottom, from the most significant to the 50th most significant and the blue-white-red color scale indicates lower, no and higher variation of gene expression for each gene. These genes are clearly overexpressed in laboratory sf-R and underexpressed in laboratory sf-C. But no clear pattern is observable in other RNAseq experiments or from field collections.

Fig. S14 - Heatmap of 50 most DE genes overexpressed on corn plant for sf-C (same as **Fig. S6**). Each raw represents z-score normalized expression for one gene across all RT and field samples. Genes are ordered from top to bottom, from the most significant to the 50th most significant and the blue-white-red color scale indicates lower, no and higher variation of gene expression for each gene. These genes are clearly overexpressed in laboratory sf-C but are mostly underexpressed in all other experiments.

Fig. S15 - Annotation of the *osiris* gene family in *Spodoptera frugiperda*.

This diagram shows the relative positions, size and orientation of *osiris* genes in *Drosophila melanogaster* (top) (Shah et al. 2012) and in *S. frugiperda* genomes after reciprocal blast analysis. In *S. frugiperda*, one big cluster contains most *osi* genes in the same order as in *Drosophila*, with an expansion of the *osi9* paralogs. Other *S.*

frugiperda osi members have been found in the genome but their localization within or outside the cluster could not be determined.

Fig. S16 - Sf-C associated gene expression across all RNAseq experiments.

73 genes (**Table S2G**) have a sf-C specific expression in laboratory experiments as well as in field collection samples. This heatmap shows the relative expression of each of these genes across all RNAseq samples analyzed (z-scores).

Fig. S17 - Annotation of COI-*numt* in the *S. frugiperda* genome

A. Webapollo screenshot showing the GSSPFG00006578001-RA predicted gene on scaffold-722 and RNAseq coverage underneath. In the yellow track, the part that has a sequence homology with mitochondrial COI gene is shown in magenta. **B**. log2 fold changes of expression of the COI-numt in all RNAseq samples showing their sf-R associated expression.

Fig. S18 - Staging of L4 larvae

A. Actual size chart that was used after calibration in laboratory conditions to stage *S. frugiperda* larvae. The width of the lines should correspond to the width of cephalic capsule. **B**. In field collections, larvae were placed on the chart printouts so that their body follows a line. To be considered an L4 larva, the width of the head should be the same size or slightly bigger than the width of the line.

Table S1 - Sequencing and alignment statistics of RNAseq experiments

This table is presenting the number of reads processed per sample and their different alignment statistics with bowtie2 (Langmead and Salzberg 2012).

Table S2 - Gene lists of the 50 most differentially expressed genes in each trial sorted by p-value and manually verified and annotated. This Excel file contains several tabs with gene lists whose expression is associated to the following conditions. **A**. sf-C on rice **B**. sf-C on corn **C**. sf-R on rice **D**. sf-R on corn **E**. sf-R **F**. sf-C **G**. sf-C laboratory and field **H**. sf-R laboratory and field. In G and F only the most significantly expressed have been manually annotated. Each tab contains the following columns: `OGS2.2` refers to the gene identifier in the OGS2.2 annotation version of the sf-C reference genome (Gouin et al. 2017). The coordinates of the gene on the reference genome is given in the `scaffold:start:end` columns. `baseMean:log2FoldChange:padj` are extracted from the result files of the DESeq2 analyses. `Annotation` is a verified manual annotation for each gene based on homology and InterPro analysis. Names in bold in this last column indicate previously published gene manual annotations (Gouin et al. 2017).

Table S3 - Comparison of RNAseq data and qPCR

This table is a list of 30 genes that are found overexpressed in sf-R compared to sf-C in the RT experiment. Last two columns on the right indicate the log2 Fold Change observed in RNAseq experiments and the $\Delta\Delta$Ct values obtained by qPCR. Except for peroxidase, all genes tested show a confirmed overexpression of these genes in sf-R.

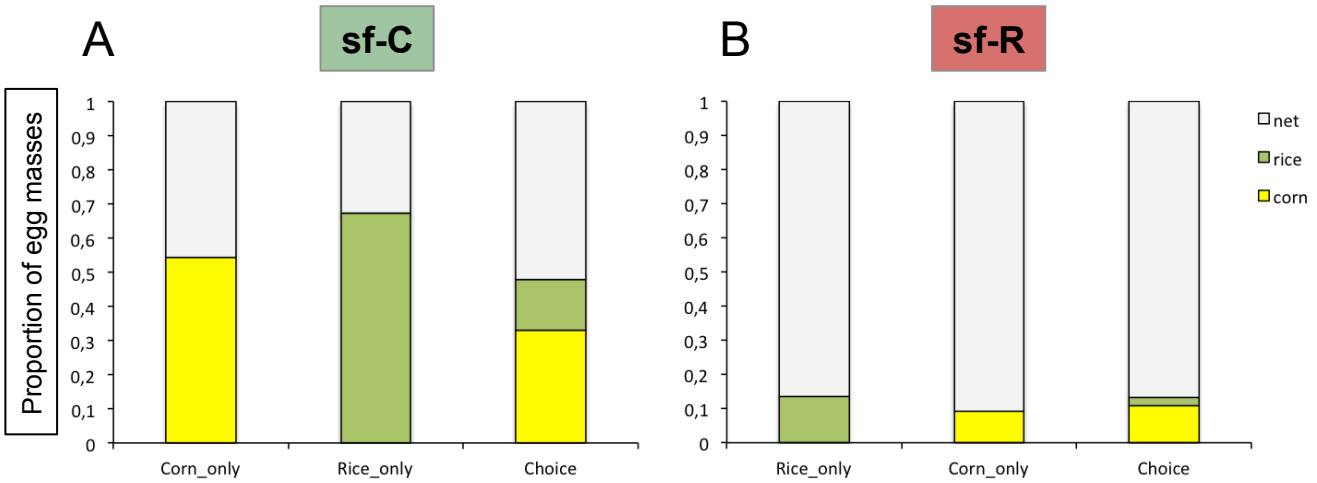


FIGURE 1

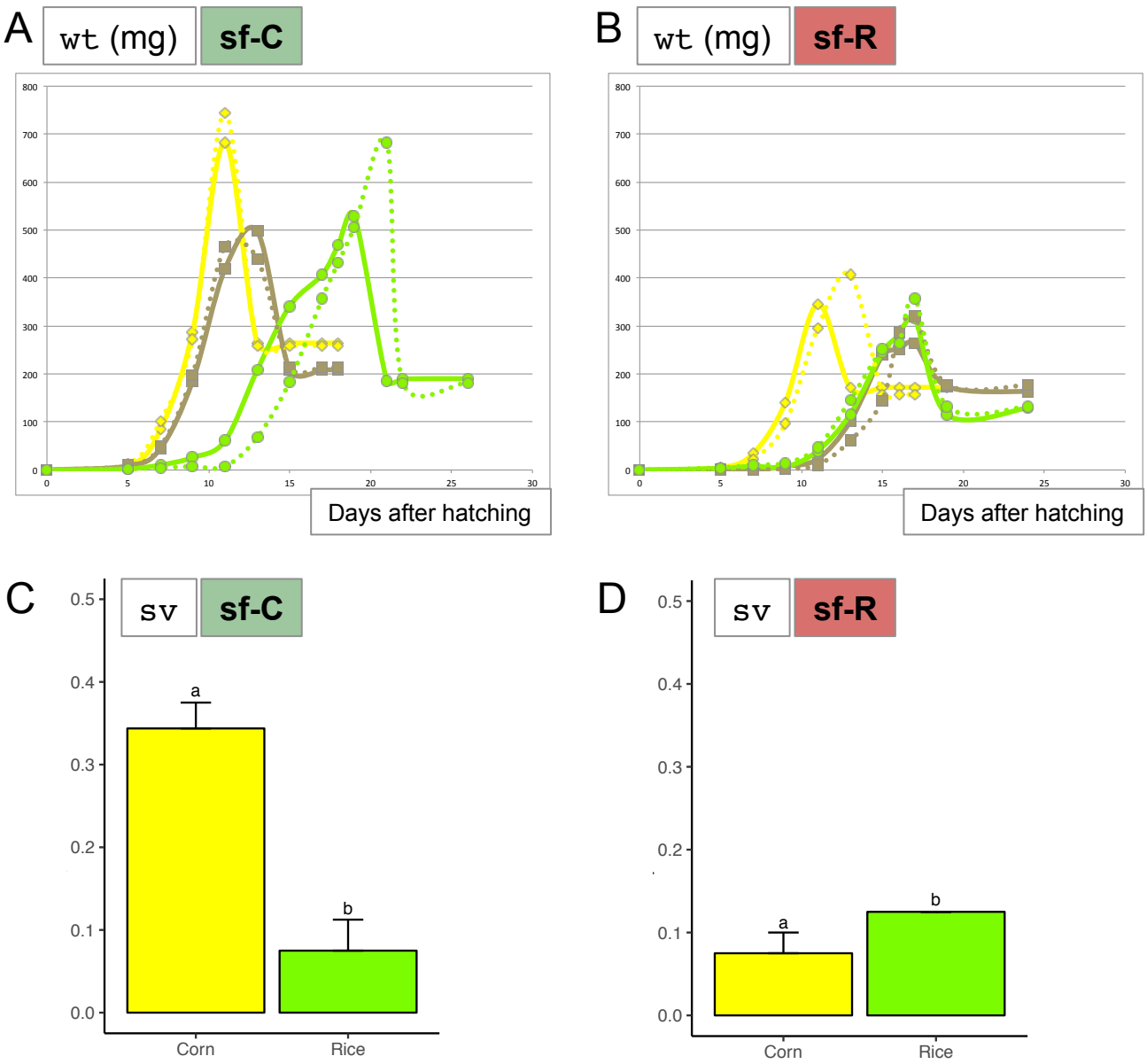

A
wt (mg)
sf-C
B
wt (mg)
sf-R
Days after hatching
Days after hatching
C
sv
sf-C
D
sv
sf-R
a
b
a
b
Corn
Rice
Corn
Rice

FIGURE 2

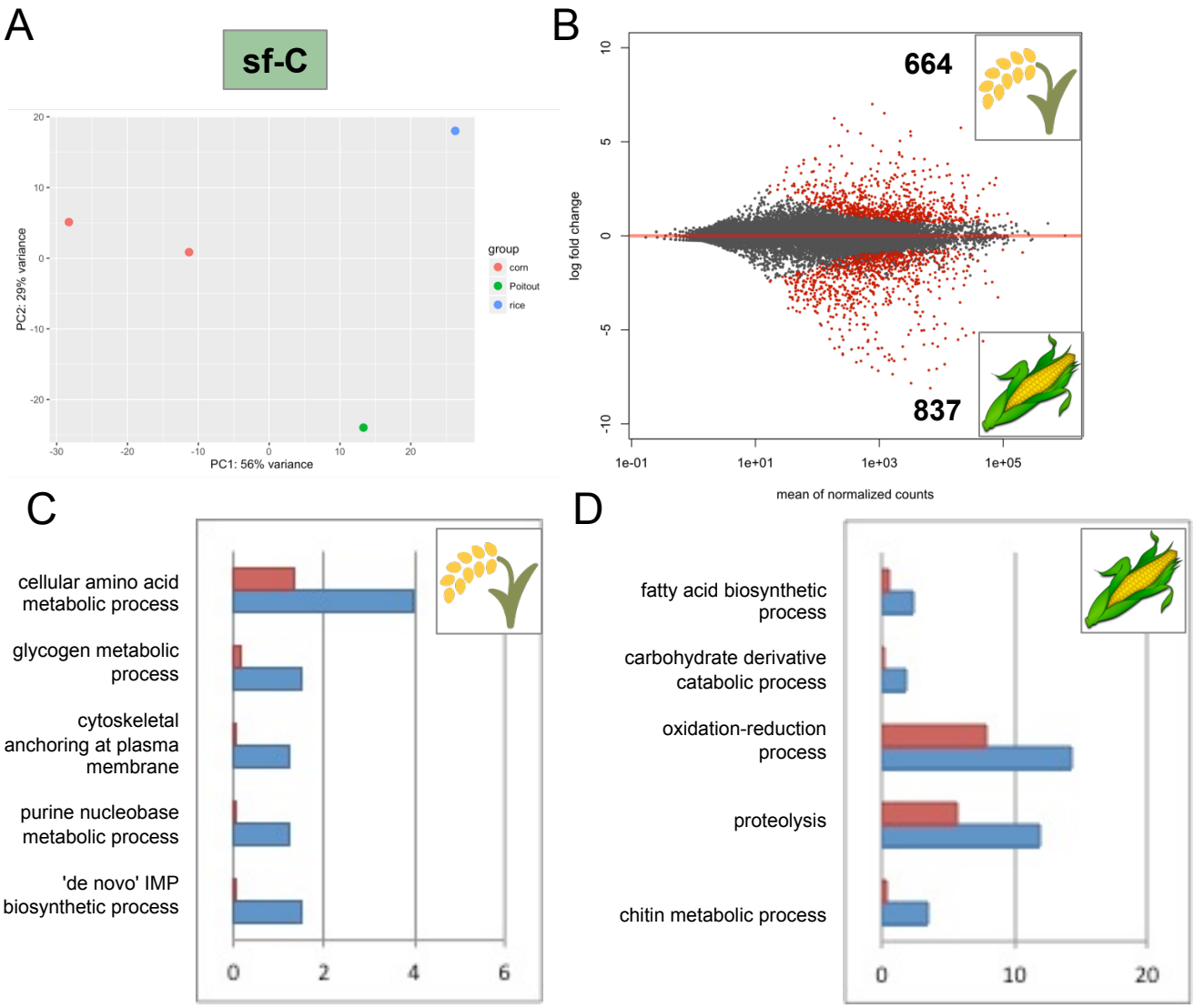

A
sf-C
PC2: 29% variance
PC1: 56% variance
group
corn
Poitout
rice
B
664
837
log fold change
mean of normalized counts
C
cellular amino acid metabolic process
glycogen metabolic process
cytoskeletal anchoring at plasma membrane
purine nucleobase metabolic process
'de novo' IMP biosynthetic process
D
fatty acid biosynthetic process
carbohydrate derivative catabolic process
oxidation-reduction process
proteolysis
chitin metabolic process

FIGURE 3

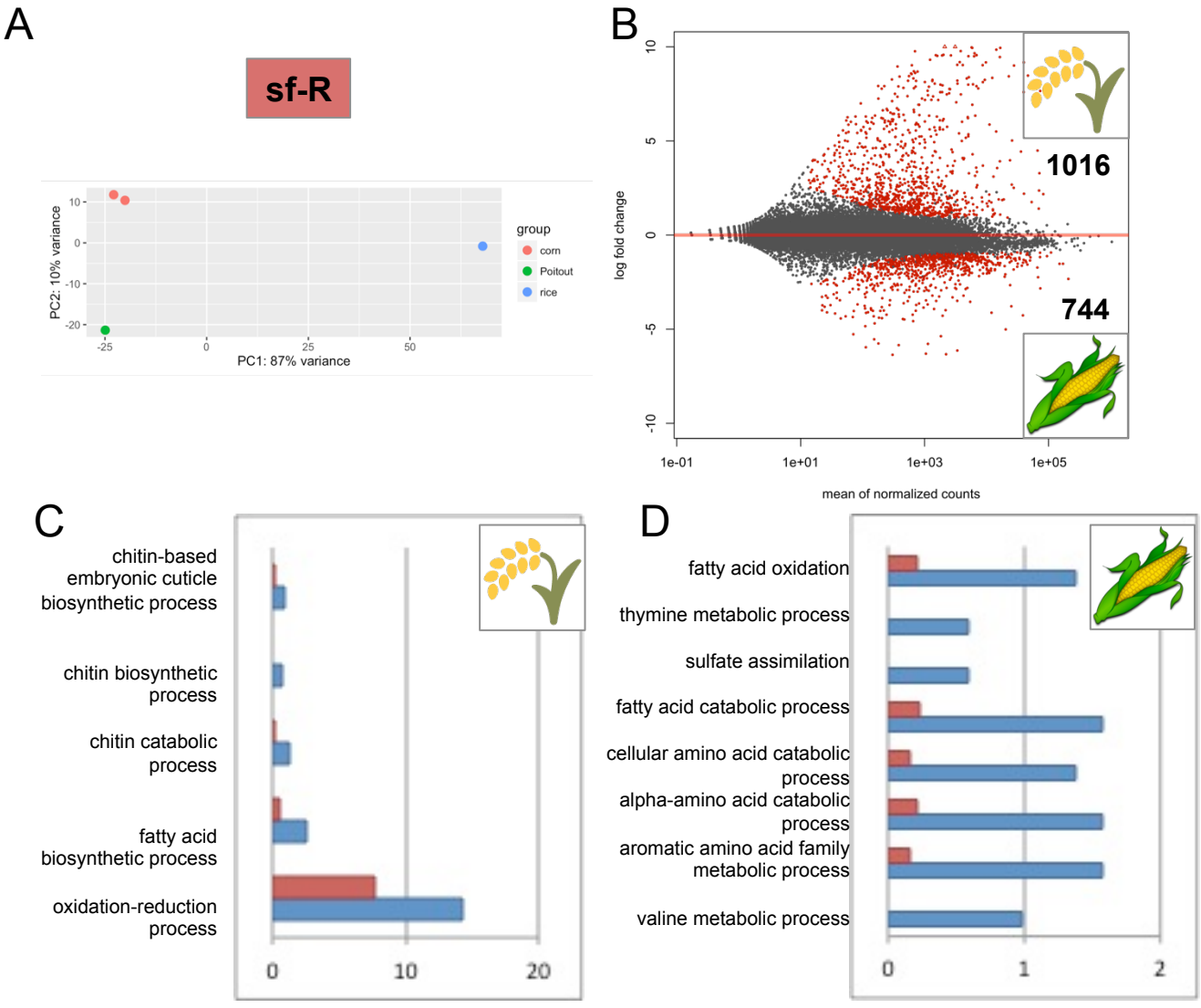
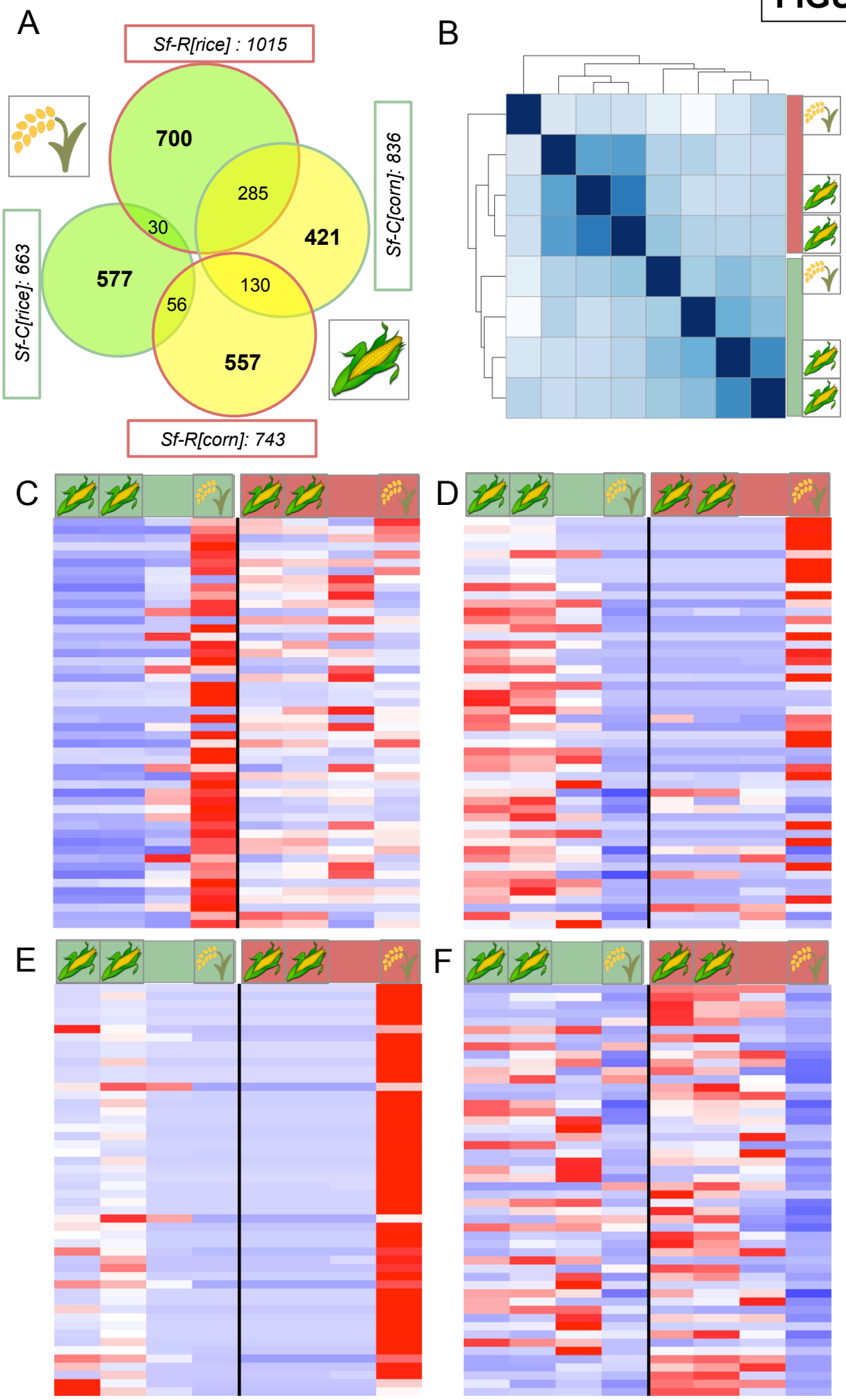
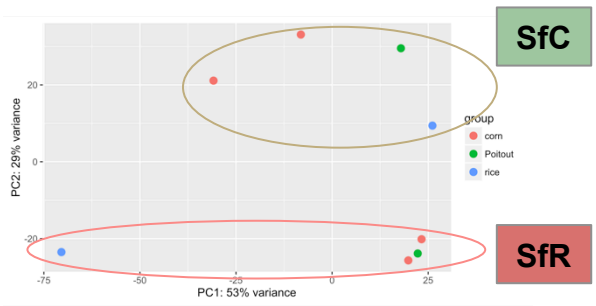

A
sf-R
PC2: 10% variance
PC1: 87% variance
group
corn
Poitout
rice
B
log fold change
mean of normalized counts
1016
744
C
chitin-based embryonic cuticle biosynthetic process
chitin biosynthetic process
chitin catabolic process
fatty acid biosynthetic process
oxidation-reduction process
D
fatty acid oxidation
thymine metabolic process
sulfate assimilation
fatty acid catabolic process
cellular amino acid catabolic process
alpha-amino acid catabolic process
aromatic amino acid family metabolic process
valine metabolic process

FIGURE 4

FIGURE 5


A
Sf-R[rice] : 1015
Sf-C[corn]: 836
Sf-C[rice]: 663
Sf-R[corn]: 743
700
285
30
421
577
130
56
557
B
C
D
E
F

A


SfC
SfR
group
corn
Poitout
rice
PC2: 29% variance
PC1: 53% variance
20
0
-20
-75
-50
-25
0
25

B

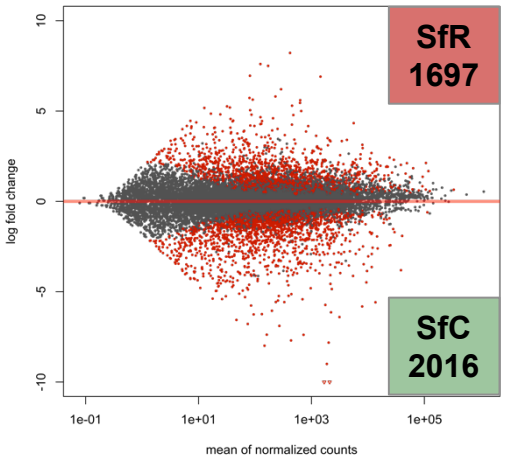

SfR
1697
SfC
2016
log fold change
mean of normalized counts
10
5
0
-5
-10
1e-01
1e+01
1e+03
1e+05

FIGURE 6

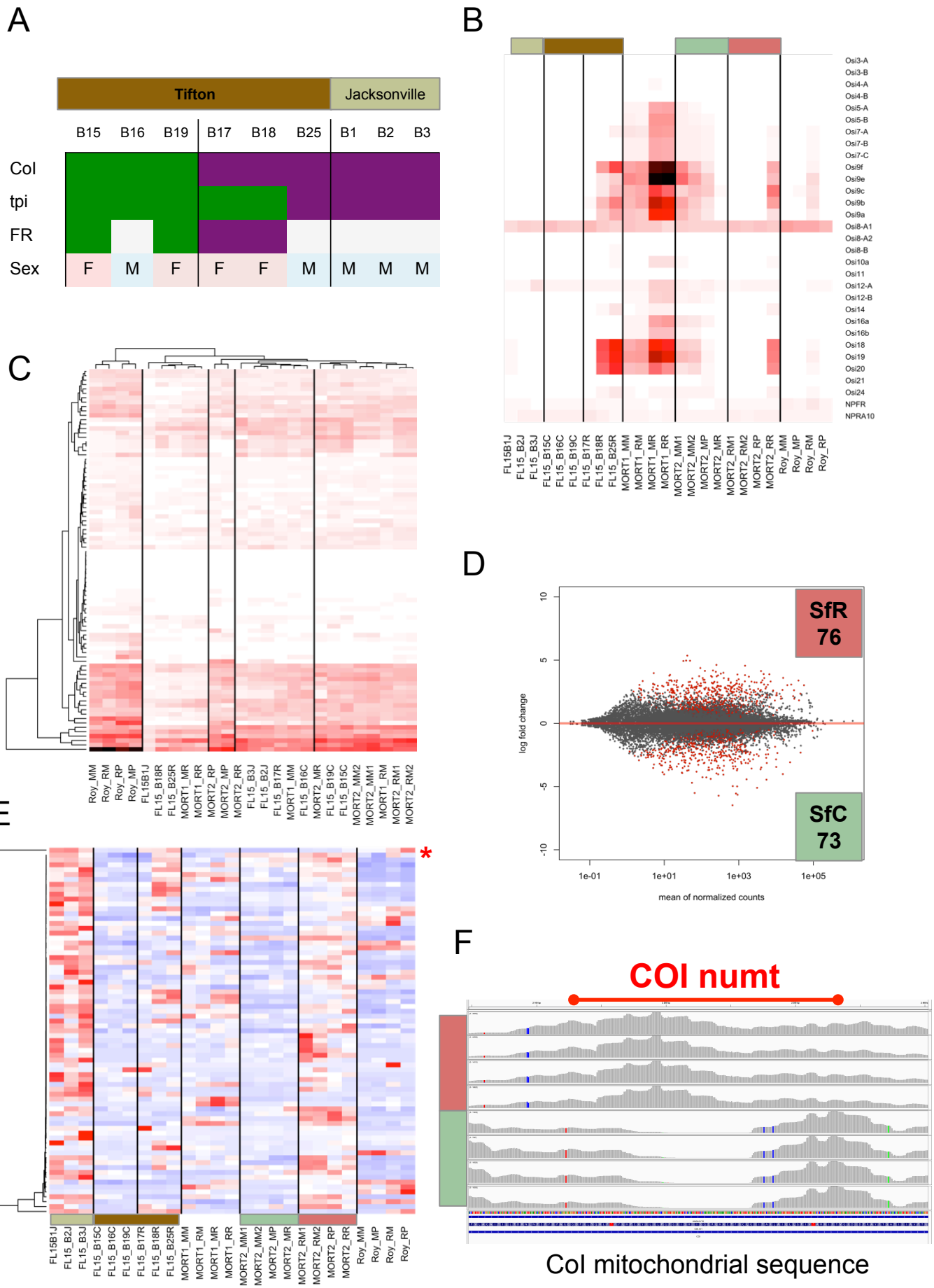

A
Tifton
Jacksonville
B15 B16 B19 B17 B18 B25 B1 B2 B3
Col
tpi
FR
Sex
F M F F F M M M M
B
Osi3-A
Osi3-B
Osi4-A
Osi4-B
Osi5-A
Osi5-B
Osi7-A
Osi7-B
Osi7-C
Osi9f
Osi9e
Osi9c
Osi9b
Osi9a
Osi8-A1
Osi8-A2
Osi8-B
Osi10a
Osi11
Osi12-A
Osi12-B
Osi14
Osi16a
Osi16b
Osi18
Osi19
Osi20
Osi21
Osi24
NPFR
NPRA10
FL15B1J
FL15_B2J
FL15_B3J
FL15_B15C
FL15_B16C
FL15_B19C
FL15_B17R
FL15_B18R
FL15_B25R
MORT1_MM
MORT1_RM
MORT1_MR
MORT1_RR
MORT2_MM1
MORT2_MM2
MORT2_MP
MORT2_MR
MORT2_RM1
MORT2_RM2
MORT2_RP
MORT2_RR
Roy_MM
Roy_MP
Roy_RM
Roy_RP
C
Roy_MM
Roy_RM
Roy_RP
Roy_MP
FL15B1J
FL15_B18R
FL15_B25R
MORT1_MR
MORT1_RR
MORT2_RP
MORT2_MP
MORT2_RR
FL15_B3J
FL15_B2J
FL15_B17R
MORT1_MM
FL15_B16C
MORT2_MR
FL15_B19C
FL15_B15C
MORT2_MM2
MORT2_MM1
MORT1_RM
MORT2_RM1
MORT2_RM2
D
SfR
76
SfC
73
log fold change
10
5
0
-5
-10
1e-01
1e+01
1e+03
1e+05
mean of normalized counts
E
*
FL15B1J
FL15_B2J
FL15_B3J
FL15_B15C
FL15_B16C
FL15_B19C
FL15_B17R
FL15_B18R
FL15_B25R
MORT1_MM
MORT1_RM
MORT1_MR
MORT1_RR
MORT2_MM1
MORT2_MM2
MORT2_MP
MORT2_MR
MORT2_RM1
MORT2_RM2
MORT2_RP
MORT2_RR
Roy_MM
Roy_MP
Roy_RM
Roy_RP
F
COI numt
CoI mitochondrial sequence

FIGURE 7