

Transcriptional differences between the two host strains of *Spodoptera* *frugiperda* (Lepidoptera: Noctuidae)

Marion Orsucci^{1,2}, Yves Moné¹, Philippe Audiot², Sylvie Gimenez¹, Sandra Nhim¹, Rima Naït-Saïdi¹, Marie Frayssinet¹, Guillaume Dumont¹, Jean-Paul Boudon³, Marin Vabre³, Stéphanie Rialle⁴, Rachid Koual⁴, Gael J. Kergoat², Rodney N. Nagoshi⁵, Robert L. Meagher⁵, Emmanuelle d'Alençon¹ & Nicolas Nègre^{1§}

¹DGIMI, Univ Montpellier, INRA, Montpellier, France

²CBGP, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France

³UE Diascope, INRA, 34130 Manguio, France

⁴MGX, Univ Montpellier, CNRS, INSERM, Montpellier, France

⁵USDA-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, India and East-Asia, 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. To test this hypothesis, we first 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 identify constitutive transcriptional differences between strains, regardless of diet, in laboratory as well as in natural populations. We found that mitochondrial transcription is the main difference between the two strains. Since mitochondrial genotypes are also the main genetic variation between the strains, we propose that the 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 plant-interacting organisms were prone to exploit these new niches. For example, some phytophagous insects were able to adapt to cultivated plants and, with the intensification of production based on monoculture activities, these insects eventually became 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). It has since been reported in India, South-East Asia and China (see the most recent report maps at <https://www.cabi.org/isc/fallarmyworm>), threatening to become a world-wide menace.

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).

In order to understand if plant adaptation is indeed at the origin of the differences between the strains, 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. In parallel, we performed RNA-Seq experiments to search for genes constitutively differently transcribed between strains, in laboratory as well as in natural populations, that could indicate which selective pressure led to strains divergence. Surprisingly, we identified a major difference in the transcription of the mitochondrial genome. Since mitochondrial genotypes are also the main genetic variation between the strains, we propose that the mitochondrial genome was the primary target of selection between the two strains.

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, $\chi^2 = 0.17$, $df = 1$, $P = 0.68$) or laying sites (LRT, $\chi^2 = 6.39$, $df = 6$, $P = 0.38$), 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**). 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 : $\chi^2 = -68.35$, $df = 1$, $P < 0.001$; LRT for rice trial : $\chi^2 = -90.10$, $df = 1$, $P < 0.001$.; LRT for choice trial : $\chi^2 = -39.53$, $df = 2$, $P < 0.001$). For sf-C, our model shows no difference in the proportion of egg masses between the net and corn plants in corn trial (LRT, $\chi^2 = -1.30$, $df = 1$,

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

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. This lack of preference for their preferred host plant is surprising because *S. frugiperda* is a species subdivided into two strains according to the host plant on which the individuals were found preferentially (*i.e.* sf-R on *Oriza sativa*, *Bermuda grass*, *Cynodon* spp. and *Medicago sativa* whereas sf-C consumes mainly *Zea mays*, *Sorghum* spp. and *Gossypium hirsutum*; Pashley 1986). The question of qualifying them as two distinct species has already been raised (Dumas et al. 2015). However, although two variants are defined, *S. frugiperda* is mainly qualified as a polyphagous species found on about 100 different host plants belonging to 27 different families (Pogue 2002). Despite these host plant preferences observed in natural populations, both strains can be sampled on the same plants (Juárez et al. 2012). About 19% of sf-R individuals are present on maize and 5% of sf-C individuals are present on

various herbaceous plants (Prowell et al. 2004). This lack of striking female preferences could be accentuated by working on laboratory strains, forced for several generations to lay on filter paper.

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 (w_t), the developmental stage to measure the time intervals (d_t) and the survival (s_v).

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 (d_t) was explained only by the host plant (LRT, $\chi^2 = -37.41$, $df = 1$, $P < 0.001$) and there was no strain effect (LRT, $\chi^2 = -0.93$, $df = 1$, $P = 0.335$; **Fig. S2**). 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. S2A**). 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. S2B**). Finally, both strains exhibited a similar pattern for d_t from 1st larval instar to adult emergence, with both strains having a longer d_t

feeding on rice than on corn (LRT, $F = 28.88$, $df = 1$, $P < 0.0001$; **Fig. S2**). 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 (w_t) at the pupal stage was explained by host plant (LRT: $\chi^2 = 555.25$, $df = 1$, $P < 0.001$), moth strain and sex, with a significant interaction between the last two variables (LRT: $\chi^2 = -6.61$, $df = 1$, $P = 0.012$). Indeed, we observed, except for sf-C on corn, that males were heavier than females (**Fig. 2A-D**). Both strains had heavier pupae from feeding on corn than feeding on rice (for sf-C: LRT, $\chi^2 = 67.107$, $df = 2$, $P < 0.001$; for sf-R: LRT, $\chi^2 = 27.18$, $df = 2$, $P < 0.0001$, **Fig. 2A-D**). Pupal weights were higher on corn condition (around 260 mg) than on rice (around 185 mg; **Fig. 2A-D**). 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.

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, $\chi^2 = -24.22$, $df = 1$, $P < 0.0001$; **Fig. 2E-F**). The survival of sf-C was significantly greater on corn (about 34%) than on rice (about 7.5%; **Fig. 2E**). However, although sf-R tended to have higher survival on rice (LRT, $\chi^2 = 2.53$, $P = 0.11$), sv was not significantly different between the two host plants (7.5% on "corn" vs 12.5% on "rice"; **Fig. 2F**). We noted that the survival rates on plant experimental set-ups were relatively low. These absolute numbers cannot be related to controlled conditions where artificial rearing is designed to provide as much survival of the population as possible (**Figure S3**). Similarly, it can not be compared to survival rates in the wild, for which we have no estimate. Host-plant, but also variable environmental parameters and interactions with competitors, predators, parasites

and pathogens can affect the survival and are an essential component of the host-plant as an ecological niche. Here, we can only conclude on the relative survival rates between similar experimental conditions, which we think reveals intrinsic adaptation to the host-plant.

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. Once again, as noted in the plant preference, the absence of plant cues during laboratory breeding over several generations could have allowed a relaxed selection of host plant characteristics. Moreover, the artificial diet is based on corn flour and therefore Sf-R has not been confronted with rice compounds for many years. Sf-R has therefore been able to adapt to certain compounds of corn explaining that differences between these two plants are not detected.

Gene expression in RT experiment

When confronted with different host plants, polyphagous insects will respond by expressing different sets of genes, some of them can be associated to a better adaptation to the host plant. Such adaptation genes in insect are known to be involved in chemosensory, digestion, detoxification and immunity processes among others (Simon et al. 2015; Celorio-Mancera et al. 2016). In order to understand if the two *S. frugiperda* strains express different adaptation genes to host plant diet, we performed RNA-Seq experiments from the larvae of the RT experiments. RNA was extracted from 4th instar larvae from the same RT experimental setup as the one on

which LHT were measured. We could perform for each strain two replicates on the corn diet, 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**).

Constitutive transcriptional differences between sf-C and sf-R

PCA analysis of the RNA-Seq data shows that the samples are grouped by strain (29% of explained variance on PC2; **Fig. 3A**), 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. 3A**). We used DESeq2 (Love, Anders, and Huber 2014) to identify constitutive 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. 3B**). We verified by q-PCR on independent samples raised on artificial diet that this strain-specific difference of expression is stable. We selected and annotated (**Fig. S4**) 50 genes overexpressed in sf-R compared to sf-C in our RNA-Seq experiments (**Fig. S5-S6**), all except one (peroxidase), were systematically overexpressed in sf-R when measured by qPCR (**Table S2-S3 & Fig. S7**).

The GO enrichment analysis did not detect any significant enrichment of either Biological Process or Molecular Function 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 (**Fig. S5**) 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) (**Fig. S6**). 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 messenger RNAs of mitochondrial origin, whose reads also align on the *numt* region. In practice, *numts* show differences of expression at the level of mitochondria.

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 (COI) 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, in laboratory conditions, we can control the genetic background of insects, the environmental conditions as well as the plant types and supply, while natural populations experience many more variables. Their genetic background might be different from one another, they may be infected or parasitized, they may be individually 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 produce a dataset (named FL15) of RNA-Seq experiments with 3 sf-C individuals from Tifton, 3 sf-R individuals from Tifton and 3 sf-R individuals from Jacksonville (**Fig. 4A**). We recovered from 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 dataset only, replicates of the same "trial + strain" individuals group well together with the FL15_B1J individual 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 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**).

Strain specific expression in laboratory and in field collections

We took advantage of a large dataset to ask again a simple question: what are the genes whose expression is constitutive of 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. 7B**). 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. 4C** and **Fig. S15**). 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 4C**). 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 (**Fig. S16**). 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. 4D**). The implication of this result on the metabolism of the larvae remains to be established, but nevertheless, it may explain 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 differentiation of *S. frugiperda* in two strains - sf-C and sf-R - is a result of their adaptation to different host plant diet. 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 performed RNA-Seq measurements of gene expression variations of L4 larvae during controlled RT experiments in the laboratory and in natural populations. The integration of these datasets allowed us to reveal constitutive differences between sf-C and sf-R.

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.

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 is used as a genetic marker for strain identification in all fall armyworm related publications, including the survey of invasive populations in Africa (Rodney N. Nagoshi et al. 2018). The fact that it is also constitutively 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. 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. 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 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 around 50 pupae sampled in Guadeloupe in 2001 for sf-C and in Florida (Hardee County) in 2012 for sf-R. From the time of their collection they have been reared under laboratory conditions on artificial diet (from Poitout et al. 1972, principal components: 77% H₂O, 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, 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*. All individuals released had emerged the night before the oviposition choice experiment. 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, within the quarantine platform (22°C, 50% humidity, natural dark-light conditions - in November around 14h dark:10h light- with fluorescent light bulbs).

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

508 rice pots, which were changed before the 4th larval instar and each day after this
509 instar until the pupation.

510 From a batch of eggs reared on artificial diet, we subdivided the progeny on the three
511 different diets (corn plant, rice plant and artificial diet). A total of 80 larvae (which
512 hatched the morning of the experiment) were deposited in each cage.

513 Two generations have been conducted on plants; during the first generation we
514 measured life history traits (LHT) for each strain in native and alternative conditions
515 and during the second generation, the larvae had been sampled at 4th larval instar for
516 RNA-Seq experiments.

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

- 524 • Survival (*sv*) is the number of emerging adults counted over the initial number
525 of larvae;
- 526 • Developmental time (*dt*) is the number of days between the beginning of the
527 experimental start until adult emergence (mean on all emerging adult in same
528 cage);
- 529 • The weight (*wt*) of individual larvae and of individual pupae of each sex in mg.
530 The day of plant infestation, we weighed the pool of 80 larvae. Then, from the
531 2nd larval stage, the weight was quantified every other day and each larvae
532 was individually weighed.

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. In parallel, and as a reference point, we performed the same experimental design and measurements on standard rearing conditions on artificial diet (Poitout and Bues 1974). Two replicates of each strain on artificial diet have been set-up from the same batch of L1 larvae from our laboratory strains. Compared to plant conditions, rearing has been performed in a square plastic box with mesh filter for aeration and food supplied *ad libitum*. Since the rearing conditions differ significantly from plant assays, we considered those experiments as reference and not as control.

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, corresponding to offspring of the larvae used to estimate the different components of fitness (survival, weight and developmental time). 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 (**Figure S.17**), 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). Bioanalyzer using 1 µl of total RNA from each sub-pool of three larvae permitted to estimate RNA quantity. The ratio of absorbance 260/280 and 260/230 was used to assess the purity of RNA in each sample. The sub-pools of three larvae, having a good quality (between 1.35 and 2) and quantity (>200 ng/ µl), were pooled again to obtain samples corresponding to the four experimental set-ups. On the one hand, the samples from rice plant containing only three larvae because of the survival problem on rice for both strains. On the other hand, the samples on artificial diet and on maize contained 12 larvae (*i.e.* 4 sub-pools of 3 larvae).

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

581 the TruSeq Stranded mRNA Sample Preparation Kit (Illumina) following
582 manufacturer's protocol.

583 *Reference and annotation*

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

590 *Differential expression analysis*

591 To identify differentially expressed genes, we first mapped reads on gene prediction
592 using Bowtie2 (Langmead and Salzberg 2012). We chose to use the same reference
593 for both the sf-C and the sf-R strain samples. For read mapping we used "very
594 sensitive" parameter setting in Bowtie2, which allowed searching extensively for the
595 best alignment for each read. Counting of aligned reads number to each gene is
596 produced by SAMtools program (Li et al. 2009). Then to detect the genes
597 differentially expressed we used DESeq2 (R package; Love, Anders, and Huber
598 2014). To measure gene expression variations between conditions, DESeq2 uses a
599 negative binomial generalized mixed model. The estimates of dispersion and the
600 logarithmic fold-changes incorporate data-driven prior distributions. Genes were
601 considered differentially expressed if they satisfy a false discovery rate lesser than
602 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 $FDR < 0.01$) 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 **Fig. S17**, 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 individually in a screw-cap 2ml tube containing 1ml of RNAlater (Sigma; R0901).

DNA/RNA extractions

Larvae from field collections were placed in a 1.5ml Eppendorf tube with RLT buffer from Qiagen. Individual 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 (JM-77: ATC ACC TCC ACC TGC AGG ATC and JM-76: 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 (**Fig. S8A**).

For the Tpi genotyping we used the following primers as described (Rodney N. Nagoshi 2010): *Tpi*-56 F (5'-CAAAATGGGTCGCAAATTCG-3') and *Tpi*-850gR (5'-AATTTTATTACCTGCTGTGG-3'). Digestion of the PCR product was made with the Avall enzyme (**Fig. S9A**).

FR1 repeat genotyping was based on PCR amplification only, as described (Rod N. Nagoshi and Meagher 2003a) with the following primers : FR-c (5'-TCGTGTAAAACGTACTTTCTT- 3'), and FR-2 (5'-GACATAGAAGAGCACGTTT-3'). Amplification is then analyzed on agarose gel (**Fig. S10**)

Quantitative PCR

For reverse transcription quantitative PCR, we used the candidate transcript sequence, as retrieved from BIPAA platform* -for example by searching GSSPFG00029721001-RA from **Table S2**- as a template for primer design using Primer3 and asking for a 50 nt amplicon. Primers used are specified in **Table S3**.

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 $\Delta\Delta C_t$ method with the laboratory sf-C strain as a reference point for each gene.

* https://bipaa.genouest.org/sp/spodoptera_frugiperda_pub/

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 funding 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.d'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, 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.

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 51. <https://doi.org/10.18637/jss.v067.i01>.
- 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). <https://doi.org/10.1186/s12862-016-0627-y>.
- 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. <https://doi.org/10.1111/j.1365-2311.2010.01248.x>.
- Crawley, Michael J. 2007. *The R Book*. *The R Book*. <https://doi.org/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). <https://doi.org/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. <https://doi.org/10.1007/s10709-015-9829-2>.
- Goergen, Georg, P. Lava Kumar, Sagnia B. Sankung, Abou Togola, and Manuele Tam?? 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). <https://doi.org/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). <https://doi.org/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*. <https://doi.org/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. <https://doi.org/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*. <https://doi.org/10.1371/journal.pgen.1000834>.
- Hill, Geoffrey E. 2015. "Mitonuclear Ecology." *Molecular Biology and Evolution* 32 (8): 1917–27. <https://doi.org/10.1093/molbev/msv104>.
- Jaenike, J. 1990. "Host Specialization in Phytophagous Insects." *Annual Review of Ecology and Systematics*. <https://doi.org/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). <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1111/evo.12940>.
- Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1186/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1016/j.ibmb.2006.05.006>.
- 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. [https://doi.org/10.1653/0015-4040\(2003\)086](https://doi.org/10.1653/0015-4040(2003)086).
- 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. [https://doi.org/10.1653/0015-4040\(2004\)087\[0454:LDOFAL\]2.0.CO;2](https://doi.org/10.1653/0015-4040(2004)087[0454:LDOFAL]2.0.CO;2).
- 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. <https://doi.org/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*, 095596. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1046/j.1365-2583.2003.00429.x>.
- 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. [https://doi.org/10.1603/0013-8746\(2003\)096\[0329:FTSIFA\]2.0.CO;2](https://doi.org/10.1603/0013-8746(2003)096[0329:FTSIFA]2.0.CO;2).
- 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*. <https://doi.org/10.1603/0046-225X-33.4.881>.
- 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*. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1603/AN09046>.
- Nagoshi, Rodney N., Georg Goergen, Kodjo Agbeko Tounou, Komi Agboka, Djima Koffi, and Robert L. Meagher. 2018. "Analysis of Strain Distribution, Migratory Potential, and Invasion History of Fall Armyworm Populations in Northern Sub-Saharan Africa." *Scientific Reports*. <https://doi.org/10.1038/s41598-018-21954-1>.
- O'Brien, Michael J, and Kevin N Laland. 2012. "Genes, Culture and Agriculture: An Example of Human Niche Construction." *Current Anthropology* 53 (4): 434–70. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1093/aesa/80.6.731>.
- 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 Lépidoptères Noctuidae et de Deux Espèces d'Arctiidae Sur Milieu Artificiel Simple. Particularités de l'élevage Selon Les Espèces ." *Annales de Zoologie, Ecologie Animale* 6 (3): 431-41.
- Powell, 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. [https://doi.org/10.1603/0013-8746\(2004\)097\[1034:MGAOHU\]2.0.CO;2](https://doi.org/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. <https://doi.org/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. <https://doi.org/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." *Genes & Genomes Genetics* 2 (2): 313-19. <https://doi.org/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): 1-15. <https://doi.org/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*. <https://doi.org/10.1093/bfgp/elv015>.
- Sparks, Alton N. 1979. "A Review of the Biology of the Fall Armyworm." *The Florida*

- Entomologist*. <https://doi.org/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. <https://doi.org/10.1146/annurev.es.08.110177.001045>.
- Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PLoS ONE* 6 (7). <https://doi.org/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*. <https://doi.org/10.1111/j.1570-7458.1988.tb02275.x>.
- 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. <https://doi.org/10.2307/3495426>.
- 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*. <https://doi.org/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*. <https://doi.org/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-C-E**) and sf-R (**B-D-F**) 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. Different letters above bars indicate significant differences in the proportion of egg masses laid per plant/cage netting ($P < 0.05$).

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

(**A-D**) Pupal weight (wt) is measured in duplicate for sf-C (**A-C**) and sf-R (**B-D**) according to plant diet: corn (yellow) and rice (green). We measured separately females (**A-B**) and male (**C-D**) pupae.

(**E,F**) Survival (sv) rate is measured from the 1st larval instar to adult emergence for sf-C (**E**) and sf-R (**F**) 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-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 4: RNA-Seq of individual larvae from the fields

A. Genotypes of the individual L4 larvae from natural populations used for RNAseq studies. Col, 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. Heterozygosity indicates a male. **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 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. **C.** Heatmap of expression variations (expressed as z-scores) of the sf-R specific expressed genes across all RNAseq experiments. 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. **D.** 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.

Supplementary Information

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 - Developmental time until adult emergence in sf-C (**A**) and sf-R (**B**) according to the diet: corn plant (yellow), rice plant (green). The variation between replicates is represented by the standard error and the different letters above bars indicate significant differences between plant diets for each strain ($P < 0.05$).

Fig. S3 - Survival from egg hatching for 50 individuals reared on artificial diet with low (exp. #1) or high (exp. #2) hygrometry

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 ($\Delta\Delta C_t$ 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 *Avall* 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 ($\Delta\Delta C_t$ 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 row 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 row 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 - Sf-C associated gene expression across all RNAseq experiments.

These genes 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. S16 - 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. S17 - 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 - 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

1118 observed in RNAseq experiments and the $\Delta\Delta C_t$ values obtained by qPCR. Except for
1119 peroxidase, all genes tested show a confirmed overexpression of these genes in sf-
1120 R.

1121

1122 **Table S3** - Candidate genes primers sequences for qPCR used in **Table S2**

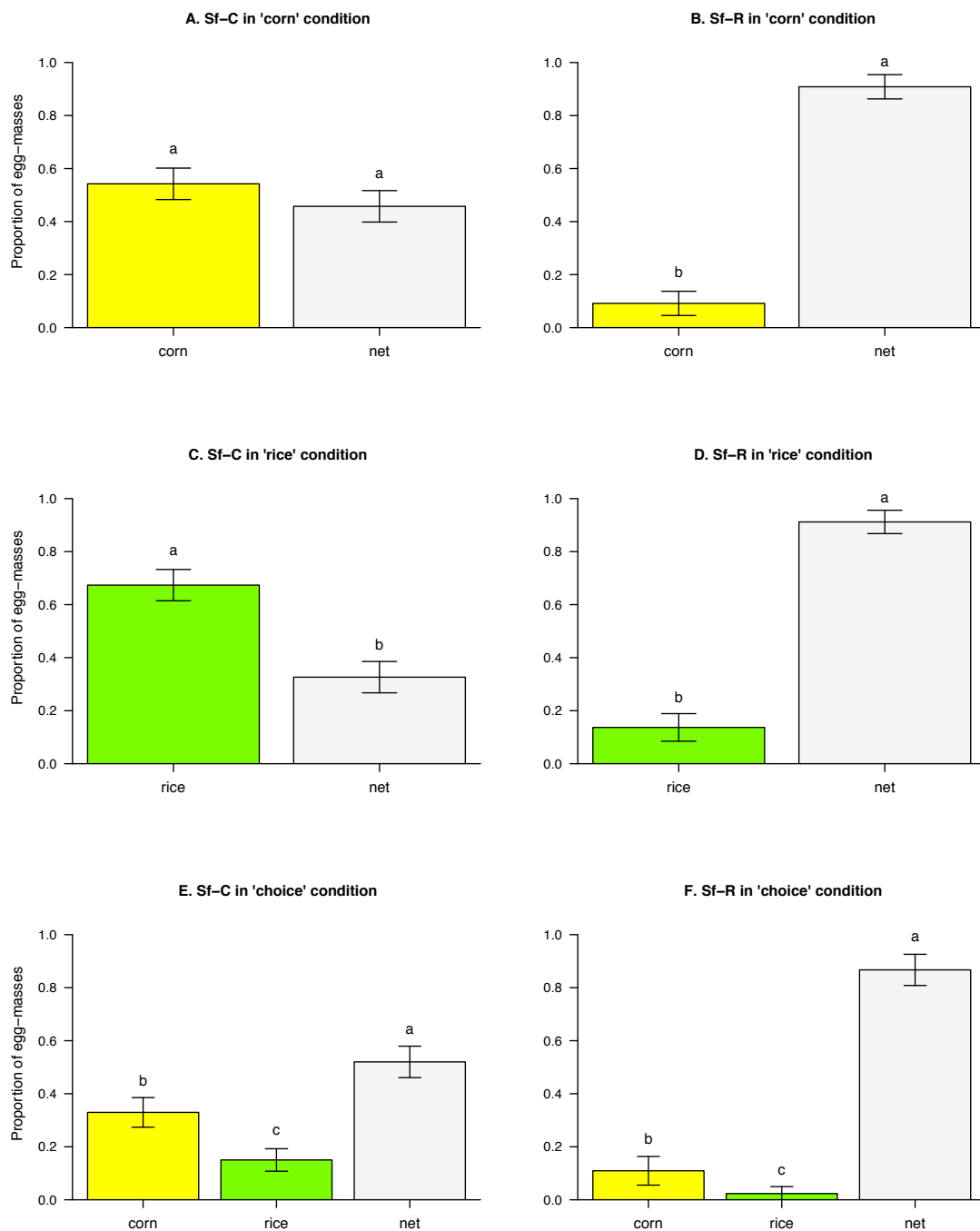


FIGURE 1

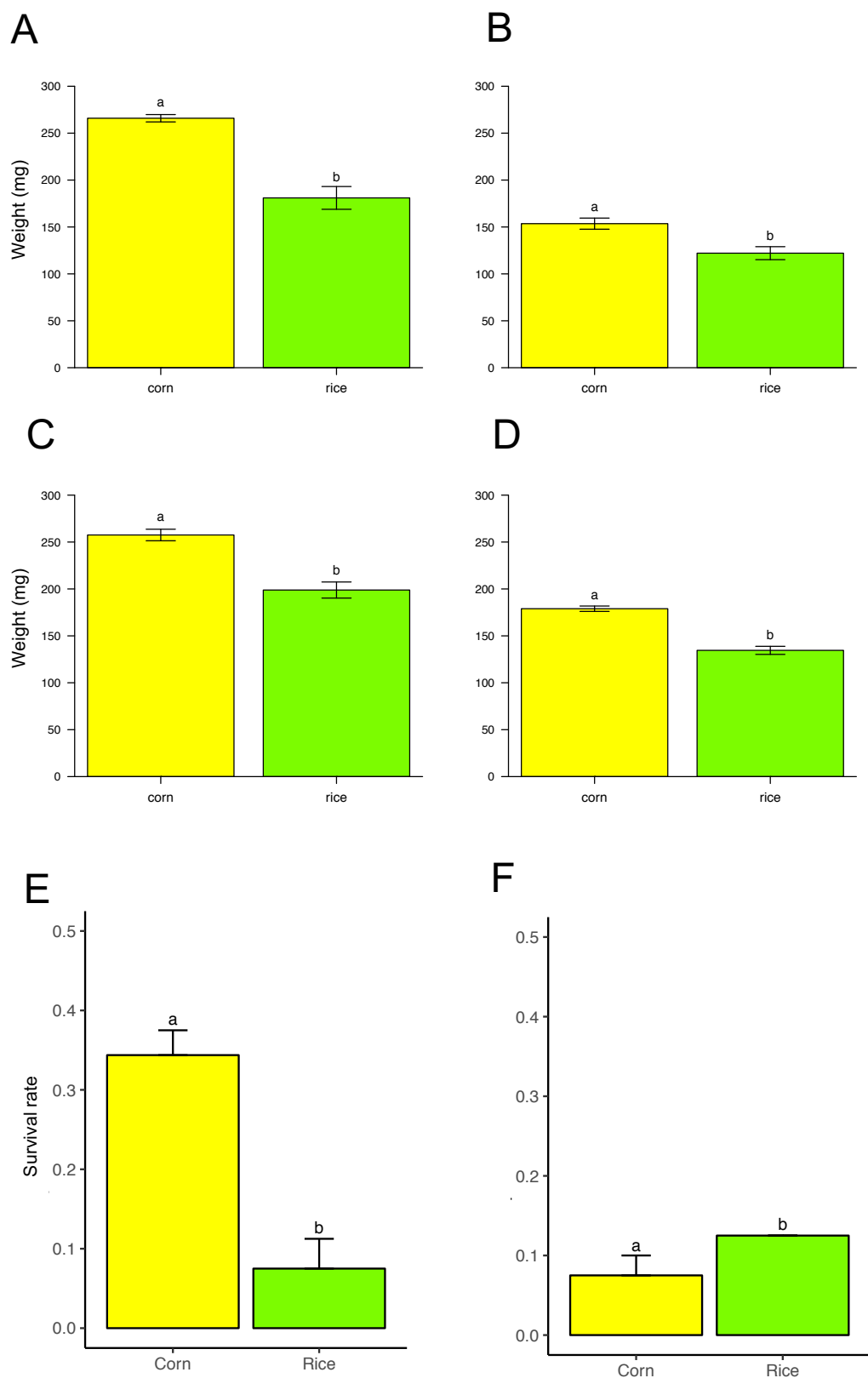
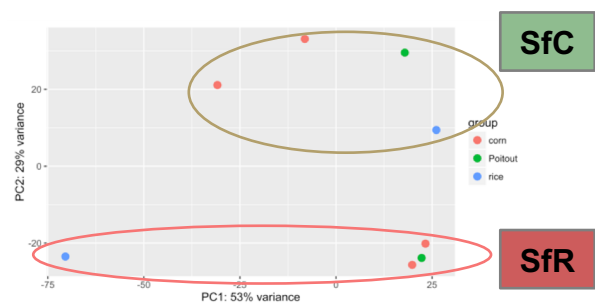


FIGURE 2

A



B

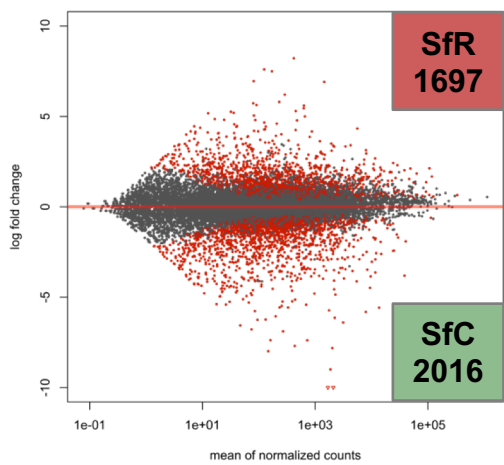
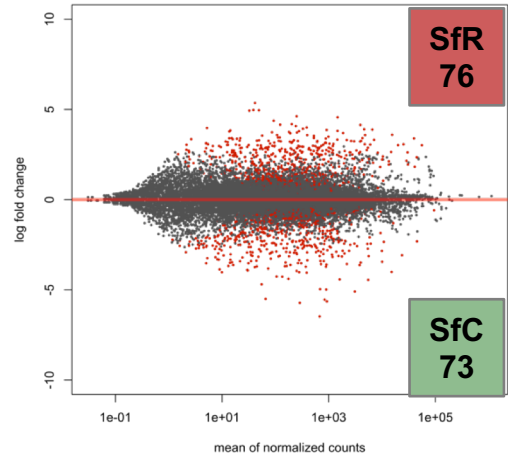


FIGURE 3

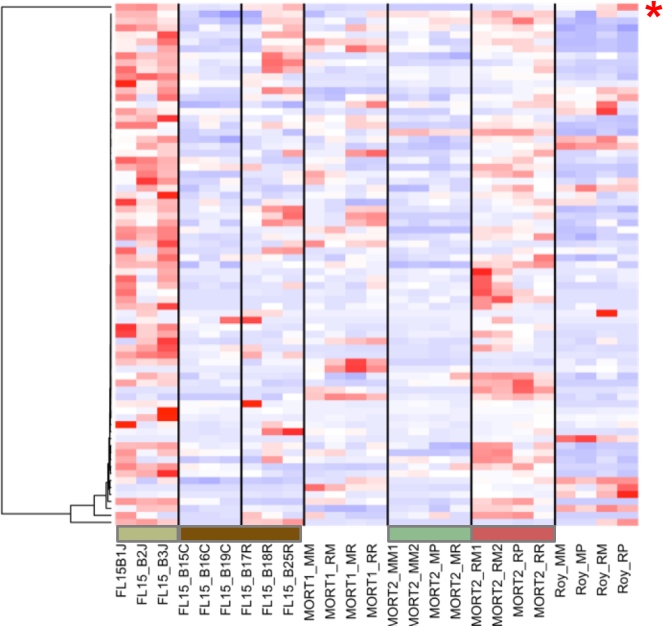
A

Tifton							Jacksonville		
	B15	B16	B19	B17	B18	B25	B1	B2	B3
Col	Green			Purple			Purple		
tpi									
FR									
Sex	F	M	F	F	F	M	M	M	M

B



C



D

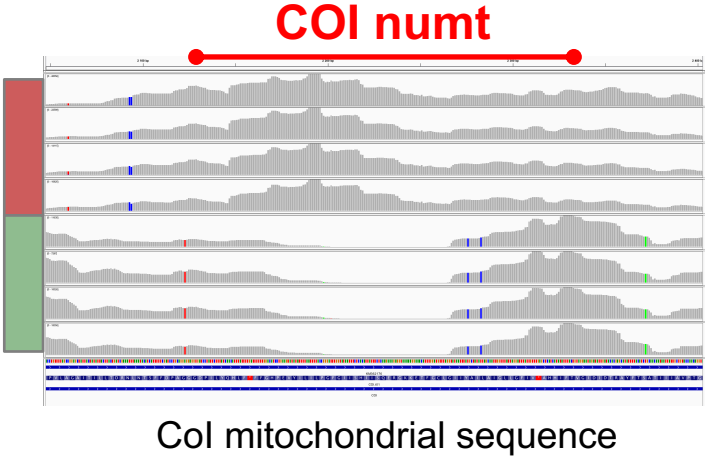


FIGURE 4

FIGURE S1

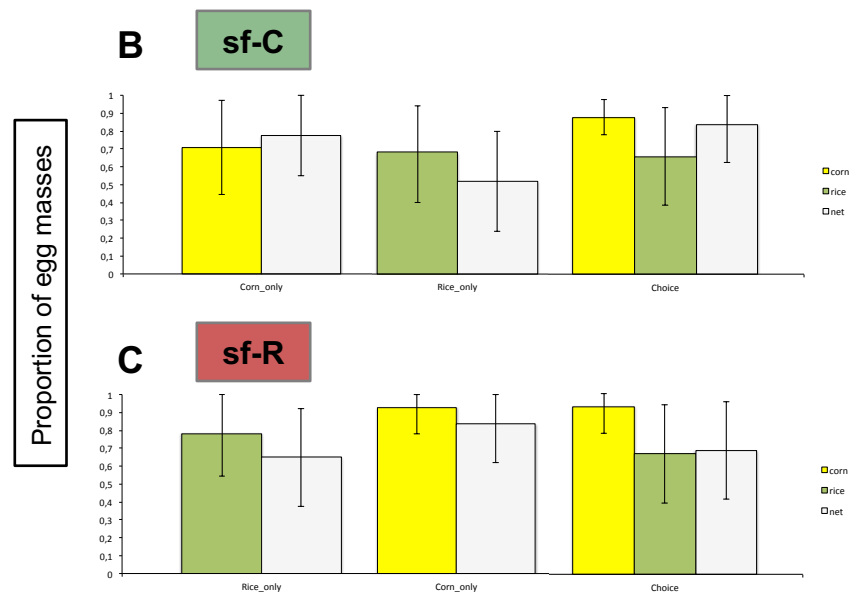
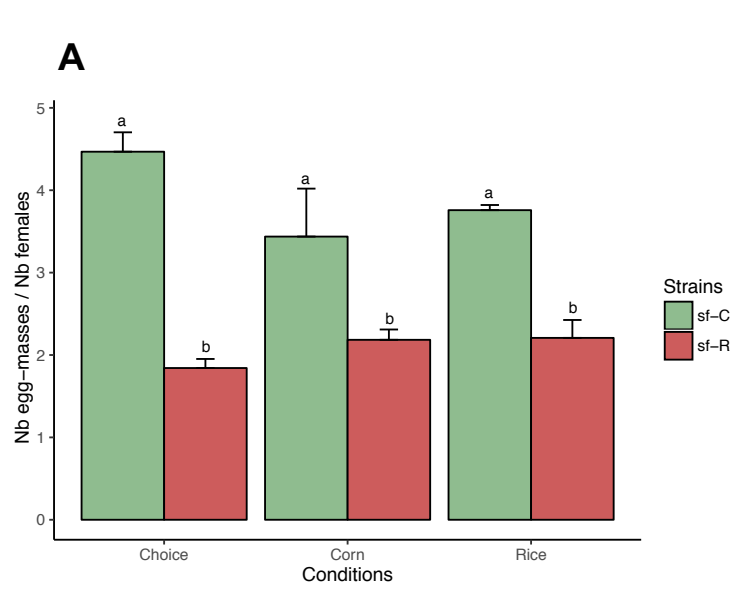


FIGURE S2

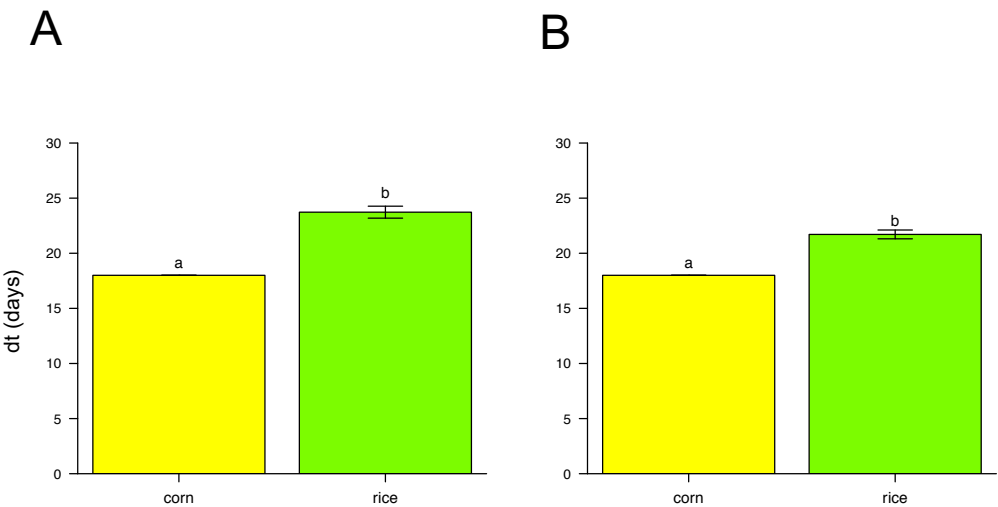


FIGURE S3

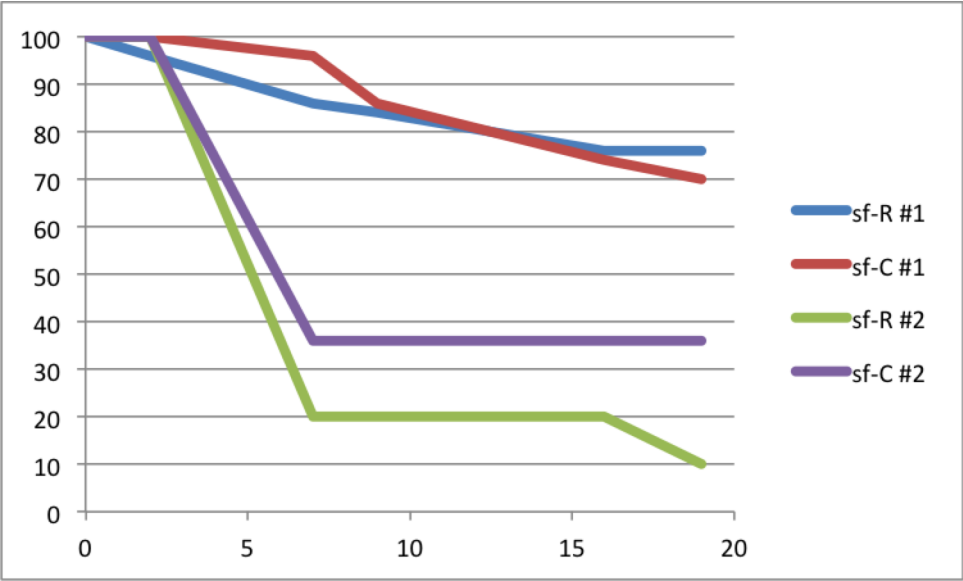
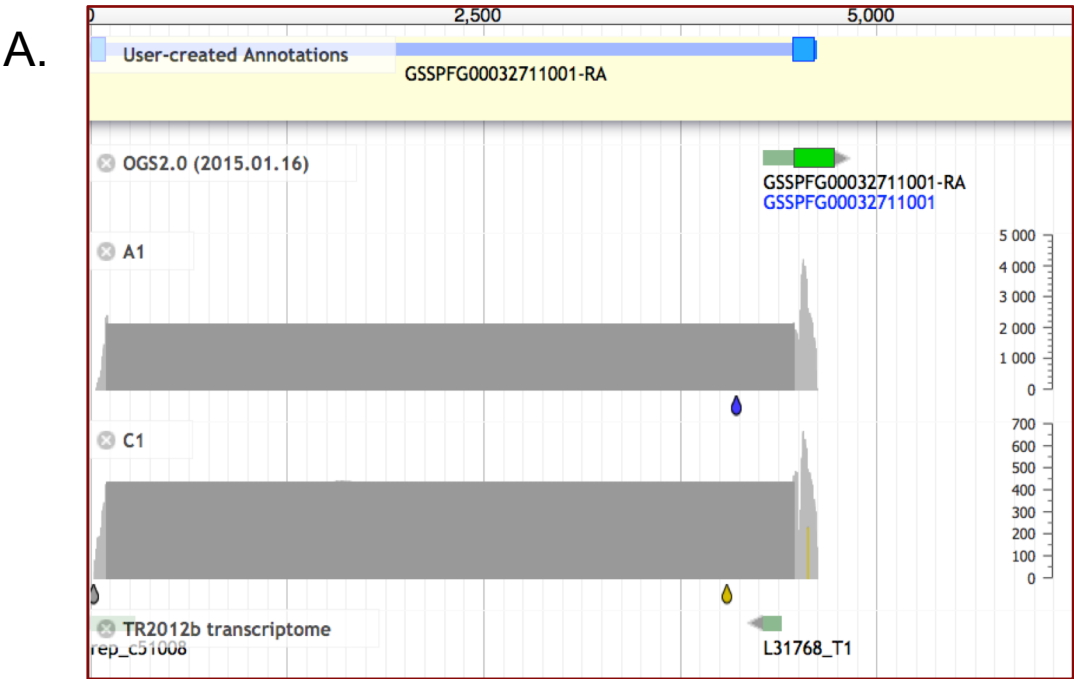


FIGURE S4



B.

Sequences producing significant alignments:		(Bits)	Value
gb AEA76321.1	polycalin [Mamestra configurata]	62.8	4e-09
gb AJQ81210.1	polycalin [Helicoverpa armigera]	56.6	4e-07
ref XP_012553082.1	PREDICTED: chlorophyllide A binding prote...	57.0	4e-07
ref NP_001037071.1	chlorophyllide A binding protein precurs...	57.0	4e-07
gb AGM34046.1	chlorophyllide A binding protein precursor [Bo...	57.0	4e-07
ref XP_012553081.1	PREDICTED: chlorophyllide A binding prote...	57.0	4e-07
gb ACB54957.2	polycalin [Helicoverpa armigera]	56.2	5e-07
gb ACB54956.1	polycalin [Helicoverpa armigera]	56.2	6e-07
gb ACB54951.1	polycalin [Helicoverpa armigera]	56.2	6e-07
gb ABU98612.1	multi-domain lipocalin [Helicoverpa armigera]	55.8	7e-07

FIGURE S5

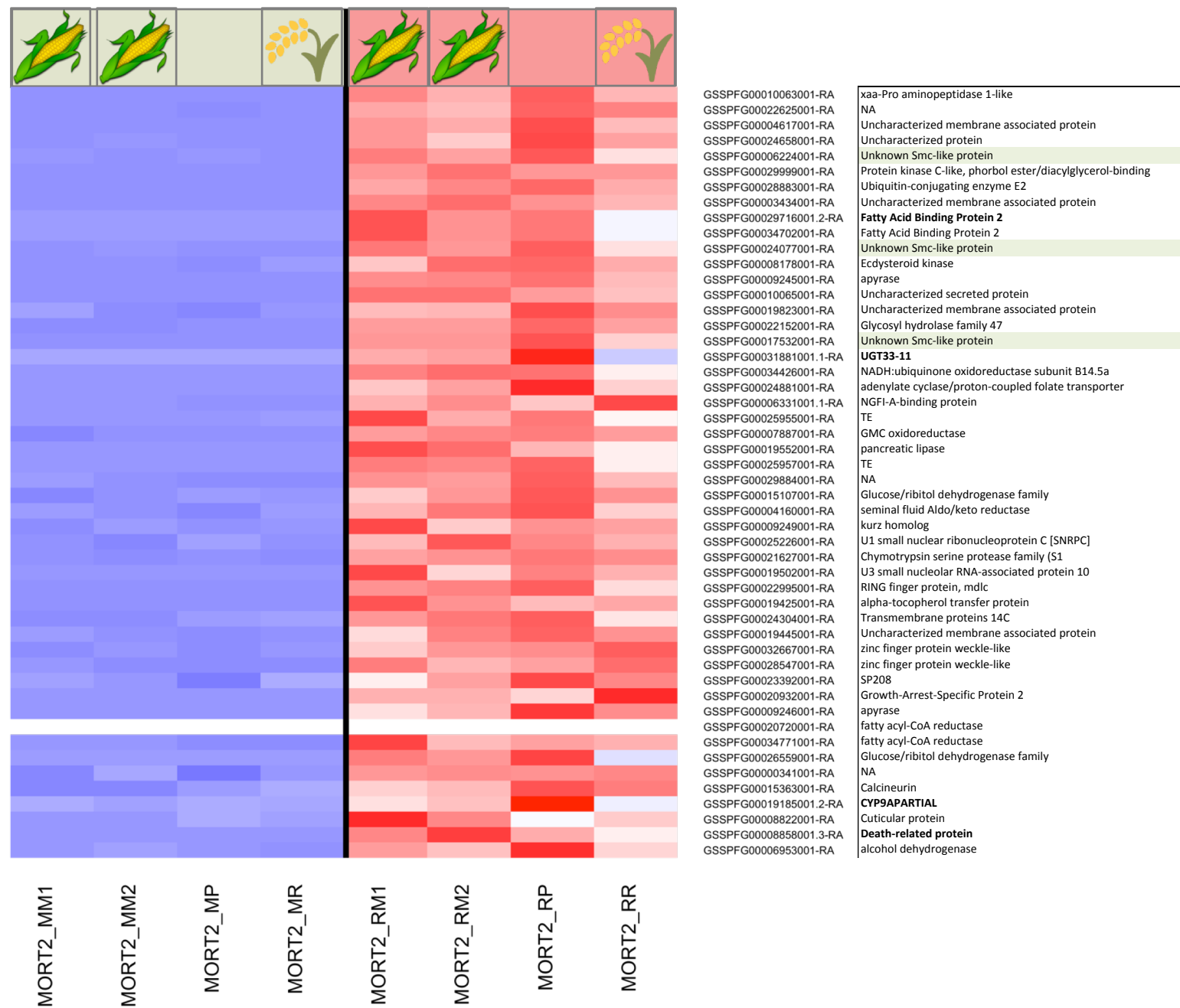


FIGURE S6

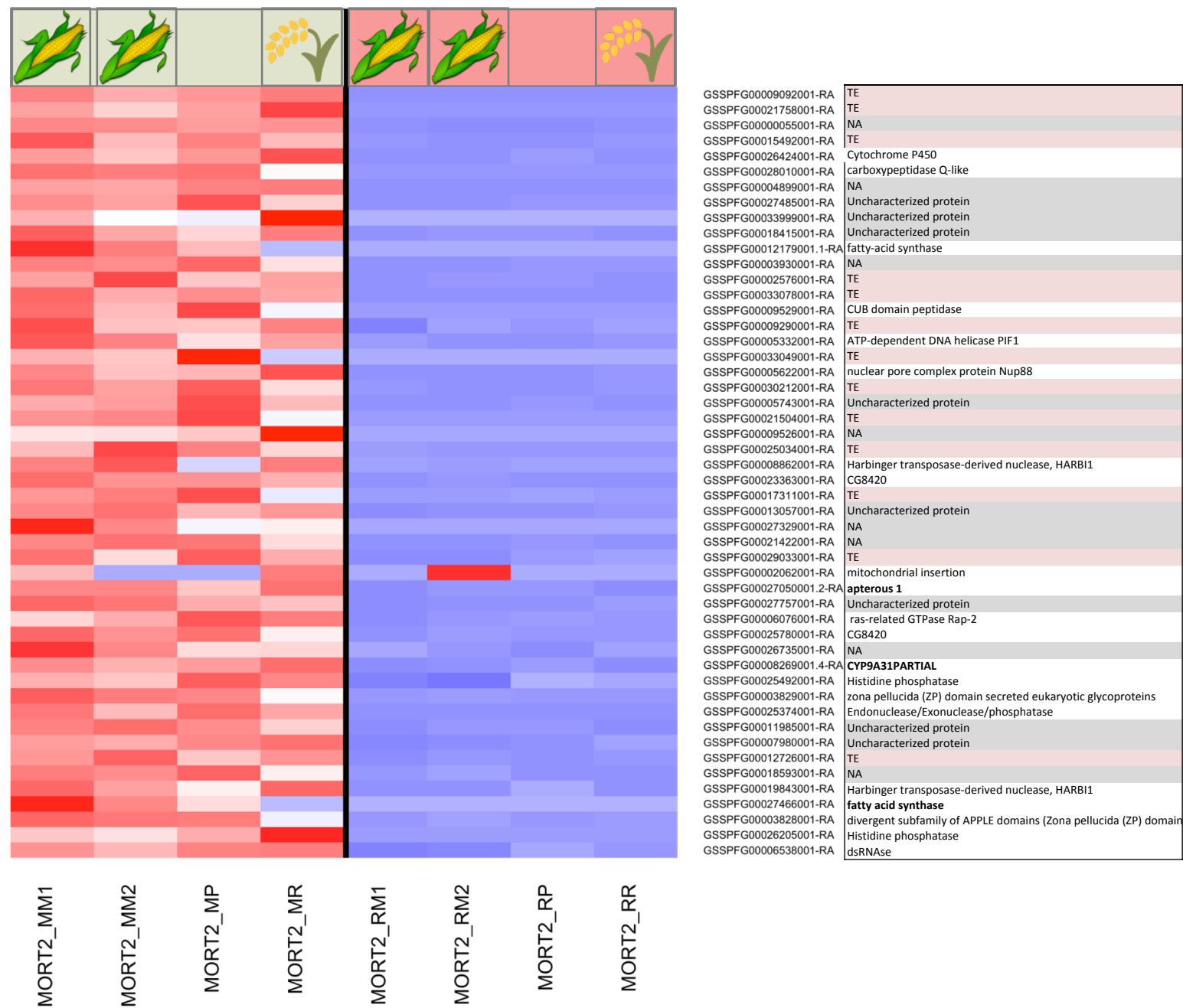
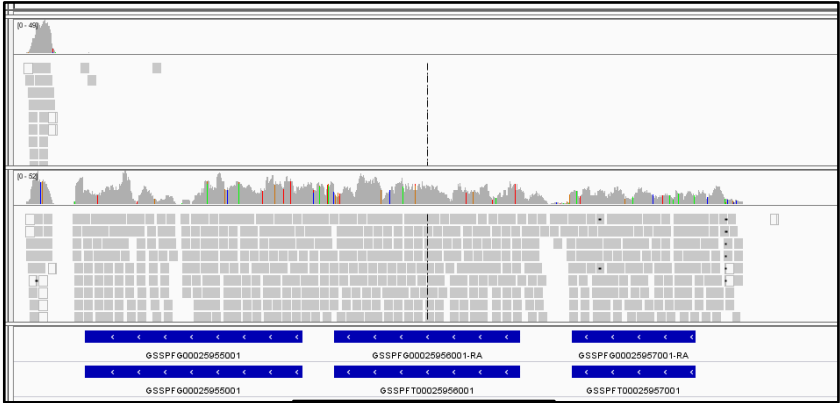


FIGURE S7

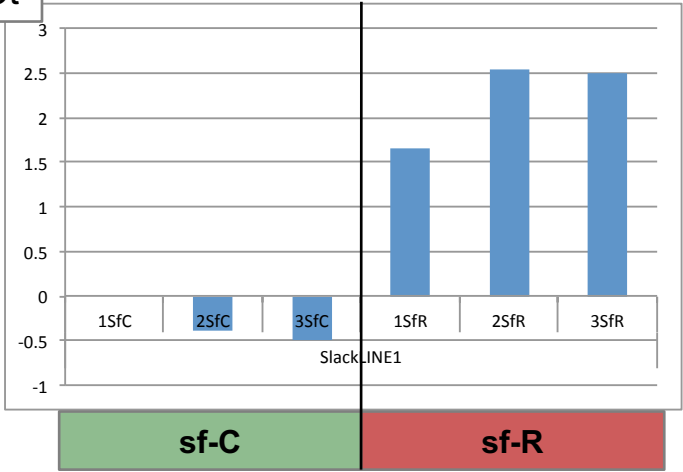
sf-C

sf-R



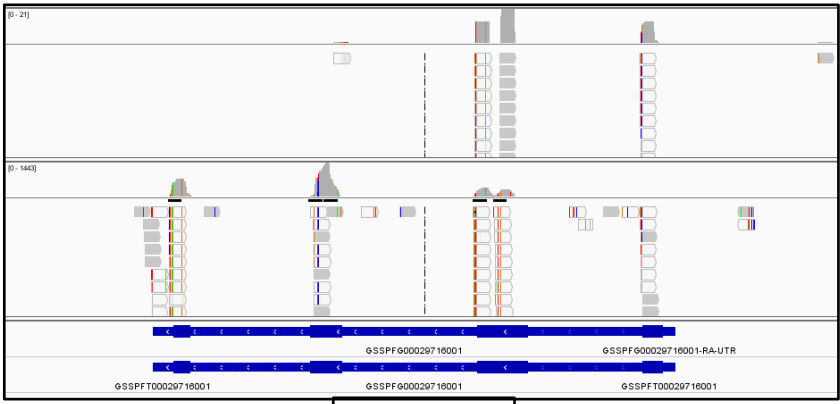
slack_LINE1

$\Delta\Delta Ct$



sf-C

sf-R



FABP10

$\Delta\Delta Ct$

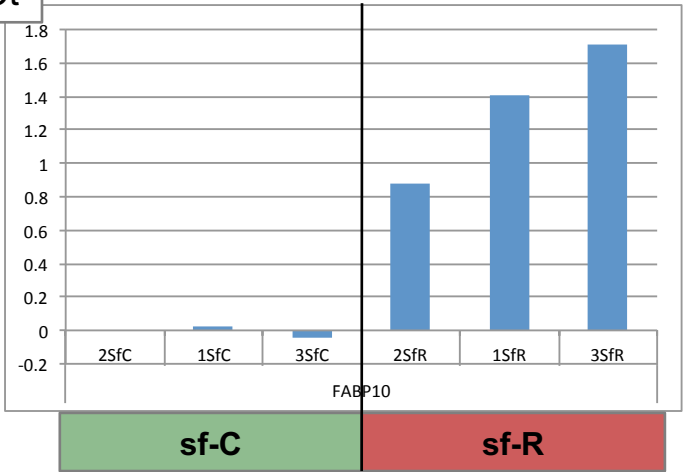
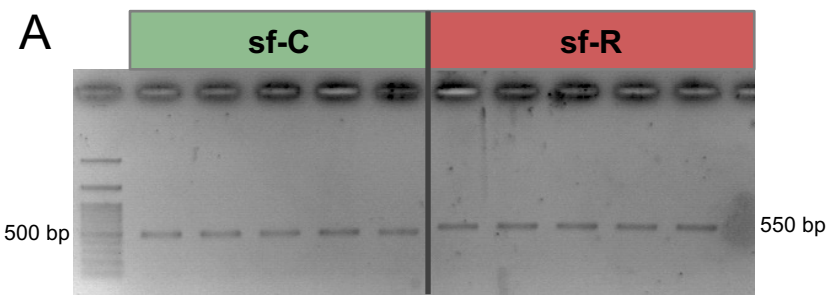
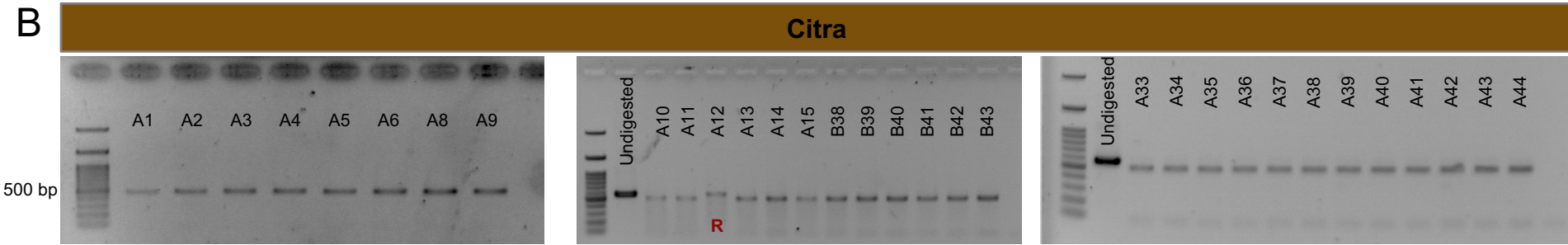


FIGURE S8

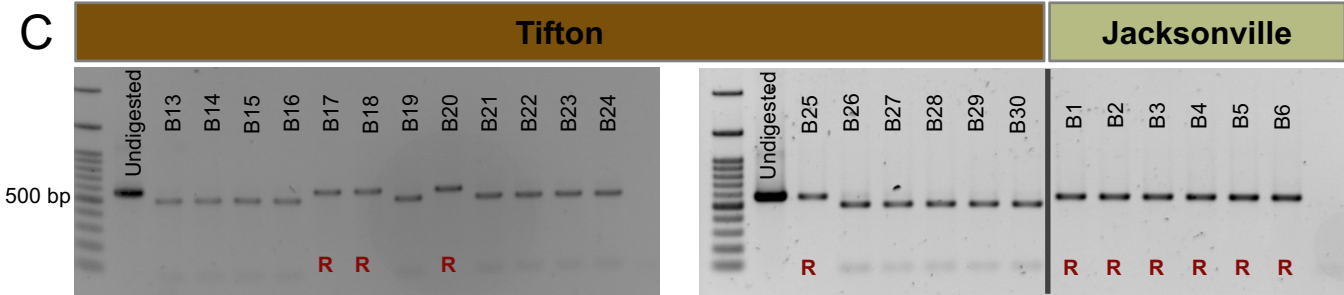
A



B



C

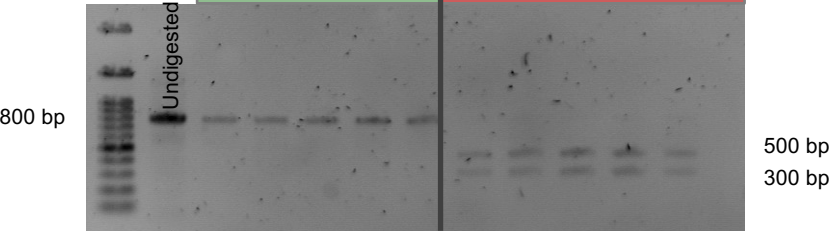


D

	Citra		Tifton		Jacksonville	
	SfC	SfR	SfC	SfR	SfC	SfR
Col	31	1	14	4	0	6

FIGURE S9

A



B

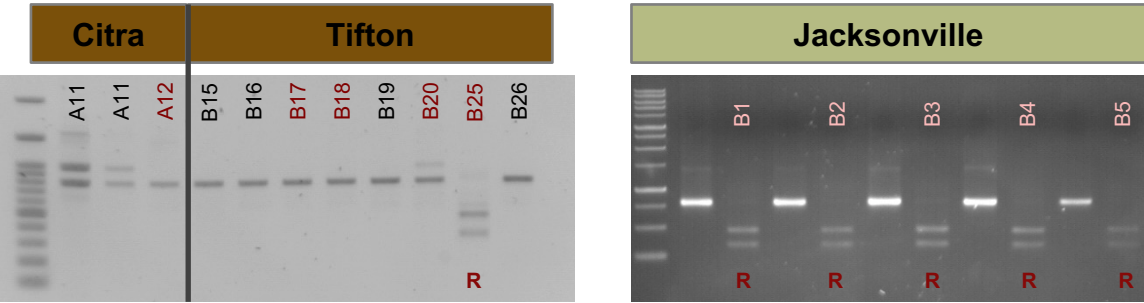
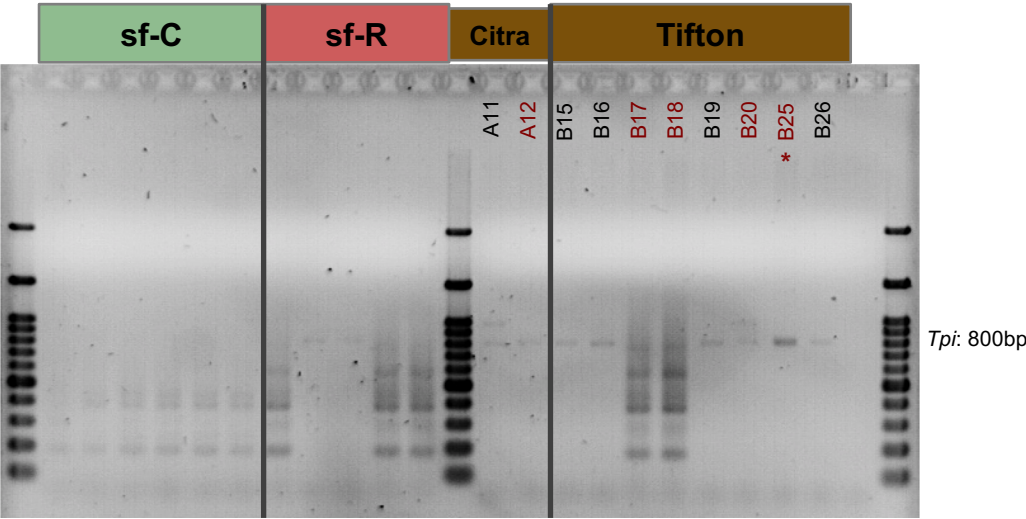


FIGURE S10

A



B

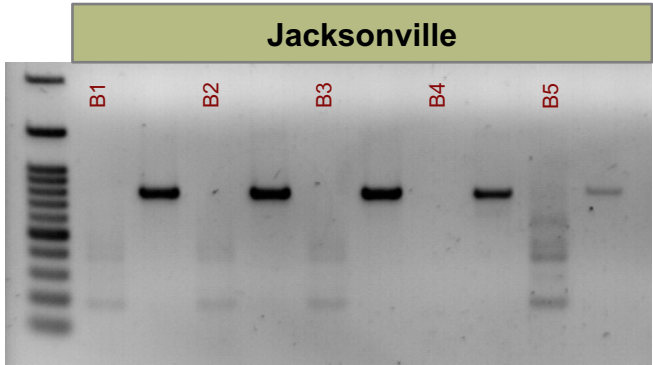


FIGURE S11

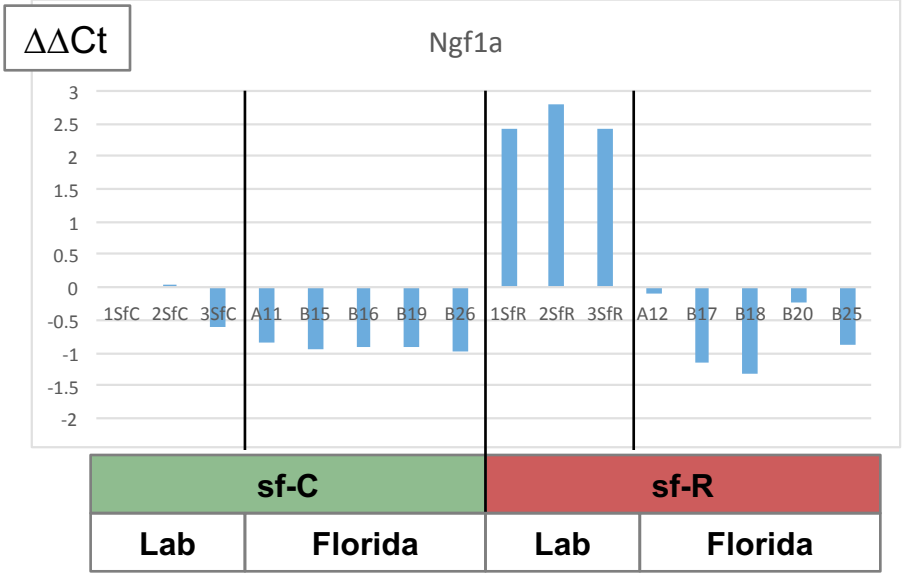
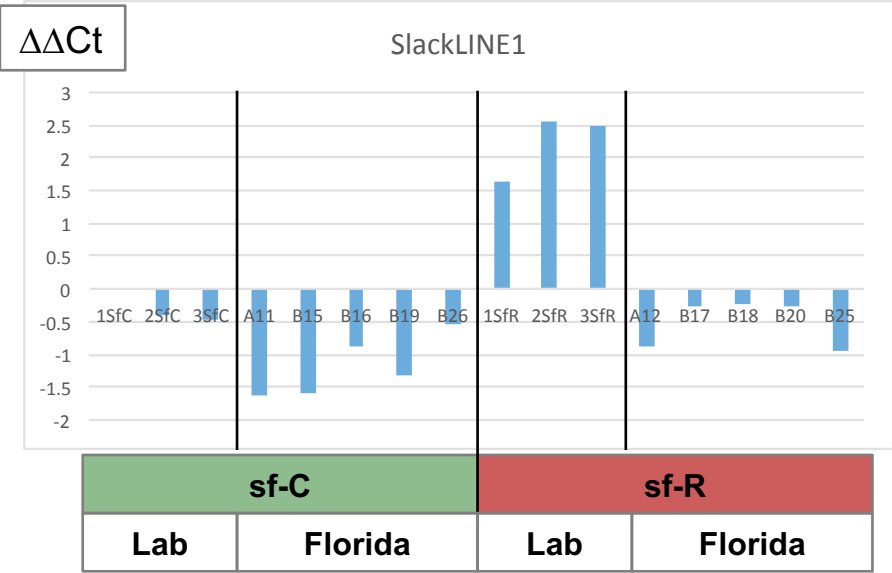
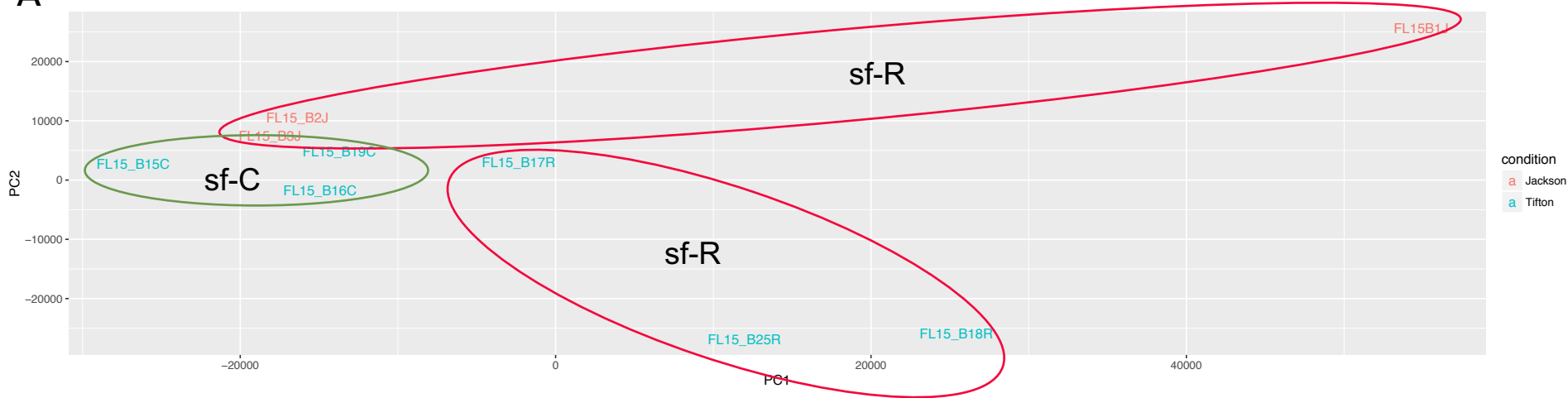
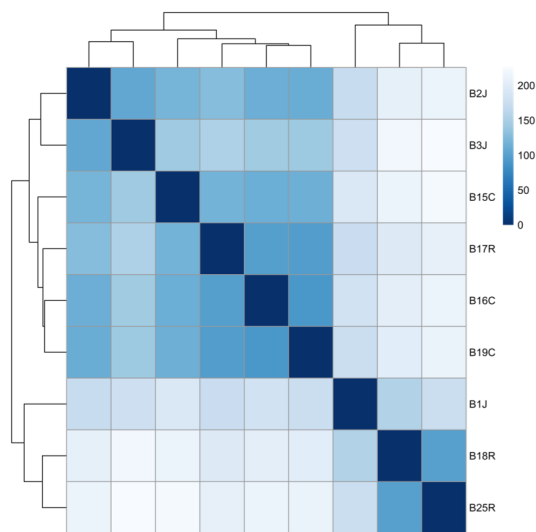


FIGURE S12

A



B



C

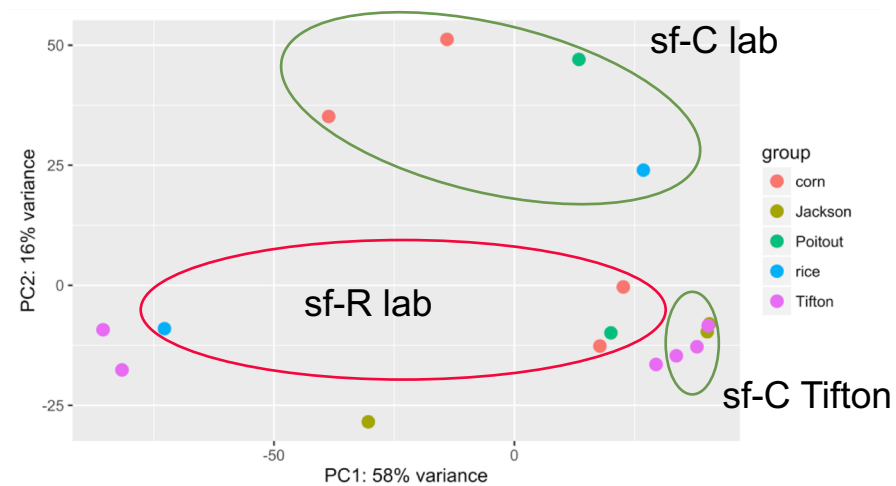


FIGURE S13

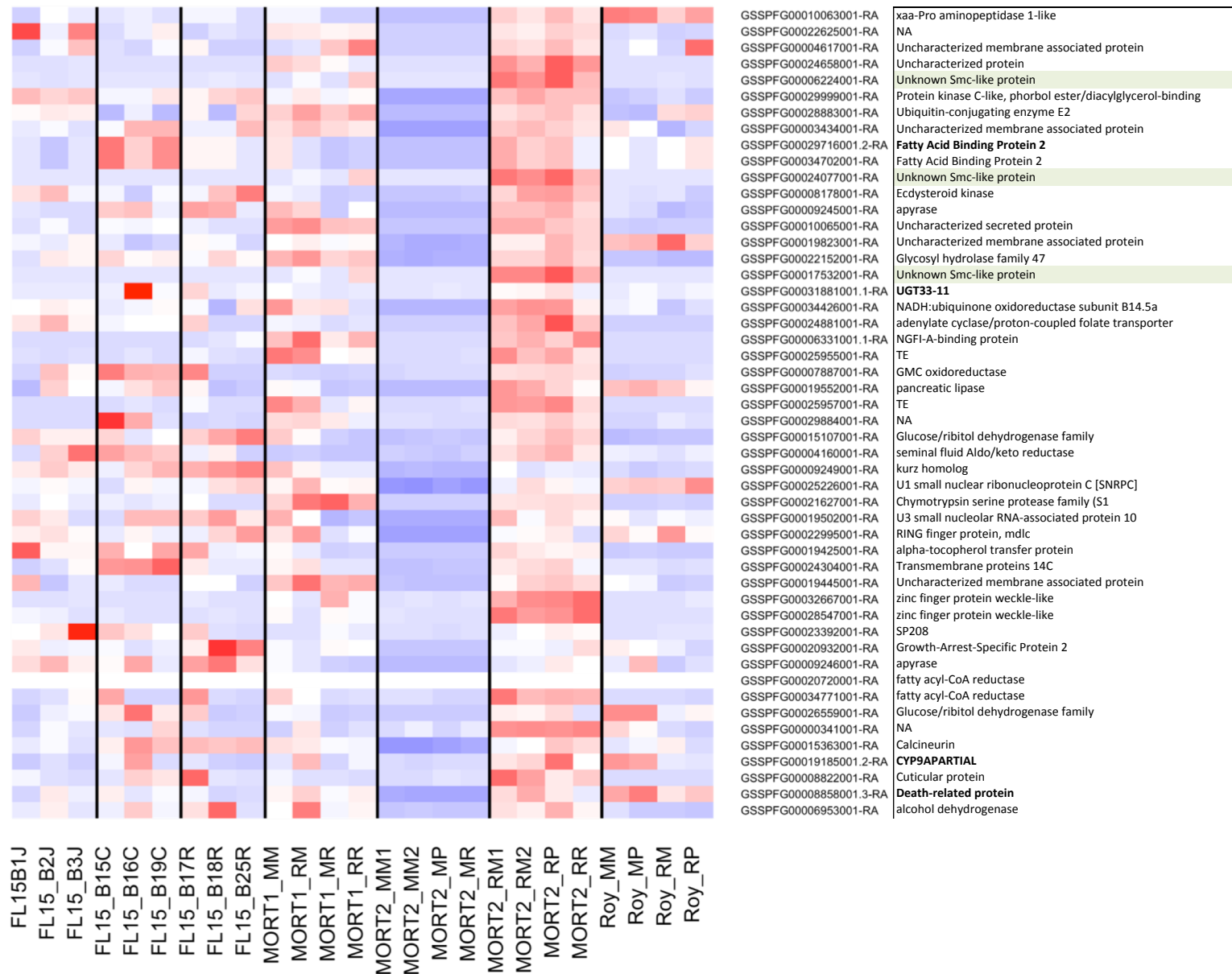


FIGURE S14

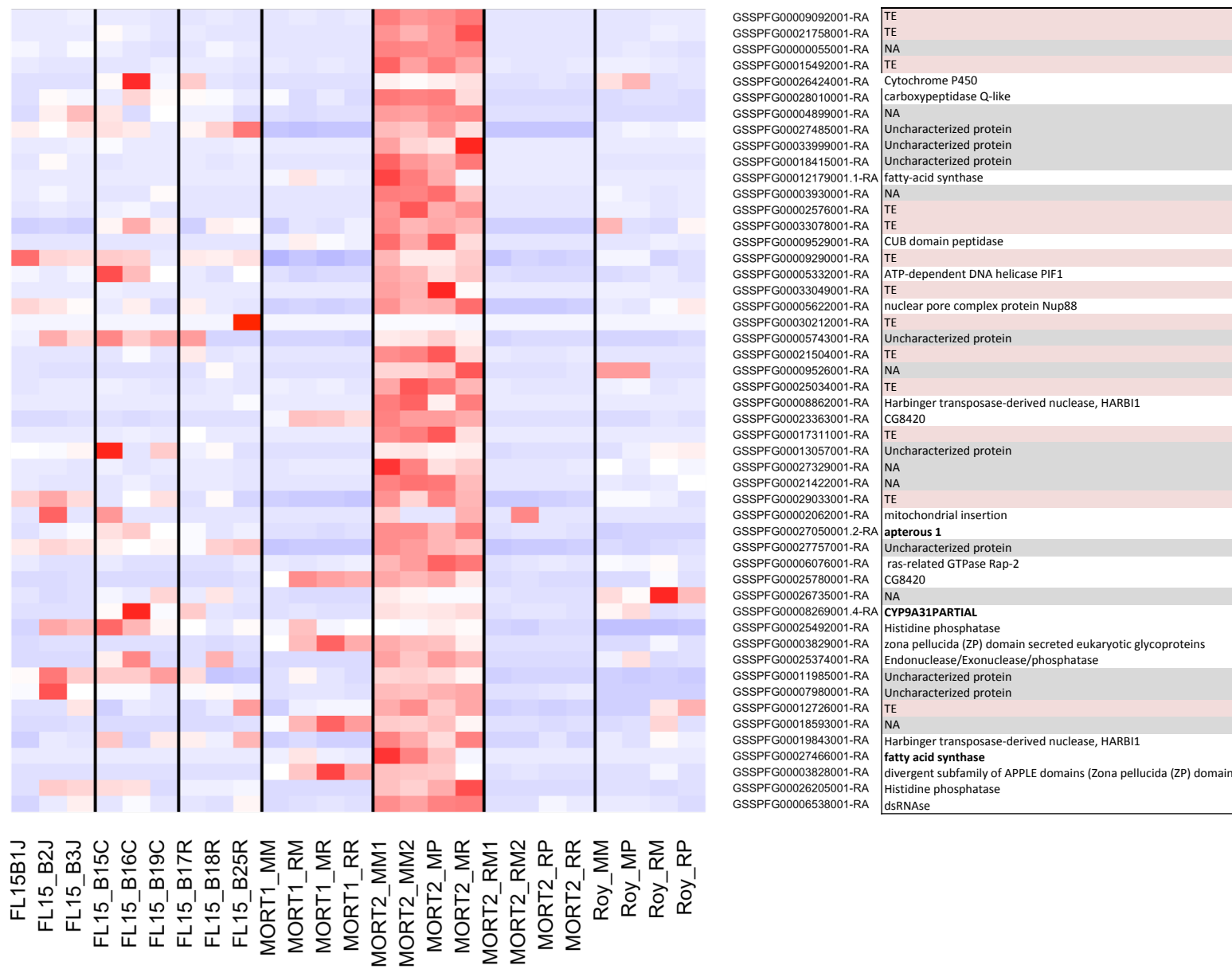


FIGURE S15

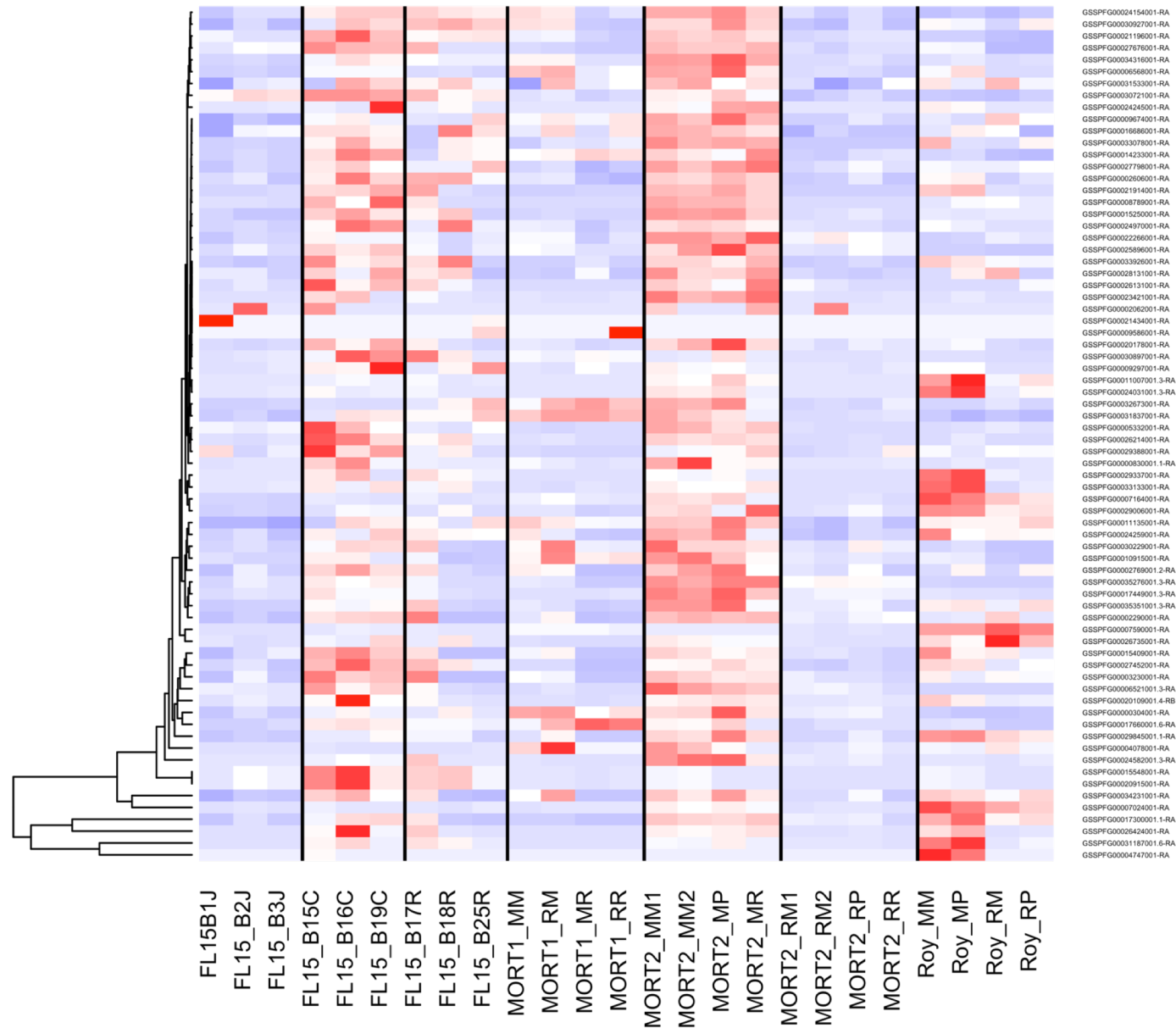
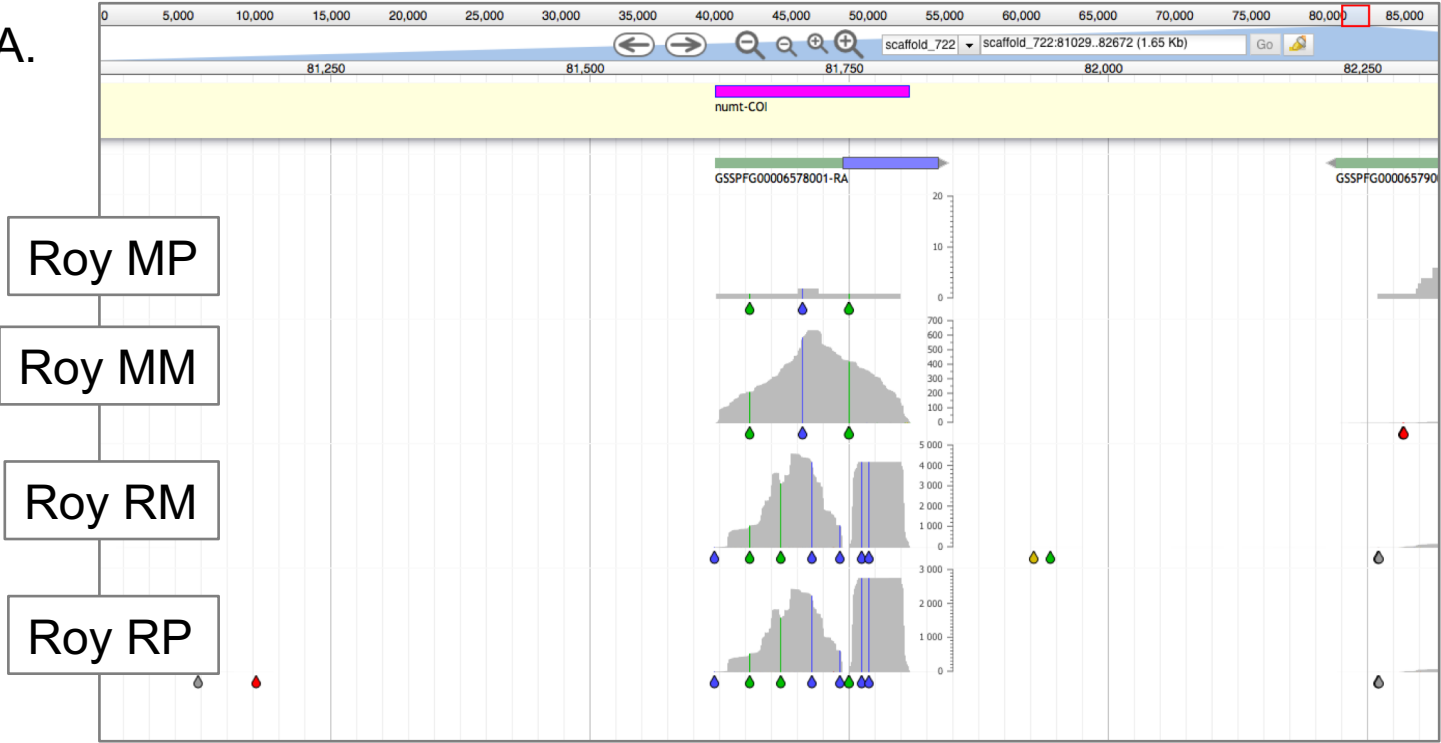


FIGURE S16

A.



B.

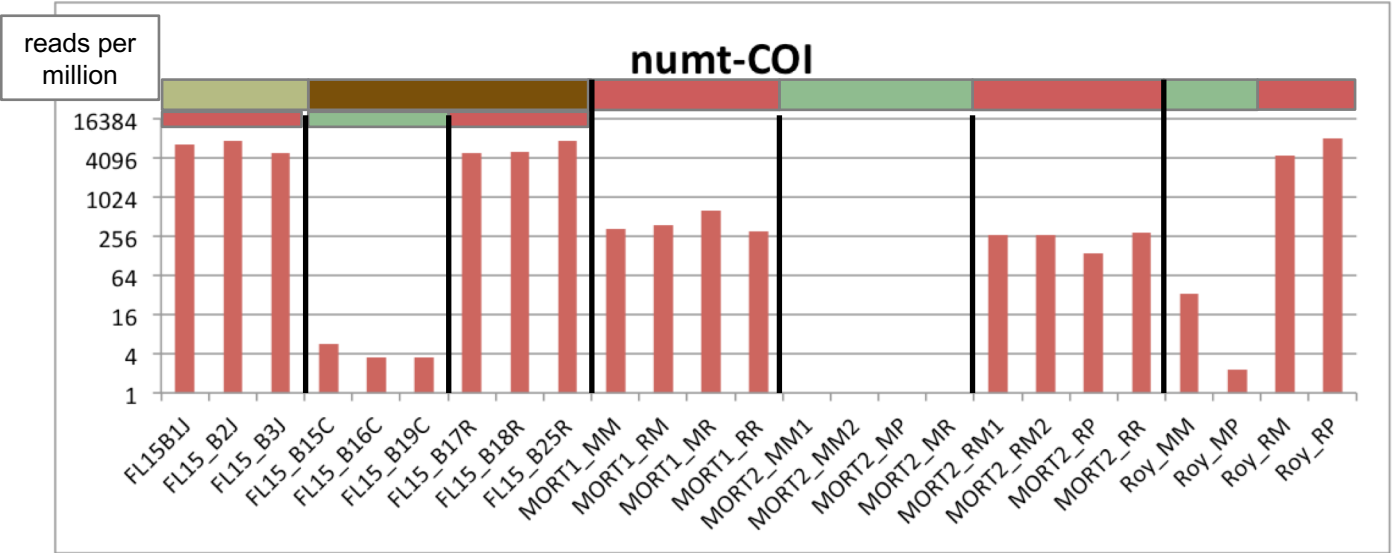
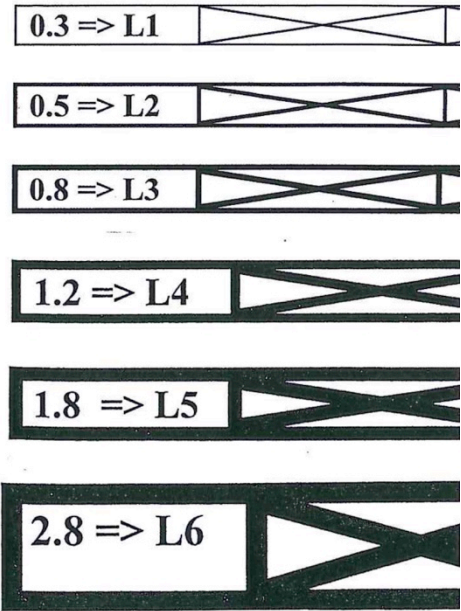


FIGURE S17

A.



B.



Table S1

sample	Strain	Diet	Total_reads	Bowtie2 alignment			
				aligned 0 times	aligned exactly 1 time	aligned > 1 times	fraction_mapped
RT experiment							
MORT2_MM1	sf-C	Corn	67150390	18011798 (26.82%)	30746676 (45.79%)	18391916 (27.39%)	73.18%
MORT2_MM2	sf-C	Corn	43617452	11210535 (25.70%)	20344243 (46.64%)	12062674 (27.66%)	74.30%
MORT2_MP	sf-C	Poitout	31801441	8475906 (26.65%)	14912937 (46.89%)	8412598 (26.45%)	73.35%
MORT2_MR	sf-C	Rice	48323710	12899202 (26.69%)	22473995 (46.51%)	12950513 (26.80%)	73.31%
MORT2_RM1	sf-R	Corn	33742585	8997746 (26.67%)	15013813 (44.50%)	9731026 (28.84%)	73.33%
MORT2_RM2	sf-R	Corn	35347649	9334932 (26.41%)	15943900 (45.11%)	10068817 (28.49%)	73.59%
MORT2_RP	sf-R	Poitout	63139685	19832315 (31.41%)	27649786 (43.79%)	15657584 (24.80%)	68.59%
MORT2_RR	sf-R	Rice	70682628	21958766 (31.07%)	32163009 (45.50%)	16560853 (23.43%)	68.93%
Second generation RT experiment							
MORT1_MM	sf-R	Corn	36304954	9214731 (25.38%)	20459886 (56.36%)	6630337 (18.26%)	74.62%
MORT1_MR	sf-R	Rice	46719601	12461844 (26.67%)	27211195 (58.24%)	7046562 (15.08%)	73.33%
MORT1_RM	sf-R	Corn	41858774	9927539 (23.72%)	24195882 (57.80%)	7735353 (18.48%)	76.28%
MORT1_RR	sf-R	Rice	37354506	8593642 (23.01%)	22722574 (60.83%)	6038290 (16.16%)	76.99%
Natural populations							
FL15_B15C	sf-C	Corn	58940405	27961728 (47.44%)	23352881 (39.62%)	7625796 (12.94%)	52.56%
FL15_B16C	sf-C	Corn	74388159	28113552 (37.79%)	37128525 (49.91%)	9146082 (12.30%)	62.21%
FL15_B19C	sf-C	Corn	33627219	12941686 (38.49%)	16468020 (48.97%)	4217513 (12.54%)	61.51%
FL15_B17R	sf-R	Corn	39842098	15290479 (38.38%)	19573496 (49.13%)	4978123 (12.49%)	61.62%
FL15_B18R	sf-R	Corn	78623719	28419568 (36.15%)	41790045 (53.15%)	8414106 (10.70%)	63.85%
FL15_B25R	sf-R	Corn	23392758	8936331 (38.20%)	12240977 (52.33%)	2215450 (9.47%)	61.80%
FL15_B2J	sf-R	Grass	33537139	12307865 (36.70%)	17130166 (51.08%)	4099108 (12.22%)	63.30%
FL15_B3J	sf-R	Grass	42191185	16293397 (38.62%)	20834698 (49.38%)	5063090 (12.00%)	61.38%
FL15_B1J	sf-R	Grass	24904583	10145929 (40.74%)	11875708 (47.68%)	2882946 (11.58%)	59.26%

Table S2

OGS2.2	Annotation	Abbreviation	log2FC	$\Delta\Delta Ct$
GSSPFG00029721001-RA	S01.UNA + repeat motif	501VNA	4.841451941	3.362002334
GSSPFG00024881001-RA	adenylate cyclase	adenylate cyclase	4.636752333	1.816379812
GSSPFG00010063001-RA	xaa-Pro aminopeptidase 1-like	aminopeptidase	5.128397542	2.967979508
GSSPFG00035209001.5-RA	carboxylesterase 016c	carboxylesterase	6.776645953	2.548145672
GSSPFG00004817001.2-RA	Polycalin1_other-exons	cohesin4817	7.782372143	2.582019629
GSSPFG00031119001.2-RA	CYP340L	CYP	6.046285317	6.149297371
GSSPFG00017290001.2-RC	CYP340L1	CYP340L1	5.379910588	3.556942464
GSSPFG00002985001-RA	delta-24-sterol reductase	d245reductase	6.111188925	2.522842306
GSSPFG00029999001-RA	DEF8	Def8	4.771051161	2.347034643
GSSPFG00031106001.2-RA	DUF4602; C1orf131 homolog	DUF4601	6.932457522	1.390933337
GSSPFG00002727001-RA	Lipocalin - nitrobinding domain - DUF1794 protein	DVF1794	5.584081573	12.00773469
GSSPFG00029716001.2-RA	FABP	FABP10	6.361577058	1.334187653
GSSPFG00034702001-RA	FABP	FABP12	6.605986931	1.368224598
GSSPFG00020720001-RA	FAR	FAR-X	5.826344504	1.674146
GSSPFG00018006001-RA	Glycogen synthase	glyc synt	5.635375812	2.062591316
GSSPFG00024097001-RA	Hemicentin 2	hemicentin2	5.658948506	3.155072607
GSSPFG00006331001.1-RA	NGFI-A-binding protein	Ngf1a	5.296521038	2.729999007
GSSPFG00008932001-RA	intraflagellar transport protein 52 homolog isoform X2	p52	5.109329722	2.97790438
GSSPFG00022903001-RA	Peroxidase	peroxydase	5.70457573	-0.42053074
GSSPFG00020440001-RA	Polycalin1	polycalin	5.257370575	1.437914219
GSSPFG00035966001.2-RB	Polycalin1	polycalin1p3	7.373488645	1.753858172
GSSPFG00002897001-RA	putative inorganic phosphate cotransporter	Ptransporter	5.338207499	0.85946203
GSSPFG00014224001-RA	Rpb8	rbp8	5.334190603	1.498294088
GSSPFG00019426001-RA	phosphatidylinositol transfer protein (Sec14p)	Sec14P	5.052587086	0.438937448
GSSPFG00025955001-RA	Slack-LINE1	SlackLINE1	7.55220001	2.518786492
GSSPFG00025956001-RA	Slack-LINE2	SlackLINE2	8.408523346	3.316992707
GSSPFG00017532001-RA	putative cohesin	smc2	7.417473777	9.842471525
GSSPFG00004617001-RA	UGT33-11	UGT3311	8.168349428	4.136506264
GSSPFG00035441001.3-RA	UDP-glycosyltransferase-33-23	UGT3323	5.473349608	2.391054548
GSSPFG00031881001.1-RA	UDP-glycosyltransferase 33J2	UGT33J2	8.291597518	3.633597123

Table S3

OGS2.2	Abbreviation	Primer Orientation	Primer Seq	Primer Orientation	Primer Seq	Prod Size
GSSPFG00029721001-RA	501VNA	FORWARD	CCAAGGAACTGATGGATTGG	REVERSE	GGGATCATGACAGAGGACACA	56
GSSPFG00024881001-RA	adenylate cyclase	FORWARD	CACGGTGGACACACTACCAG	REVERSE	TCATAACCCCTCCCAGCATA	50
GSSPFG00010063001-RA	aminopeptidase	FORWARD	ACTGGACGCAATTTGAGGAG	REVERSE	GCTTCATCAGCTTCCAGAGG	54
GSSPFG00035209001.5-RA	carboxylesterase	FORWARD	TTGTGATACCTGGCGATGAA	REVERSE	GGGGGTGTAGACATTGAGGA	50
GSSPFG00004817001.2-RA	cohesin4817	FORWARD	CGGGTGTTCTCTGGAGAATTA	REVERSE	TCGACTGTGCATCATTGGAT	51
GSSPFG00031119001.2-RA	CYP	FORWARD	GGGGTTTGATCGCTCATCTA	REVERSE	CGTCAAATGGCTCTTTACCC	51
GSSPFG00017290001.2-RC	CYP340L1	FORWARD	TTAAACCGGAGCGATGGTTA	REVERSE	GCATTCCGGTTTTCTGGTAA	52
GSSPFG00002985001-RA	d245reductase	FORWARD	ATCATCGTGATGGTGGCTCT	REVERSE	CCAGATCTTCCAAACCAAGG	51
GSSPFG00029999001-RA	Def8	FORWARD	GTGCCAAACCGCATTAACTT	REVERSE	ATAATCGCGGTTTCATTCCAC	50
GSSPFG00031106001.2-RA	DUF4601	FORWARD	GTTTGGAATGTCGGGTTTTG	REVERSE	CTATCCGCGCTTCTTCTTTC	50
GSSPFG00002727001-RA	DVF1794	FORWARD	ATCAAACCTGGAACGAACGA	REVERSE	GCCCATGTTATGACTGACGA	51
GSSPFG00029716001.2-RA	FABP10	FORWARD	GTGTCCCCGATGACAAGATT	REVERSE	TCTGGTCTGGGGGTAGACTC	51
GSSPFG00034702001-RA	FABP12	FORWARD	GTGTCCCCGATGACAAGATT	REVERSE	TCTGGTCTGGGGGTAGACTC	51
GSSPFG00020720001-RA	FAR-X	FORWARD	CGGAGTACCGTATTCCTGA	REVERSE	TGAGCTGCTTCCCAAGAAAT	53
GSSPFG00018006001-RA	glyc synt	FORWARD	GCTCCGACATGACAGTGGTA	REVERSE	TATTCGTCTTGGCAGGGAAG	51
GSSPFG00024097001-RA	hemicentin2	FORWARD	TGTGGTGCTGAAGAACCTT	REVERSE	TGGGCCCATATTTCTATCA	50
GSSPFG00006331001.1-RA	Ngf1a	FORWARD	TTAATAACCCCGCCCTTTTC	REVERSE	CAGTTGGGCAGAGGTTAGGA	54
GSSPFG00008932001-RA	p52	FORWARD	ATCCAAAAGAATGCCACGTC	REVERSE	GGTGACGGCTCGGTTTAGTA	50
GSSPFG00022903001-RA	peroxydase	FORWARD	TAGCGCAATCTGGTGATGAG	REVERSE	GGTTGAGACGGACGGTTCTA	51
GSSPFG00020440001-RA	polycalin	FORWARD	GGGCCAAACGATTGTTTCTA	REVERSE	TATTGCCATGTCGGATCAAA	50
GSSPFG00035966001.2-RB	polycalin1p3	FORWARD	TGGTGGTGGCATCTCAGTAA	REVERSE	CGTTGCAAGTCTTTGGTTCA	55
GSSPFG00002897001-RA	Ptransporter	FORWARD	TCCAATTCTACTGAAGCCAGAG	REVERSE	TTACATCCTCAGCTCTTTCTACG	52
GSSPFG00014224001-RA	rbp8	FORWARD	AATGGCCGGTGTATTATTCG	REVERSE	CCGGGTCAATATCTTTCACG	53
GSSPFG00019426001-RA	Sec14P	FORWARD	ACCGCTGTTCCAAATTTTCAT	REVERSE	TCCTAACGTCAAAACAGCTGAA	51
GSSPFG00025955001-RA	SlackLINE1	FORWARD	GGAGAAGGGTGGCAAAAGAT	REVERSE	GGCCTCCTCTAACGACTTCC	50
GSSPFG00025956001-RA	SlackLINE2	FORWARD	CCCCAACAGAGAAAGATCCA	REVERSE	TTGTGCATAGAATGGCCTTG	50
GSSPFG00017532001-RA	smc2	FORWARD	CCATGGCCAATGGTATTAGG	REVERSE	CATCACCTGTTTCCTCGACA	53
GSSPFG00004617001-RA	UGT3311	FORWARD	GGTGTGCAAAAATGGGATT	REVERSE	CACGAGTCCAACCAAAACAA	57
GSSPFG00035441001.3-RA	UGT3323	FORWARD	CAGTTCCTTTGGTGGAGCTT	REVERSE	CTGAAGCGCCAATATTCTCA	50
GSSPFG00031881001.1-RA	UGT33J2	FORWARD	CTCTGGAAGTGGGACAAGGA	REVERSE	TCTGATGTTTCGCTGATTTGC	51