BRILL

Genome-wide analysis revisits incipient sympatric and allopatric speciation in a beetle

Wei Hong[a,*], Kexin Li[b,c,*], Kamal Sharaf[c], Xiaoying Song[b,c], Tomàš Pavlìcek[c], Huabin Zhao[a,d,**,***] and
Eviatar Nevo[c,**,***]

[a]*Department of Ecology, Tibetan Centre for Ecology and Conservation at WHU-TU, Hubei Key Laboratory of Cell Homeostasis,
College of Life Sciences, Wuhan University, Wuhan 430072, China*
[b]*State Key Laboratory of Grassland Agro-ecosystem, Institute of Innovation Ecology, Lanzhou University, Lanzhou 730000, China*
[c]*Institute of Evolution, University of Haifa, Haifa 3498838, Israel*
[d]*College of Science, Tibet University, Lhasa 850000, China*

Abstract The grain beetle, *Oryzaephilus surinamensis*, is a widespread species distributed in the wild and in granaries. Our earlier extensive biological studies indicated that the beetle shows *incipient* sympatric speciation (SS) in the wild at Evolution Canyon I (EC-I), Israel, and allopatric speciation, in a granary. Here we provide genome-wide evidence supporting our adaptive evolution scenario involving two models of speciation, SS in the wild, and allopatric in the granary. The EC-I microsite is a hot spot of SS across life from bacteria to mammals caused by the sharp opposite microclimates. The tropical hot, dry and savannoid biome dubbed the "African" slope (AS), sharply contrasts with the opposite temperate, cool, humid, and forested biome on the European" slope (ES), separated by only ~250 meters. The third allopatric granary population is 26 km north of EC-I. The granary population showed larger genomic, morphological, and behavioral distances, smaller genome size, more unique transposable elements, and reproductive isolation, displaying faster genomic divergence than between the wild populations at EC-I. The incipient SS of the wild populations, and the speciation of the granary population are reinforced by the substantial genomic divergence among the three beetle populations, supporting again the evolutionary scenario of incipient SS with gene flow at EC-I, and allopatric speciation in the granary population. We propose additional studies in Israel, the Mediterranean basin, and worldwide, to negate alternative explanations, based on a broader sampling and analysis.

Keywords ecological adaptation; genomic divergence; natural selection; population genetics; incipient sympatric and allopatric speciation

Introduction

The current genome-wide evidence of the saw-toothed grain beetle *Oryzaephilus surinamensis* follows and elaborates our earlier studies (Sharaf et al. 2008, 2010, 2013; Sharaf 2013). The earlier studies involved morphology, ecology, demography, genome size, nuclear and mitochondrial preliminary comparisons, including AT genomic proportions and cross-breeding between the three beetle populations. The three populations include two wild incipiently sympatric populations from Evolution Canyon-I, Mount Carmel, Israel, a hot spot of sympatric speciation (see Nevo 2014), and the third *allopatric* population from an isolated granary (silo, S) in the Haifa Bay. *Oryzaephilus surinamensis*, a cosmopolitan beetle species, is common in the wild in warm climates in the Meditarrenan, subtropical and tropical regions and abundant worldwide in granaries. It is associated with wild cereal Neolithic domestication. We compared and contrasted the beetle, first in the wild at Evolution Canyon-I (EC-I), Lower Nahal Oren, Mount Carmel, Israel. EC-I microsite is located in the Mediterranean Climate zone of Israel, south of Haifa

(see Supplementary Fig. S1), sharing its geology and *macro* climates, on the opposite canyon slopes. By contrast, its interslope microclimates are dramatically divergent, as is also true in additional three Evolution canyons we study in Israel, in the Galilee, Golan, and south Negev desert Mountains. These Evolution Canyons became an optimal model to study biodiversity evolution, including adaptive evolution, and incipient or full sympatric speciation across life (Nevo 2014). Since initiation of the Evolution Canyon model project in 1990, we identified in the four Evolution Canyons (EC-I-IV), 4000 species, and in EC-I at Mount Carmel, the most studied EC, 2500 species from soil bacteria to soil fungi, plants and animals up to mammals. We published on these long-term projects of the Evolution Canyons model 250 scientific papers listed under Nevo Evolution Canyons at http://evolution@haifa.ac.il. These interdisciplinary studies include many species, with ~15 major model organisms studied in depth both phenotypically and genotypically, with the major objective to assess their adaptation and incipient sympatric mode of speciation. These organisms include the cyanobacterium,

---

Nostoc linckia (Krugman et al. 2001; Satish et al. 2001; Dvornyk et al. 2002); soil bacterium, *Bacillus simplex* (Sikorski and Nevo 2005), wild barley, *Hordeum spontaneum* (Nevo 2006; Parnas 2006), fruit-fly, *Drosophila melanogaster* (Korol et al. 2006), grain beetle, *Oryzaephilus surinamensis* (Sharaf et al. 2008, 2010, 2013; Sharaf 2013), and spiny mouse, *Acomys cahirinus* (Hadid et al. 2014; Li et al. 2016), displaying adaptive incipient sympatric speciation at EC I. Recently, we added to this list two additional plant species, the crucifer *Ricotia lunaria*, in Hebrew Carmelit Naa, related to *Arabidopsis,* and wild emmer wheat*, Triticum dicoccoides,* the progenitor of cultivated wheat (Wang et al. 2020). EC- I microsite consists of the African slope (AS) tropical, hot, dry, and savanoid south facing slope (SFS), abutting and contrasting with the European slope (ES), temperate, cool, humid, and forested, north facing slope (NFS). The AS and ES are separated at bottom by 100 m, at top, 400 m, and at mid-slopes 250 m. Remarkably, they represent *locally,* due to sharp microclimatic divergence, global, continental, divergence (Nevo 1995; Pavlícek et al. 2003). Hence the EC model, is an hot-spot optimal natural model to highlight adaptive incipient or full sympatric speciation. The sympatric speciation model was first proposed by Darwin (1859), but was hotly debated as a rare model of speciation by many students of speciation, foremost by Ernst Mayr (1963), who changed his mind later, realizing its extensive occurrence in nature (Provine and Mayr 2004). Nevertheless, the sympatric speciation model was considered by many evolutionary biologists a rare model occurring only under very constrained conditions (e.g., Coyne and Orr 2004). Mounting studies (see Appendix 1) support sympatric speciation as a realistic model. This has been highlighted both theoretically (Gavrilets 2004), and experimentally (Appendix 1). Our own work in the microclimatic Evolution Canyon model (5), and its edaphic extensions in Evolution Plateau (EP) (Hadid et al. 2013; Li et al. 2015, 2016) and Evolution Slope (ES) in the Galilee Mountains, support the commonality in nature of the sympatric model of speciation. Microclimatic, microgeological, microedaphic, and microbiotic ecological divergences abound in nature, and provide optimal opportunities for the sympatric speciation model to generate new species (Nevo 2014).

Our earlier studies are discussed extensively with significant statistics in the PhD of Sharaf, 2013 available at the Haifa University library (Sharaf 2013). The results of these earlier studies suggested incipient sympatric speciation of the beetle *Oryzaephilus surinamensis* between the AS and ES populations at ECI, substantiated by postzygotic reproductive isolation, based on intraslope and interslope crossbreeding, between AS and ES wild populations at EC-I, and between both wild incipiently sympatric populations and granary (Silo, S) domesticated population that speciated allopatrically in granaries (nevo 2015).

In the present extended wide genome comparison between all three populations, two in the wild EC-I, and the third in the silo human granary, we provide new wide genomic evidence indicating substantial genomic divergence among the three beetle populations. The evidence includes SNPs, linkage disequilibria, LD, selected genes, and transposable elements, further demonstrating the genomic differences between the AS and ES populations at EC-I, and strongly selected between both wild EC- I populations, and the isolated, by a limited gene flow, silo population. We again support our earlier results and hypothesis on the incipient sympatric speciation of the beetles at EC-I, and the allopatric speciation in the isolated grain silo, based on genome-wide findings, including reproductive genes and mating behavior. However, to negate alternative explanations, we propose further studies of the beetles in the wild in Israel, Mediterranean basin, and worldwide, to obtain a broader comparative evolutionary perspectives.

Results

Genome assembly and population sequencing

O. surinamensis (Fig. 1a) is a cosmopolitan grain beetle (Supplementary Fig. S1). Eight grain beetles were collected from the "European", temperate, cool and humid, forested ES, also designated North Facing Slope, NFS (midslope population, station no. 6,) and eight grain beetles from the opposite "African", tropical, savannoid, AS, also designated South Facing Slope, SFS (midslope population, station no. 2) in Evolution Canyon I (EC-I) (Fig. 1a), and eight beetles from a grain silo (S) in Haifa Bay (Fig. 1a), at a distance of 26 km north of EC-I, respectively, in May 2014. Genomic DNA was isolated from the whole body using the Qiagen DNeasy kit. A reference genome was generated by the Illumina MiSeq platform with 300-bp paired-end sequencing. Whole genome resequencing data were produced by the Illumina HiSeq 2500 platform with 150-bp paired-end sequencing. After strict filtering, 41.4 Gbp clean data were retained for downstream analysis. The sequencing depth of the individu**als** used for genome assembly was 39.8-folds of the genome, and we obtained 374.7-folds of the genome for the three populations in total (Supplementary Tables S1-S2). The genome size of this grain beetle was estimated to be approximately 138 Mb by the K-mer frequency method (Supplementary Fig. S2; Chikhi and Medvedev 2013), and the assembly size was 104 Mb. The contig and scaffold N50 statistics were 19.6 Kb and 21.9 Kb (Supplementary Table S1), respectively. Genome annotation was conducted and 10,045 protein coding genes were identified.

Whole genome resequencing data were mapped to the reference genome; the mapping rate, coverage, and data size appear in Supplementary Table S2. We identified 1.63, 1.67 and 1.38 million single-nucleotide polymorphisms (SNPs) from the SFS, NFS, and S populations, respectively; and 331, 330, and 295 thousand indels were identified from the SFS, NFS, and S populations, respectively (Supplementary Fig. S3).

Genomic divergence analysis

Population divergences among the three populations (SFS, NFS, and S for granary) were explored using all identified SNPs. A neighbor-joining tree was constructed. We found
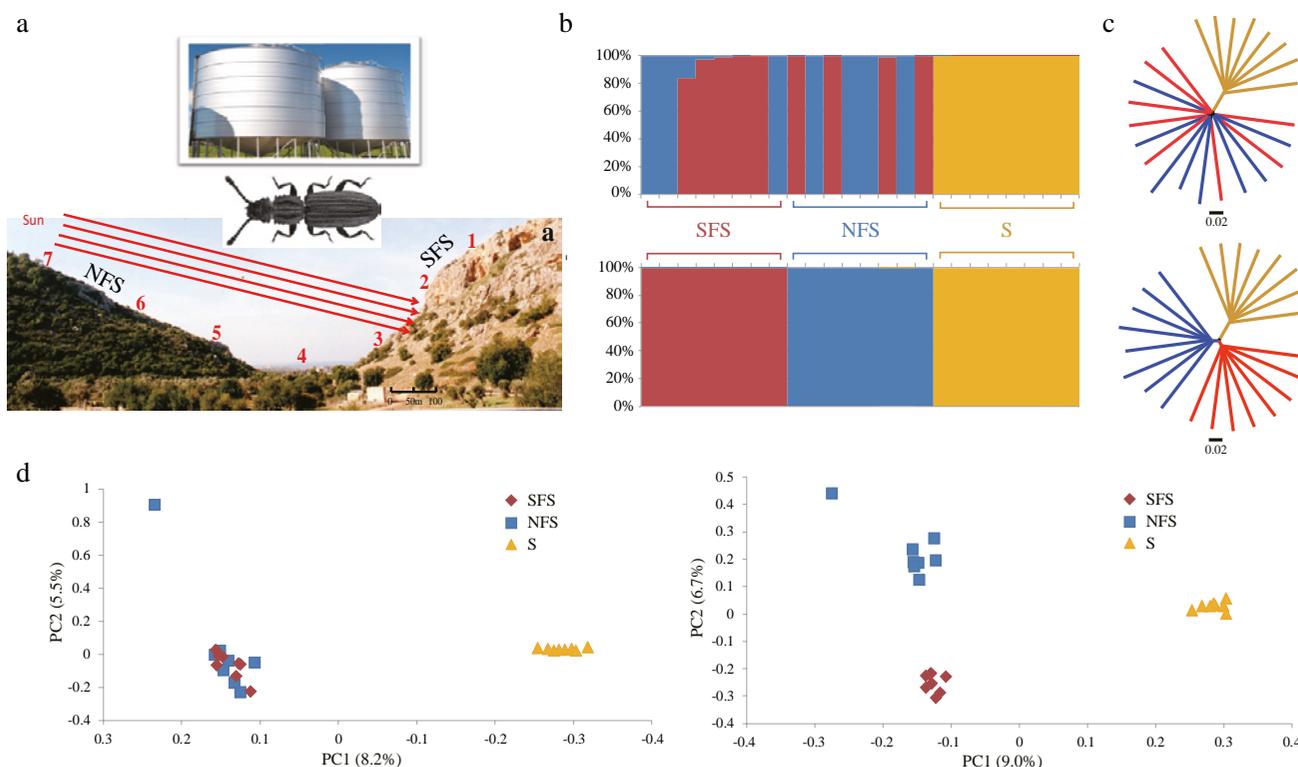
Figure 1.   Population divergence of SFS, NFS and S populations. (a) *Oryzaephilus surinamensis* and sampling sites. *Oryzaephilus surinamensis* (middle) samples were collected from a grain storage silo (top) and the north-facing slope (NFS) and south-facing slope (SFS) of "Evolution Canyon I" (bottom). (b) Genetic structure of the three populations at the whole genome level (top) and in the highly divergent genomic regions (HDGRs) (bottom). (c) Neighbor-joining tree of the SFS (red), NFS (blue), and S (yellow) populations, constructed with all SNPs (top) and the SNPs in HDGRs (bottom). (d) Principal component analysis (PCA) of the three populations, based on all SNPs (left) and the SNPs in HDGRs (right).

that all individuals from the S population were clustered together but distantly related from the populations of EC-I (SFS and NFS), while the individuals from SFS and NFS populations cannot be separated at the whole genome level (Fig. 1c). However, when we chose the SNPs from the highly divergent genomic regions (HDGRs, which were defined as the genomic regions with the highest 5% of F_{ST} values between SFS and NFS), the AS = SFS and ES = NFS populations could be clustered genetically into two populations (Fig. 1c), which is consistent with their ecological origin divergent biomes. A principal component analysis (PCA) was performed on all individuals using two alternative SNP datasets, one with all the SNPs and the other with SNPs in the HDGRs (Fig. 1d). The PCA results on all SNPs indicate that the S (granary) population is distinct from the wild ECI populations, while the genome divergence between the AS = SFS and ES = NFS populations is not complete (Fig. 1d). The PCA results demonstrated that the AS = SFS and ES = NFS populations could be separated using the SNPs from the HDGRs (Fig. 1d). Population genetic structure analysis showed similar results as shown in both NJ tree and PCA (Fig. 1b).

Genetic diversity, linkage disequilibrium, and recombination rate

Genetic diversity was estimated by mean theta (θ): 4.619×10^{-3}, 4.701×10^{-3}, 3.924×10^{-3}, and 6.155×10^{-3} for SFS, NFS, S and the combined wild population (SFS+NFS),

respectively. The genetic diversity for the wild population is significantly higher than the S population (P-value $< 2.2 \times 10^{-16}$, Mann-Whitney U test); the NFS population displays significantly (P-value $= 1.622 \times 10^{-15}$, Mann-Whitney U test) higher genetic diversity than the SFS population, and both SFS (P-value $= 2.2 \times 10^{-16}$, Mann-Whitney U Test) and NFS (P-value $= 2.2 \times 10^{-16}$, Mann-Whitney U Test) populations show significantly higher genetic diersity than the S population. The order of genetic diversity is as follows: wild>NFS>SFS>S.

Linkage disequilibrium (LD), measured by calculating the square value of the correlation coefficient (r^2), decreased rapidly to its bottom within 5,000 bp in all populations of the beetle but different populations show variations in LD decay. The domesticated S population (in orange) shows the highest LD (Fig. 2a) and dropped to its lowest value in the longest physical distance; while the population from the SFS (in red) and NFS (in blue) display nearly the same LD. The wild populations (in green) exhibit lower LD than each of the three populations, and all the combined populations (SFS+NFS+S, in black) has the lowest LD. The mean recombination rate for each individual was calculated and listed in Supplementary Table S3, and that for the SFS, NFS, and S populations was 1.95, 2.34, and 1.31 per kilo bases, respectively. It is significantly higher in the NFS population than in the SFS ($P = 0.039$, T-test) and S populations ($P = 0.001$, T-test). The mean recombination rate of the SFS population is significantly ($P = 8.62e-5$, T-test) higher than in the S granary population.
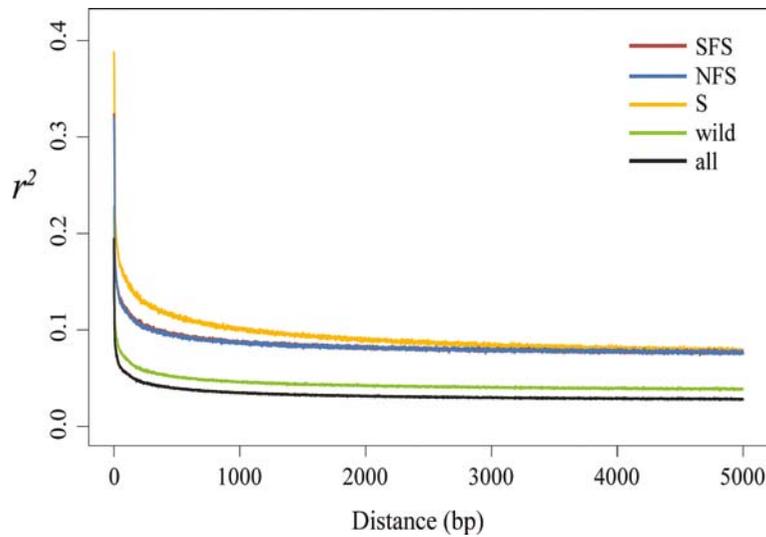
Figure 2.   Linkage disequilibrium. X-axis denotes physical distance and Y-axis stands for wild = SFS+NFS+S.

Natural selection in putatively selected regions, and functional enrichment

Putatively selected regions (PSRs) of each of the three populations were identified by using combined highest 5% F_{ST} and lowest 5% Tajima's D with a sliding-window approach. In all the populations, PSRs showed positive and significantly higher spatial autocorrelation level than random situation (SFS: $P < 2.2\times10^{-16}$; NFS: $P < 2.2\times10^{-16}$; S: $P < 2.2\times10^{-16}$; wild: $P < 2.2\times10^{-16}$, Z-test). In the PSRs, we identified 64 and 70 genes from SFS and NFS populations,

respectively; when the S, granary population was compared to the wild population, 196 and 101 genes were identified as the putatively selected genes (PSGs) (Supplementary Table S4). Gene ontology (GO) analysis demonstrates that compared to the SFS population, the NFS population was characterized by body musculature, development and neurogenesis (Fig. 3). Compared to the wild populations, the PSGs of the S granary population were enriched in neurogenesis, metabolism, chemical sensory, detoxification, reproduction and damage repair (Fig. 4).
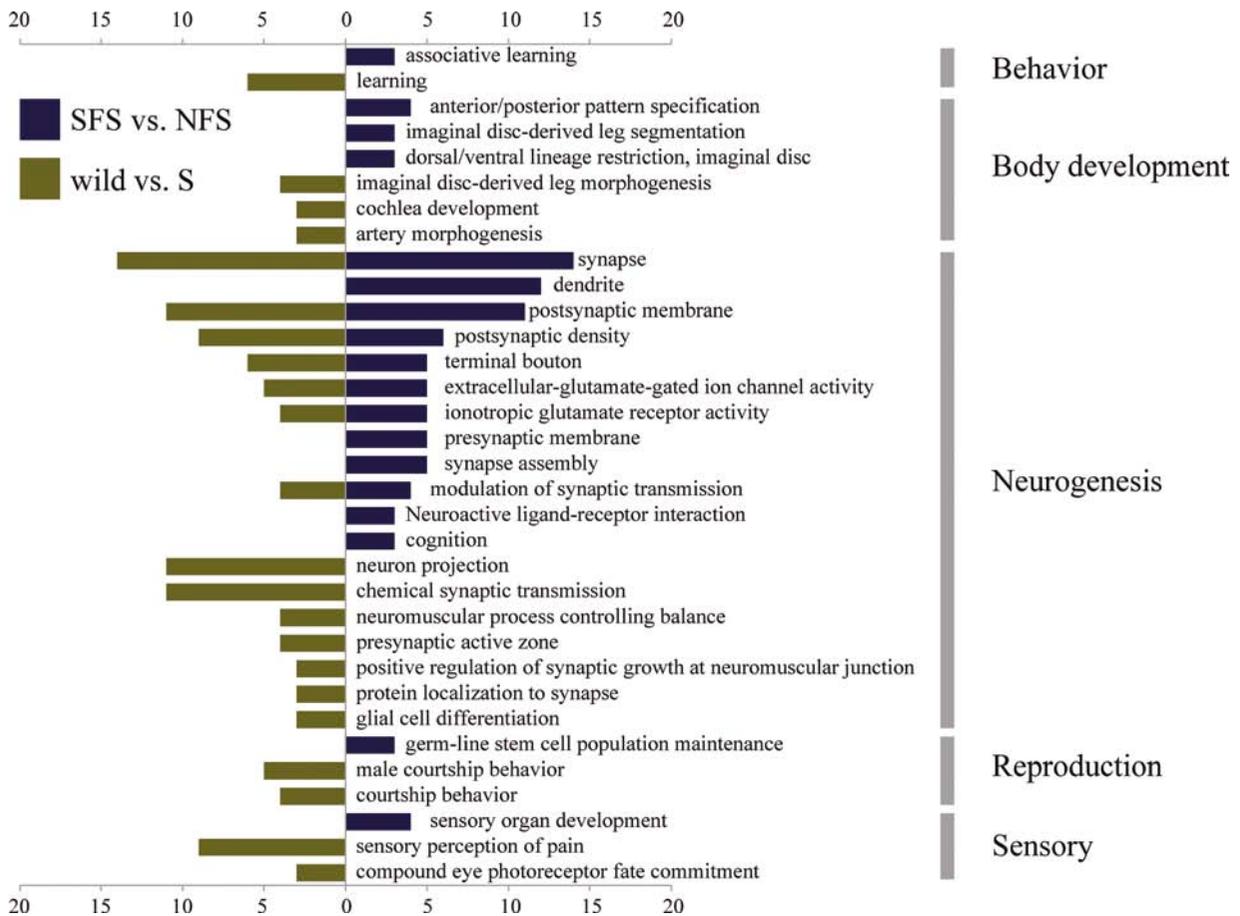

Figure 3.   Functional enrichment analysis of putatively selected genes (PSGs) in SFS and NFS populations.
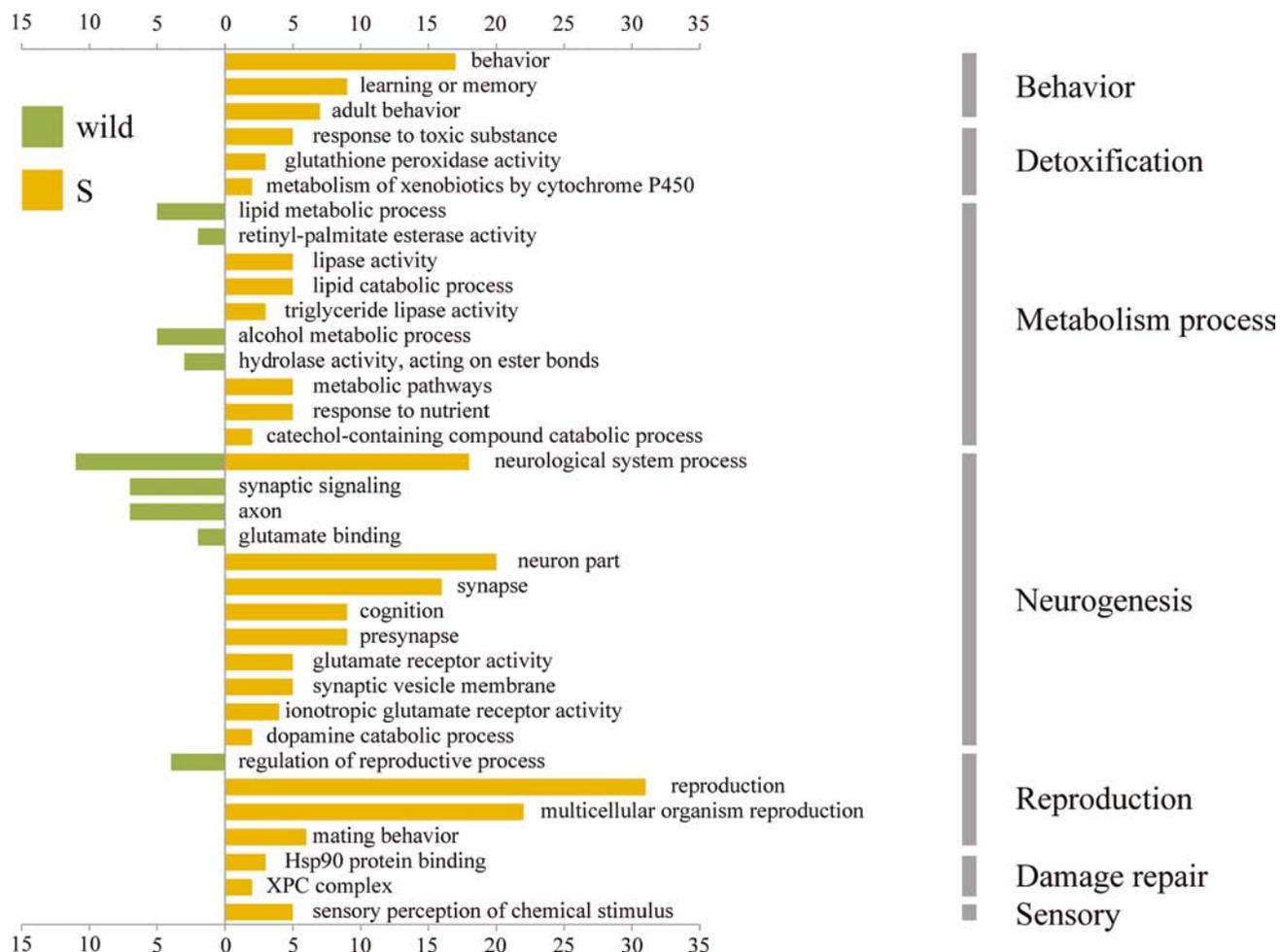
Figure 4. Functional enrichment analysis of putatively selected genes (PSGs) in the combined wild (SFS+NFS) and S populations. X-axis denotes number of genes.

Transposable elements divergence

Transposable elements (TE) are comprised of two classes on the basis of the transposition mechanisms: one is DNA-based transposons and the other is RNA-based retrotransposons (27). Transposons could move to new sites directly by a cut-and-paste mechanism, while retrotransposons are transferred to a new place in the genome through an RNA intermediate (28). All of the transposable elements, including DNA transposons, short-interspersed nuclear elements (SINEs), long-interspersed nuclear elements (LINEs), long terminal repeats (LTRs), tandem repeats, and some unknown TEs were identified from each population (Supplementary Table S5) and the population-unique TEs were listed in Fig. 5. There are 46,606, 46,692, 47,265 and 46,802 TEs harboring in SFS, NFS, S, and the combined wild populations (NFS+SFS), respectively, and 110, 196, 769, and 271 TEs are unique to SFS, NFS, S, and the combined wild populations, respectively (Fig. 5). The S granary population harbors the most unique TEs of all different kinds, followed by the wild populations, NFS, and SFS populations (Fig. 5). The number of LINEs and LTRs unique to each population is small (Fig. 5). Notably, there are no unique SINEs to any of the populations (Fig. 5). TE-related genes, in other words, genes with TEs located in their promoters, introns, or exons were listed in Supplementary Table S5.

Functional enrichment was performed on the genes related to the TEs differentiated between the SFS and NFS, and between the wild and S populations (Supplementary Fig. S4). For SFS and NFS population pair, the difference mainly exists in neurogenesis and body development, while in the wild and S granary population pairs the difference was in neurogenesis, body development, sensory and reproduction (Supplementary Fig. S4).

Discussion

The important background results (Sharaf et al. 2008, 2010, 2013; Sharaf 2013; Wu and Ting 2004)

Earlier results (Sharaf et al. 2008, 2010, 2013), detailed and analyzed in Sharaf's PhD thesis (Sharaf 2013), are of cardinal importance and contribute much biological background to the wide-genome results of the present study. Based on ecology (natural habitats and niches), morphology (body size and occurrence of horns), significant inter-population genome size, comparisons of genome size in other organisms at EC-I, larval differential survivorship in the lab common garden experiment, genetic differentiation of AFLP and mtDNA, and, most importantly cross-breeding between all three populations. The following general conclusions have been drawn. Sharp adaptive ecologic, morphologic, and behavioral divergences have been identified
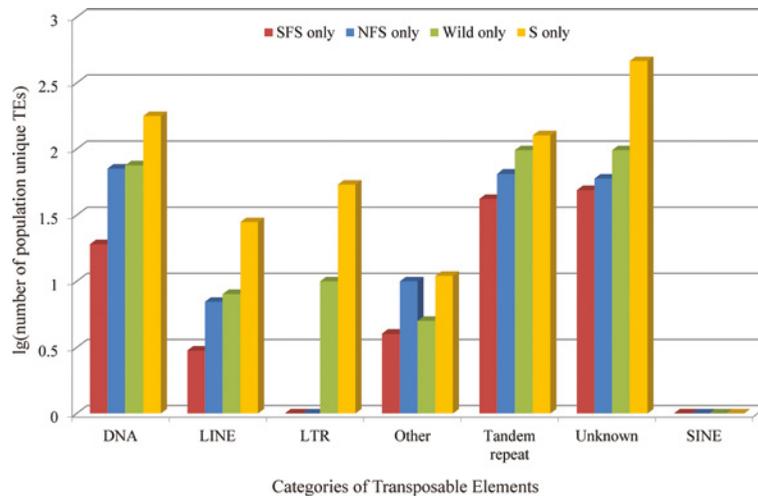
Figure 5.   Distribution of transposable elements unique to each population.

statistically between the wild AS and ES interslope populations of EC-I, and between each of them and the isolated grain silo, S population. These statistically significant differences justify to hypothesize that the AS and ES populations from EC I are adaptively divergent evolutionarily, and undergo early stages of incipient sympatric speciation, with postzygotic reproductive isolation. These early results and hypotheses, have been complemented substantially in the current genome-wide analysis as discussed below, with new important evidence.

Whole genome sequencing and population divergence

The reference genome of *O. surinamensis* was estimated to be 138 Mbp, which is comparable to our previously reported 151.5-154 Mbp (Sharaf et al. 2013) estimated by flow cytometry. The *de novo* assembly size of 104 Mbp is smaller than the estimated genome size, with a relatively shorter scaffold N50 (21.9 Kb). This may be due to the relatively high heterozygosity (Supplementary Fig. S2), low sequencing depth, and the lack of additional sequencing libraries with different insert sizes. We did not construct multiple sequencing libraries due to the small size of the beetle and the high heterozygosity of the genome (Supplementary Fig. S2). Nevertheless, about 95% of ultra-conserved core eukaryotic genes were found to be complete in the assembly (Supplemenatry Table S1), suggesting that the reference genome is of good quality. In addition, we used conservative and stringent strategy to annotate the genome, which resulted in 10,045 gene models. Our predicted gene number was smaller than that of the two known beetle genomes (13,088 in *Dendroctonus ponderosae* (Keeling et al. 2013) and 16,404 in *Tribolium castaneum* (Richards et al. 2008), but the gene numbers in the three beetle genomes are still comparable. As a result, we believe our draft genome of the beetle is adequate for subsequent analyses.

Neighbor-joining tree, PCA, and population structuring analyses demonstrate that the individuals from the grain silo, S population were significantly divergent from the ES = NFS and AS = SFS and clustered as a distinct population (Fig. 1). This is congruent with a newly and

totally divergent ecological niche in the grain silo, S. The S granary population feeds on cultivated cereals, while the NFS and SFS populations feed primarily on live-oak acorns and other food resources present in the field. We can expect that difference in food leads to a drastic population differentiation (e.g. in body size and genome size) and genome divergence (Vanbergen et al. 2003). The AS = SFS and ES = NFS populations were not divergent at the whole genome scale, which was indicated by several individuals' mixture (Fig. 1b-d). This might be due to the interslope migration of the beetles, which keeps homogenizing the genetic background of the two populations. However, the AS = SFS and ES = NFS populations could be separated by all of the three methods: NJ tree, PCA, and population structuring when we used the SNPs in highly divergent regions (Fig. 1b-d) across the genome. Further, we randomly separated SFS and NFS samples into two equal groups (8 samples of each group) 1000 times, calculated the F_{ST} and Tajima's D of selected regions in SFS and NFS, and performed a permutation test. The result suggested that these regions have significantly higher F_{ST} and lower Tajima's D than random background, indicating that these regions may be truly under selection between SFS and NFS rather than displaying false positive caused by random sampling error (Fig. 6). Moreover, the enriched GO function of these divergent genes may have resulted from their unique, slope-specific, ecological-genomics functional divergence due to the environmental stresses temperate, cool-humid, forested in the European-like ES = NFS as compared to the tropical, hot-dry, savannoid African-like AS = SFS (Fig. 3). Some loci may have divergently adapted to the interslope contrasting microclimates ecologies, and biomes of Evolution Canyon-I (e.g. the AS = SFS and ES = NFS of EC I), however, gene flow continued to partly homogenize the retained parts of the genome (Wu and Ting 2004). This full genome result is consistent with the earlier findings of our previous studies listed above, which showed that the divergence between AS = SFS and ES = NFS has not been completed, but that these two populations were genetically distinct from the S granary population (Sharaf et al. 2013). The interslope divergence of AS and ES
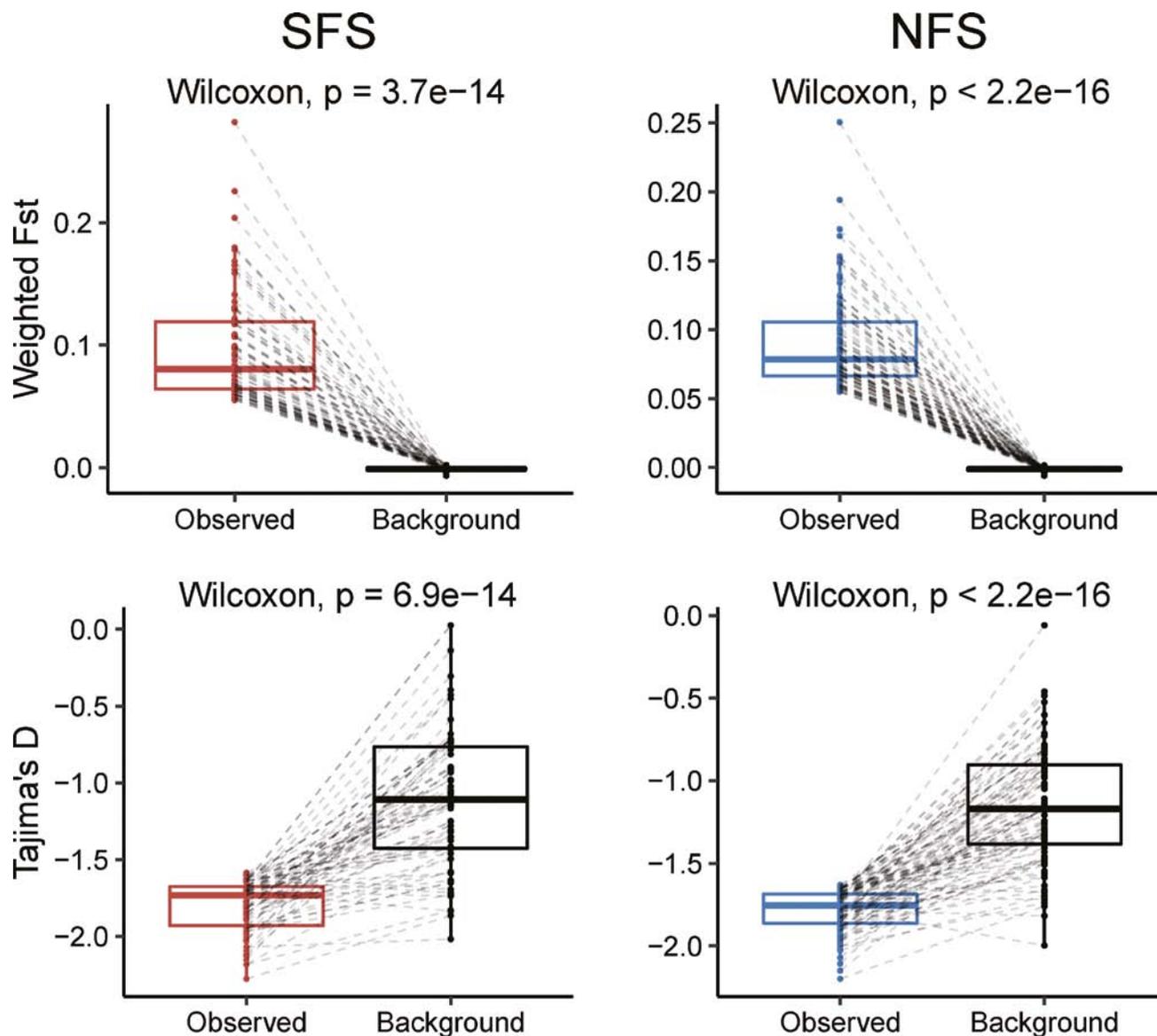
Figure 6.   Permutation test of selected regions between SFS and NFS. Samples from SFS and NFS were randomly separated into two groups equally 1000 times. The "Observed" box showed observed Fst and Tajima's D values from selected region in SFS and NFS, while the "Background" box showed the mean value of 1000 bootstraps in corresponding regions. The significance was tested by paired Wilcox test.

populations is adaptively caused by their sharply divergent microclimates, although the similar acorn oak fruits consumed by ES and AS populations may diminish the divergence.

Genetic diversity and linkage disequilibrium

Genetic diversity was lower in the domesticated grain silo population than that in the wild, either AS or ES populations, suggesting that the genetic diversity was reduced during the process of domestication and the relatively homogenized and stable environment in the grain silo (Qi et al. 2013; Zhou et al. 2015). The genomic diversity is significantly higher in the ES=NFS population than in the AS = SFS population, suggesting higher environmental stresses for the ES population due to the cool and humid environment, ample with viral, bacterial and fungal pathogens (Nevo 2006). This result is different from our previous study based on AFLP molecular markers, in which the

AS = SFS population harbored the highest genetic diversity (Sharaf et al. 2013). The discrepancy between the earlier and current study in the estimates of genetic diversity may derive from several reasons, individually or in concert. First, the number of molecular markers in the earlier study was not sufficient. Second, the sample size in our study is not enough. The higher genetic diversity in ES = NFS than in AS = SFS population is possibly caused by two reasons: First, the population size on the AS = SFS is very limited (2), due to the limited number of live-oak trees in the savannoid AS = SFS (Nevo et al. 1999), which may result in low genetic diversity. Second, as *O. surinamensis* is a warm-adapted species, when **it** moved to a new, cool, and humid NFS environment, especially under 20°C, the development of the species was restricted (Halstead 1980), and the novel strong stresses selected for higher genetic diversity (Nevo 1998). Remarkably, this is true for the ES = NFS slope, which is distinct ecologically from the low humidity of the grain silo population (Figs. 3, 4). Similar

higher genetic diversity on the ES occurs in *Drosophila melanogaster* which is also incipiently speciating sympatrically at Evolution Canyon I (Korol et al. 2006; Hübner et al. 2013; Kim et al. 2014).

Multiple factors could contribute to the mosaic genome divergence during the processes of population differentiation, such as recombination rate (Nachman and Payseur 2012), diversifying selection (Li et al. 2015, 2016), and genetic drift (Ohta 1992). In the present study, LD was higher in the domesticated silo population than in each of the AS and ES populations, the combined wild populations, and all combined populations (Fig. 2a). The same pattern was observed in other organisms, where domestication could increase the LD (Qi et al. 2013). In addition to the possible bottleneck during domestication, beetles in the silo were also subjected to different food resources, stable but very different microclimate than in the wild at EC I, change of daily and seasonal photoperiod, and circadian rhythm, and the lack of water; all of these environmental changes increasing the stress and, consequently, selected for increased LD. The LD is lower in both the combined wild populations and all combined populations, suggesting that the LDs are environmentally specifically selected and decreased when populations were combined (Li et al. 2015). The recombination rate in the three populations was different possibly due to different selective regimes (Otto and Michalakis 1998) imposed by the three contrasting environmental stresses. The recombination rate and genetic diversity show the same trend as follows: NFS > SFS > S.

Natural selection and functional enrichment

The difference of selection pressure between populations may lead to high F_{ST} values (Holsinger and Weir 2009); likewise, negative Tajima's D may result from positive selection (Tajima 1989). However, genetic drift may also cause changes of these statistics. Due to the selective sweep, selected regions tend to cluster in linkage blocks (48), while genetic drift appears randomly in the genome (Lynch et al. 2011). In all the populations, PSRs show positive and significantly higher spatial autocorrelation level than random pattern, indicating that the PSRs were more likely under selection rather than genetic drift. The number of PSGs was significantly less than random value in each population (SFS: $P < 2.2\times10^{-16}$; NFS: $P < 2.2\times10^{-16}$; S: $P < 2.2\times10^{-16}$; wild: $P < 2.2\times10^{-16}$, Z-test); probably due to the clustering, PSRs tend to cover the same gene rather than several genes. PSGs from the ES = NFS population were enriched in musculature possibly because there were abundant foods on the forested north-facing slope (NFS), covered with a mixed forest of *Quercus calliprinos*, whose acorns are the major food of the beetle, and *Pistacia palaestina*, and therefore animals could acquire enough nutrition. Likewise, the body development is enriched in the ES = NFS population probably due to fast growth caused by abundant food resources, primarily lots of live-oaks acorns, *Quercus calliprinos*, on the north-facing slope. Indeed, the mean body weight of the NFS population was significantly greater than that of the AS=SFS population,

in both genders, following the Bergman rule not only in mammals but also in an insect (Sharaf 2013). The enriched function of neurogenesis in the NFS population might suggest that when the beetles moved to the ES = NFS environment, they evolved a nervous system selected by the stressful microclimatic environment, which better adapted them to the cool and humid microclimate (Zimmermann et al. 2011). However, the higher food resources on the ES = NFS compared to the AS = SFS evolved a much higher population density (Sharaf et al. 2008; Sharaf 2013). Beetles found in the grain barn also evolved a nervous system adapting them to the unique granary environment, with 22-26°C, and 14% relative humidity. When the beetles migrated to the lower, but stable temperature in the silo, the nervous system may also have helped them adapting to the cooler environment (Zimmermann et al. 2011). That trend, but in the opposite direction, was also identified in blind mole rats, *Spalax judaei* in Israel, when colonizing the dry-hot desert, the brain size increased significantly (Nevo et al. 1988).

Allopatric speciation in the Grain Silo versus incipient sympatric speciation at Evolution Canyon

The allopatric speciation mode in the grain- silo contrasts with the incipient sympatric speciation mode in the hot spot of sympatric speciation of Evolution Canyon. As food resources became abundant in the grain silo the metabolism could be strengthened. Changes of the chemical sensory and detoxification helped the beetles to shift the food resource from live-oak acorns to specific grains. Importantly, reproduction related genes, which mainly involved in gametogenesis and courtship behavior (Supplementary Table S4), were found under selection in the grain silo population, associated with postzygotic and prezygotic reproductive isolation, supporting the hypothesis reached earlier (Sharaf 2013), that the speciation mode in the silo was allopatric in contrast to the incipient sympatric speciation at Evolution Canyon-I (Nevo 2006, 2014). Another hot spot of Sympatric speciation was found in blind subterranean, mole rats, of the *Spalax ehrenbergi* superspecies at Evolution Plateau in the Upper Galilee (Li et al. 2015, 2016). These genomic results reinforce the cros-breeding results (Sharaf et al. 2013) which revealed offspring inferiority in interslope contrasting intraslope crosses at Evolution canyon (Sharaf et al. 2013). This was consistent with the grain Silo population significantly higher fecundity than that of wild populations. The grain Silo population could also develop prezygotic reproductive isolation based on chemical sensory , that needs future testing, beside the postzygotic reproductive isolation as indicated by crossing (Sharaf 2013). Likewise, the significantly higher number of unique TEs in the grain Silo's population (Supplementary Table S5, and Fig. 5, and later discussion), might have contributed to a new Neolithic *allopatric* species in human granaries, represented by the grain silo population. Strengthened detoxification, damage repair, and nervous system may also help the barn-living beetles tolerate periodic treatments with pesticides (Zettler and Arthur 2000).

Transposable elements

Transposable elements (TEs) reportedly play a pivotal role in adaptation (Casacuberta and González 2013; Schrader et al. 2014), genome evolution (Fedoroff 2013), generation of adaptive phenotypes (Casacuberta and González 2013) through transposition, retro-transposition, aberrant recombination, and ectopic transposition (Hua-Van 2011). TEs may affect gene function, generate new genes (Nekrutenko and Li 2001), and provide genetic materials and regulatory elements, which could generate stress-inducible regulatory complexes (Casacuberta and González 2013). Many TEs were reported to be taxon-specific, and their presence in some coding genes will accelerate species divergence, and speciation (Nekrutenko and Li 2001). In the present study, population-specific TEs were identified in all three AS, ES, and grain silo populations. This may have facilitated beetle adaptation and speciation, especially in the grain-silo, which had the highest unique TE abundance, but also facilitated incipient sympatric speciation in EC-I. The largest number of TEs displayed in the grain-silo population (Fig. 5) might have helped the population promote genetic differentiation and provided a potential way for rapid species divergence (Nekrutenko and Li 2001), and the TEs also may have helped the beetles to respond to their changing stressful environments (Casacuberta and González 2013). As *O. surinamensis* is a warm-adapted beetle, the bigger number of TEs in ES = NFS than in AS = SFS populations may have also facilitated population divergence and adaptation to a new cooler NFS environment (Grandaubert et al. 2014), and promote incipient sympatric speciation, as has been demonstrated at EC-I across life from bacteria to mammals (Nevo 2014). GO enrichment performed on TE-related genes demonstrated that the difference between the AS = SFS and ES = NFS was in neurogenesis and body development. This was possibly due to the adaptation to the new and climatically more stressful environment with a richer food supply for the AS population moving to the ES = NFS biome, with lower temperature and particularly high relative humidity. Compared to the wild population, the developed nervous system in the grain silo population may also have facilitated the adaptation to a cooler and drier, hence more stressful ecological niche, generating since Neolithic times, 10,000 years ago, a new allopatric species.

Conclusions and prospects

Adaptation and speciation are twin processes of evolution, and the first signal of both is genetic divergence. Here we describe genomic divergence within the cosmopolitan beetle *Oryzaephilus surinamensis* in the wild at Evolution Canyon and under human domestication, in the granary silo. We provided genomic evidence supporting genetic divergence in the three populations: two wild populations at EC-I, 250 m apart, probably undergoing adaptive incipient sympatric speciation as suggested in our earlier studies (Sharaf et al. 2013), and was shown also at Evolution Canyon-I for several species across life (Nevo 2014). These include soil bacteria *Bacilus simplex* (Sikorski and Nevo 2005), plants, wild barley (Nevo 2006; Parnas 2006),

Dropsophila melanogaster (Korol et al. 2006), and spiny mice (Hadid et al. 2014; Li et al. 2016). Two additional species that underwent sympatric speciation, demonstrated recently, were the crucifer *Ricotia lunaria* (Qian et al. 2018), and wild emmer wheat, *Triticum dicoccoides* (Wang et al. 2020). Likewise, subterranean blind mole rats, *Spalax galili* speciated sympatrically at Evolution Plateau, eastern Upper Galilee, Israel (Hadid et al. 2013; Li et al. 2015, 2016). Both Evolution Canyon, and Evolution Plateau, are two hot spots of incipient or full sympatric speciation, out of six currently studied Evolution model microsites in Israel. While both wild populations of *O. surinamensis* are at EC-I, Mount Carmel, Israel, the third population is located 26 km north of the EC-I wild populations, isolated in a silo in the Haifa Bay. The EC-I wild populations of grain beetles are separated by 250 meters, on the two abutting slopes of the EC-I: the tropical, savannoid, hot, and dry mid south-facing slope population (AS = SFS) versus the temperate cool and humid, forested mid north-facing slope population (ES = NFS). Genomically, the three populations display the following quantitative order: Genetic diversity: NFS > SFS > S; linkage disequilibrium (LD): S > SFS = NFS; recombination rate: NFS > SFS > S; number of putatively selected genes: NFS > SFS; when the granary S population was compared to the wild population, 196 and 101 genes were identified, respectively, as the putatively selected genes (PSGs) (Supplementary Table S4). Transposable elements: S > NFS > SFS; population-unique TEs: S > NFS > SFS. Moreover, crossing experiments between the three populations, AS, ES and grain silo, identified inferiority in fertility, in the EC-I interslope as compared to the intraslope crosses. Thus, postzygotic reproductive isolation, support the two speciation scenarios, *incipient* sympatric at EC I, and allopatric in the granary silo (Sharaf 2013, and Appendix 2). The beetle *Oryzaephilus surinamensis* is largely a warm adapted species (Supplementary Fig. S1). Both the north-facing slope (cool and humid) and the granary silo (cool, dry, and microclimatically stable) are novel and potentially stressful environments for the beetle. We interpret the above genetic traits' order as representing adaptations to the more stressful novel environments to the beetle ES = NFS (highest genetic diversity, recombination rate, and putatively selected genes in ES = NFS at ECI), and the granary-silo (highest LD, and TEs), associated with invasion to new stressful environments). The silo population apparently diverged from the two wild populations at EC-I, following the cultivation of wild cereals, ~10,000 years ago. In summary, the beetle *O. surinamensis* represents two ecological adaptive speciation models of divergence: *incipient sympatric speciation at the EC I hot spot of sympatric speciation,* together with other six species across life, from bacteria to mammals, at Evolution Canyon-I, and the allopatric speciation, following wild cereal Neolithic domestication in the granary silo.

Although *O. surinamensis* is a worldwide storage pest, its origin remains unknown. The beetle has been found in food storage at Greece since ancient times (~6,400 years ago) (Valamoti and Buckland 1995), implying the hypothesis that the beetle originated and started to appear in human granaries and later in silos in the east Mediterranean area,

following the emergence of old human agriculture in the Near East Fertile Crescent. The beetle may have lived in the preagricultural wild-collected seeds in human huts for a long time before the domestication of cereals. However, global human dispersal and commerce play a vital role in the distribution and population structure of storage pests. Thus, to investigate the relationship among the three populations in our previous (Sharaf et al. 2013; Sharaf 2013) and current work , the cosmopolitan grain silo populations and the other wild populations in Israel, Mediterranean basin, and elsewhere in the world, also need to be sampled to understand the evolutionary processes of this beetle. Plenty of genes and gene networks coping with local adaptation have been identified in different populations, but their detailed mechanisms are still unclear. Post-zygotic reproductive isolation exists between the three grain beetle populations at different levels (Sharaf 2013). Reproduction genes involved in gametogenesis and courtship behavior (Supplementary Table S4) were found under selection in the recent granary silo population which could be involved in reproductive isolation, hence the granary population may be regarded as a good candidate of a new allopatric species. Since the mill populations of *Tribolium castaneum* are highly divergent (Semeao et al. 2012), *O. surinamensis* living in the same habitat may show similar genetic divergence. Thus, further analysis based on more samples from different granary populations, in Israel, the Mediterranean basin, and worldwide, may unfold their evolutionary relationship with the wild populations. Future behavioral and genetic mating experiments could decide the stage of incipient sympatric speciation of the interslope wild populations at EC I. Our current genomic analysis reinforces our earlier analyses (Sharaf et al. 2008, 2010, 2013; Sharaf 2013).

Our data are consistent with the scenario of incipient sympatric and allopatric speciation models, in the wild EC I, and silo populations, respectively. However, more populations need to be sampled to eliminate various alternatives. A deeper bioinformatic analysis could be conducted on the noncoding regulatory elements, as unfolded in human ENCODE analysis. Likewise, an extensive epigenetic methylation analysis could highlight the regulation of genome dynamics. Our analysis including earlier studies (Sharaf et al. 2013; Sharaf 2013), and the current wide genome analysis, provide a solid starting point basis for future extended analyses to highlight the evolutionary dynamics of *Oryzaephilus surinamensis* in Israel, Mediterranean region, and worldwide. Finally, as first hypothesized by Darwin,we predict that sympatric speciation, either incipient or full, is a common speciation model across life, and across the planet, since Evolution canyons at microsites due to geologic, edaphic, climatic, biotic, and abiotic contrasts abound on our planet (Nevo 1995, 2014).

Materials and methods

All experiments on the beetles were approved by the Ethics Committee of Wuhan University, University of Haifa and the Institute of Agricultural Research, Chinese Academy of Agricultural Sciences, and conformed to the rules and guidelines on animal experimentation in China and Israel. Whole genomes of the beetle, *Oryzaephilus surinamensis,* from the EC-I and the silo in Israel were sequenced on Illumina HiSeq 2500. Neighbor-joining tree, PCA, and population structuring analyses were performed based on the SNPs identified by screening genomic regions that show low genetic diversity (measured by Tajima's D) in one population but high divergence (measured by F_{ST}) between two of the three populations. Genetic diversity, linkage disequilibrium, and recombination rate of each population were calculated from the same SNP dataset. PSGs of each population were identified by two combined parameters of Tajima's D and F_{ST}. Functional enrichment analysis was performed on the genes under selection. Transposable elements of each population were isolated, and genes with TEs were assessed by functional enrichment analysis. Full details of the materials and methods were described in Supplementary Text S1.

Supplementary material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.12721472

References

Barrón MG, Fiston-Lavier A-S, Petrov DA, González J. Population genomics of transposable elements in *Drosophila*. Annu Rev Genet. 2014; 48: 561–581.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol. 2013; 22: 1503–1517.

Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2013; 30: 31–37.

Coyne JA, Orr HA. Speciation. Sunderland, MA. Sinauer Associates, Inc, 2004.

Darwin C. On the origin of species. John Murray, London. 1859.

Dvornyk V, Vinogradova O, Nevo E. Long-term microclimatic stress causes rapid adaptive radiation of kaiABC clock gene family in a cyanobacterium, *Nostoc linckia*, from "Evolution Canyons" I and II, Israel. Proc Natl Acad Sci USA. 2002; 99: 2082–2087.

Fedoroff NV. Transposable elements, epigenetics, and genome evolution. Science. 2012; 338: 758–767.

Fedoroff NV. Plant transposons and genome dynamics in evolution. John Wiley & Sons, 2013.

Gavrilets S. Fitness landscapes and the origin of species (MPB-41). Princeton University Press, 2004.

Grandaubert J, et al. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. BMC Genomics. 2014; 15: 891.

Hadid Y, Pavlíček T, Beiles A, Ianovici R, Raz S, Nevo E. Sympatric incipient speciation of spiny mice *Acomys* at "Evolution Canyon", Israel. Proc Natl Acad Sci USA. 2014; 111: 1043–1048.

Hadid Y, et al. Possible incipient sympatric ecological speciation in blind mole rats (*Spalax*). Proc Natl Acad Sci USA. 2013; 110: 2587–2592.

Halstead D. A revision of the genus *Oryzaephilus* Ganglbauer, including descriptions of related genera (Coleoptera: Silvanidae). Zool J Linn Soc. 1980; 69: 271–374.

Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat Rev Genet. 2009; 10: 639.

Hübner S, et al. Genome differentiation of *Drosophila melanogaster* from a microclimate contrast in "Evolution Canyon", Israel. Proc Natl Acad Sci USA. 2013; 110: 21059–21064.

Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. The struggle for life of the genome's selfish architects. Biology Direct. 2011; 6: 19.

Kim YB, et al. Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. Proc Natl Acad Sci USA. 2014; 111: 10630–10635.

Keeling CI, et al. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol. 2013; 14: R27.

Korol A, Rashkovetsky E, Iliadi K, Nevo E. *Drosophila* flies in "Evolution Canyon" as a model for incipient sympatric speciation. Proc Natl Acad Sci USA. 2006; 103: 18184–18189.

Krugman T, Satish N, Vinogradova O, Beharav A, Kashi Y, Nevo E. Genome diversity in the cyanobacterium *Nostoc linckia* at "Evolution Canyon", Israel, revealed by inter-HIP1 size polymorphisms. Evol Ecol Res. 2001; 3: 899–915.

Li K, et al. Sympatric speciation of spiny mice, *Acomys*, unfolded transcriptomically at "Evolution Canyon", Israel. Proc Natl Acad Sci USA. 2016; 113: 8254–8259.

Li K, et al. Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. Proc Natl Acad Sci USA. 2015; 112: 11905–11910.

Li K, et al. Transcriptome, genetic editing, and microRNA divergence substantiate sympatric speciation of blind mole rat, *Spalax*. Proc Natl Acad Sci USA. 2016; 113: 7584–7589.

Li K, et al. Transcriptome, genetic editing, and microRNA divergence substantiate sympatric speciation of blind mole rat, *Spalax*. Proc Natl Acad Sci USA. 2016; 113: 7584–7589.

Li K, et al. Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. Proc Natl Acad Sci USA. 2015; 112: 11905–11910.

Lynch M, Bobay L-M, Catania F, Gout J-F, Rho M. The repatterning of eukaryotic genomes by random genetic drift. Annu Rev Genom Hum Genet. 2011; 12: 347–366.

Mayr E. Animal species and evolution. Harvard University Press; Cambridge, MA. 1963.

Nachman MW, Payseur BA. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos Trans R Soc Lond, B, Biol Sci. 2012; 367: 409–421.

Nevo E. Evolution in action: Adaptation and incipient sympatric speciation with gene flow across life at "Evolution Canyon", Israel. Isr J Ecol Evol. 2014; 60: 85–98.

Nevo E. "Evolution Canyon": a microcosm of life's evolution focusing on adaptation and speciation. Isr J Ecol Evol. 2006; 52: 501–506.

Nevo E. Asian, African and European biota meet at "Evolution Canyon" Israel: local tests of global biodiversity and genetic diversity patterns. Proc R Soc Lond B Biol Sci. 1995; 262: 149–155.

Nevo E. Molecular evolution and ecological stress at global, regional and local scales: The Israeli perspective. J Exp Zool A Comp Exp Biol. 1998; 282: 95–119.

Nevo E, Fragman O, Dafni A, Beiles A. Biodiversity and interslope divergence of vascular plants caused by microclimatic differences at "Evolution Canyon", Lower Nahal Oren, Mount Carmel, Israel. Isr J Plant Sci. 1999; 47: 49–59.

Nevo E, Pirlot P, Beiles A. Brain size diversity in adaptation and speciation of subterranean mole rats. J Zool Syst Evol Res. 1988; 26: 467–479.

Nekrutenko A, Li W-H. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 2001; 17: 619–621.

Ohta T. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 1992; 23: 263–286.

Otto SP, Michalakis Y. The evolution of recombination in changing environments. Trends Ecol Evol. 1998; 13: 145–151.

Parnas T. Evidence for incipient sympatric speciation in wild barley, *H. spontaneum*, at "Evolution Canyon", Mt. Carmel, Israel, based on hybridization and physiological and genetic diversity estimates: Thesis submitted in partial fullfilment of the requirements for the Master's Degree, University of Haifa; 2006.

Pavlícek T, Sharon D, Kravchenko V, Saaroni H, Nevo E. Microclimatic interslope differences underlying biodiversity contrasts in "Evolution Canyon", Mt. Carmel, Israel. Isr J Earth Sci. 2003; 52: 1–9.

Provine WB. Ernst Mayr: genetics and speciation. Genetics. 2004; 167: 1041–1046.

Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010; 20: R208–R215.

Qi J, et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet. 2013; 45: 1510–1515.

Qian C, et al. Transcriptomes divergence of *Ricotia lunaria* between the two micro-climatic differentiated interslopes at the "Evolution Canyon", Israel. Front Genet. 2018; 9: 506.

Richards S, et al. The genome of the model beetle and pest *Tribolium castaneum*. Nature. 2008; 452: 949–955.

Sharaf K, Bureš P, Horová L, Pavlíček T, Nevo E. Distribution of abundance and genome size variability in the grain beetle *Oryzaephilus surinamensis* (Linnaeus, 1758)(Coleoptera: Silvanidae). Zool Middle East. 2008; 45: 79–90.

Sharaf K, Hadid Y, Pavlíček T, Nevo E. Local genetic population divergence in a saw-toothed grain beetle, *Oryzaephilus surinamensis* (L.)(Coleoptera, Cucujidae). J Stored Prod Res. 2013; 53: 72–76.

Sharaf K, Horová L, Pavlíček T, Nevo E, Bureš P. Genome size and base composition in *Oryzaephilus surinamensis* (Coleoptera: Sylvanidae) and differences between native (feral) and silo pest populations in Israel. J Stored Prod Res. 2010; 46: 34–37.

Sharaf K. Adaption and speciation patterns in the saw-toothed grain beetle *Oryzaephilus surinamensis* (Linnaeus, 1758), (Coleoptera: Cucujidae). PhD thesis. 2013: 127.

Sharaf K, Hadid Y, Pavlíček T, Nevo E. Local genetic population divergence in a saw-toothed grain beetle, *Oryzaephilus surinamensis* (L.)(Coleoptera, Cucujidae). J Stored Prod Res. 2013; 53: 72–76.

Satish N, Krugman T, Vinogradova O, Nevo E, Kashi Y. Genome evolution of the cyanobacterium *Nostoc linckia* under sharp microclimatic divergence at "Evolution Canyon", Israel. Microb Ecol. 2001; 42: 306–316.

Schrader L, et al. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun. 2014; 5: 5495.

Semeao AA, Campbell JF, Beeman RW, Lorenzen MD, Whitworth RJ, Sloderbeck PE. Genetic structure of *Tribolium castaneum* (Coleoptera: Tenebrionidae) populations in mills. Ecol Entomol. 2012; 41: 188–199.

Sikorski J, Nevo E. Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. Proc Natl Acad Sci USA. 2005; 102: 15924–15929.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123: 585–595.

Vanbergen AJ, Raymond B, Pearce IS, Watt AD, Hails RS, Hartley SE. Host shifting by *Operophtera brumata* into novel environments leads to population differentiation in life-history traits. Ecol Entomol.2003; 28: 604–612.

Valamoti S, Buckland P. An early find of *Oryzaephilus surinamensis* (L.)(Coleoptera: Silvanidae) from final neolithic Mandalo, Macedonia, Greece. J Stored Prod Res. 1995; 31: 307–309.

Wang H, et al. Sympatric speciation of wild emmer wheat driven by ecology and chromosomal rearrangements. Proc Natl Acad Sci USA. 2020; 117: 5955–5963.

Wu C-I, Ting C-T. Genes and speciation. Nat Rev Genet. 2004; 5: 114–122.

Zimmermann K, et al. Transient receptor potential cation channel, subfamily C, member 5 (TRPC5) is a cold-transducer in the peripheral nervous system. Proc Natl Acad Sci USA. 2011; 108: 18114–18119.

Zettler JL, Arthur FH. Chemical control of stored product insects with fumigants and residual treatments. Crop Protect. 2000; 19: 577–582.

Zhou Z, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015; 33: 408–414.