Spodoptera littoralis genome mining brings insights on the dynamic of expansion of gustatory receptors in polyphagous noctuidae

Camille Meslin, Pauline Mainet, Nicolas Montagné, Stéphanie Robin, Fabrice Legeai, Anthony Bretaudeau, J Spencer Johnston, Fotini A. Koutroumpa, Emma Persyn, Christelle Monsempès, et al.

Spodoptera littoralis genome mining brings insights on the dynamic of expansion of gustatory receptors in polyphagous noctuidae

Camille Meslin [ID],[1] Pauline Mainet,[1] Nicolas Montagné,[1] Stéphanie Robin,[2,3] Fabrice Legeai [ID],[2,3] Anthony Bretaudeau,[2,3] J. Spencer Johnston,[4] Fotini Koutroumpa,[1,6] Emma Persyn [ID],[1,5] Christelle Monsempès,[1] Marie-Christine François,[1] Emmanuelle Jacquin-Joly [ID] [1,*]

[1]INRAE, Sorbonne Université, CNRS, IRD, UPEC, Université de Paris, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES-Paris), 78026 Versailles, France,
[2]INRAE, UMR Institut de Génétique, Environnement et Protection des Plantes (IGEPP), BioInformatics Platform for Agroecosystems Arthropods (BIPAA), Campus Beaulieu, 35042 Rennes, France,
[3]INRIA, IRISA, GenOuest Core Facility, Campus de Beaulieu, Rennes 5042, France,
[4]Department of Entomology, Texas A&M University, College Station, TX 77843, USA,
[5]CIRAD, UMR PVBMT, Réunion, France,
[6]Present address: INRAE, Université Tours, Infectiologie et Santé Publique (ISP), 37380 Nouzilly, France

*Corresponding author: INRAE, Sorbonne Université, CNRS, IRD, UPEC, Université de Paris, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES-Paris), 78026 Versailles, France. Email: emmanuelle.joly@inrae.fr

Abstract

The bitter taste, triggered via gustatory receptors, serves as an important natural defense against the ingestion of poisonous foods in animals, and the increased host breadth is usually linked to an increase in the number of gustatory receptor genes. This has been especially observed in polyphagous insect species, such as noctuid species from the *Spodoptera* genus. However, the dynamic and physical mechanisms leading to these gene expansions and the evolutionary pressures behind them remain elusive. Among major drivers of genome dynamics are the transposable elements but, surprisingly, their potential role in insect gustatory receptor expansion has not been considered yet. In this work, we hypothesized that transposable elements and possibly positive selection would be involved in the highly dynamic evolution of gustatory receptor in *Spodoptera* spp. We first sequenced de novo the full 465 Mb genome of *S. littoralis*, and manually annotated the main chemosensory genes, including a large repertoire of 373 gustatory receptor genes (including 19 pseudogenes). We also improved the completeness of *S. frugiperda* and *S. litura* gustatory receptor gene repertoires. Then, we annotated transposable elements and revealed that a particular category of class I retrotransposons, the SINE transposons, was significantly enriched in the vicinity of gustatory receptor gene clusters, suggesting a transposon-mediated mechanism for the formation of these clusters. Selection pressure analyses indicated that positive selection within the gustatory receptor gene family is cryptic, only 7 receptors being identified as positively selected. Altogether, our data provide a new good quality *Spodoptera* genome, pinpoint interesting gustatory receptor candidates for further functional studies and bring valuable genomic information on the mechanisms of gustatory receptor expansions in polyphagous insect species.

Keywords: *Spodoptera littoralis*; gustatory receptors; transposable elements

Introduction

Animals rely heavily on their sense of taste to discriminate between harmful poisonous foods, usually through the detection of bitter compounds, and beneficial sustenance. Interestingly, narrowness of food diets in animals is usually linked to a decreased number of gustatory receptors (GRs), in both mammals such as the blood-feeder bats (Hong and Zhao 2014), and in insects such as the body louse (Kirkness *et al.* 2010)—an obligate ectoparasite of human—the fig wasp *Ceratosolen solmsi* (Xiao *et al.* 2013)—specialized on *Ficus*—and many Lepidoptera specialist feeders, although mammals and insect GRs are unrelated. Reversely, the increased host breadth is usually linked to GR gene expansions. This has

been especially observed in polyphagous insects, including omnivorous species such as the American cockroach *Periplaneta americana* (Li *et al.* 2018) and herbivorous species such as noctuid species (Xu *et al.* 2016; Cheng *et al.* 2017; Gouin *et al.* 2017).

In polyphagous noctuids, the sequencing of the genomes of *Spodoptera frugiperda* and *Spodoptera litura* revealed GR repertoires of 231 and 237 genes (Cheng *et al.* 2017; Gouin *et al.* 2017), respectively, more than twice as much compared with other monophagous and oligophagous Lepidoptera species (*Bombyx mori*: 69 genes, *Heliconius melpomene*: 73 genes) (The International Silkworm Genome Consortium 2008; The Heliconius Genome Consortium 2012; Briscoe *et al.* 2013; Guo *et al.* 2017), suggesting that the number of GRs has greatly increased during evolution in polyphagous

Lepidoptera via gene tandem duplication. The genomic architecture of the GR family is thus well known in these species and, together with previous studies, it supports the evidence that the family evolved under a birth-and-death model as well as under different selective pressures depending on the clade considered (Nei and Rooney 2005; Briscoe et al. 2013; Almeida et al. 2014; Suzuki et al. 2018). Most of these GRs belong to clades grouping the so-called "bitter" receptors, but in fact the function of the majority of these GRs remains enigmatic. Although the bitter GR class exhibits the most dynamic evolution, the mechanisms leading to GR expansions and the evolutionary pressures behind them remain elusive. Among major drivers of genome dynamics are the transposable elements (TEs). TEs are very diverse and are distributed along genomes in a nonrandom way. Similar or identical TEs can induce chromosomal rearrangements such as deletions, insertions, and even duplications, features that are frequent in multigene families such as GRs. Surprisingly, their potential role in insect GR expansion has not been considered yet.

In order to study GR evolution and the potential role of TEs in GR expansion in more detail, we sequenced an additional genome of a *Spodoptera* species: *Spodoptera littoralis*. So far, only 38 GRs were identified (Poivet et al. 2013; Walker et al. 2019; Koutroumpa et al. 2021) in *S. littoralis* whereas several hundreds of GRs were annotated in its counterparts *S. litura* and *S. frugiperda*. To investigate this singularity, we report here the sequencing of the *S. littoralis* genome, its full assembly, functional automatic annotation and expert annotation of the main chemosensory gene families, namely soluble carrier proteins [odorant-binding proteins (OBPs) and chemosensory proteins (CSPs)] (Pelosi et al. 2018) and the 3 major families of insect chemosensory receptors [odorant receptors (ORs); ionotropic receptors (IRs), and GRs] (Robertson 2019). With a particular focus on gustation, we also reannotated GRs in *S. litura* and *S. frugiperda*. Then, we analyzed the evolutionary history of GRs, by looking at the enrichment for TEs in the vicinity of GRs and by analyzing selective pressures acting on the different GR clades.

Materials and methods

Estimation of S. littoralis genome size

The genome size of *S. littoralis* was estimated using flow cytometry. Genome size estimates were produced as described before (Johnston et al. 2019). In brief, the head of a *S. littoralis* adult male along with the head of a female *Drosophila virilis* standard (1C = 328 Mbp; 1C = amount of DNA in the gamete of homogametic sex) (Greilhuber et al. 2005) were placed into 1 ml of Galbrath buffer in a 2-ml Kontes Dounce and ground with 15 strokes of the A pestle. The released nuclei were filtered through a 40-μM nylon filter and stained with 25 μg/ml propidium iodide for 2 h in the dark at 4°C. The average red fluorescence of the 2C nuclei was scored with a Partec C flow cytometer emitting at 514 nm. The 1C genome size of *S. littoralis* was estimated as (average red florescence of the 2C *S. littoralis* peak)/(average fluorescence of the 2C *D. virilis* peak) ×328 Mbp.

Spodoptera littoralis genome sequencing and assembly

Biological material and genomic DNA extraction

Whole genomic DNA was extracted from 2 male larvae obtained after 2 generations resulting from a single pair of *S. littoralis* originating from an inbred laboratory colony maintained in INRAE Versailles since 2000s on a semiartificial diet (Poitout and Bues 1974) at 24 ± 2°C and 65 ± 5% relative humidity under 16:8 h

light:dark photoperiod. The sex of individuals was verified by checking for presence of testis. The gut was removed and DNA extraction was performed from whole, late-stage larvae using Qiagen Genomic-tip 500/G (Qiagen Inc., Chatsworth, CA, USA). A total of 30 μg of genomic DNA were obtained.

Sequencing

Different types of libraries were generated for 2 sequencing technologies: Illumina and PacBio. For Illumina sequencing, 5 libraries were prepared and constructed according to the Illumina manufacturer's protocol (1 library of 170, 1 of 250, and 3 of 500 bp). Illumina sequencing was performed at the BGI-tech facilities (Shenzen, China) on a HiSeq2500 machine. Around 68 Gb were obtained, representing 144× of the estimated genome size (470 Mb) (Supplementary Data 1). The raw reads were filtered at BGI to remove adapter sequences, contaminations, and low-quality reads and the quality of all raw reads was assessed using FASTQC (Andrews S, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). PacBio sequencing was performed at GenoScreen (Lille, France) by the SMRT sequencing technology on 9 SMRTcell RSII, generating 2 846 820 reads. Around 16 Gb were obtained, representing 34× of the estimated genome size (Supplementary Data 1). High-quality sequences were obtained by generating circular consensus sequencing.

Genome assembly

A first assembly was done using Platanus (v1.2.1) (Kajitani et al. 2014) with Illumina data. A second assembly was obtained by doing scaffolding with SSPACE-LR (modified) (Boetzer and Pirovano 2014) using PacBio data and gap filling using GapCloser (Boetzer and Pirovano 2012). These second assembly was evaluated using Benchmarking Universal Single-Copy Orthologue (BUSCO v5.2.2) (Simão et al. 2015) with a reference set of 5,286 genes conserved in Lepidoptera.

Structural and functional genome annotation

Structural automatic genome annotation was done with BRAKER (v1.11) (Hoff et al. 2016) using all RNAseq data described in Supplementary Data 1. RNAseq libraries were sequenced from different larvae and adult tissues from males and females including the proboscis, palps, legs, and ovipositor and sequenced by Illumina (Supplementary Data 1) (Poivet et al. 2013; Walker et al. 2019; Koutroumpa et al. 2021). Reads were trimmed using Trimmomatic (v0.36) (Bolger et al. 2014) with the following parameters: ILLUMINACLIP: TruSeq2-PE.fa: 2:30:10, LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 4:15, MINLEN: 36. Trimmed reads were mapped on the genome assembly using STAR (v.5.2a) (Dobin et al. 2013) with the default parameters except for the following parameters: outFilterMultimapNmax = 5, outFilterMismatchNmax = 3, alignIntronMin = 10, alignIntronMax = 50,000, and alignMatesGapMax = 50,000. As done for the genome assembly, gene annotation was evaluated using BUSCO v5.2.2 with a reference set of 5,286 proteins conserved in Lepidoptera. Putative functions of predicted proteins were assigned using blastp (v2.6.0) against GenBank NR (nonredundant GenBank CDS translations + PDB + SwissProt + PIR + PRF) release 2017 September, and interproscan v5.13-52.0 against Interpro. Associated GO terms were collected from blast NR and interproscan results with blast2GO (v2.5).

Annotation of OBPs, CSPs, ORs, and IRs

The annotation of genes encoding soluble transporters (OBPs and CSPs), ORs, and IRs was performed using known sequences from other species with their genome sequenced (*S. frugiperda*, *S. litura*,

B. mori, *H. melpomene*, and *Danaus plexippus*) (Zhan *et al.* 2011; Briscoe *et al.* 2013; Cheng *et al.* 2017; Gouin *et al.* 2017; Guo *et al.* 2017). For all annotations in *S. frugiperda* genome, we use the "corn" strain as reference (Gouin *et al.* 2017). For each type of gene family, the set of known amino acid sequences and the genome sequence of *S. littoralis* were uploaded on the BIPAA galaxy platform to run the following annotation workflow. First, known amino acid sequences were used to search for *S. littoralis* scaffolds potentially containing genes of interest using tblastn (Camacho *et al.* 2009). All *S. littoralis* scaffolds with significant blast hits (e-value <0.001) were retrieved to generate a subset of the genome. Amino acid sequences were then aligned to this subset of the genome using Scipio (Keller *et al.* 2008) and Exonerate (Slater *et al.* 2005) to define intron/exon boundaries and to create gene models. Outputs from Scipio and Exonerate were then visualized on a Apollo browser (Lee *et al.* 2013) available on the BIPAA platform. All gene models generated have been manually validated or corrected via Apollo. Based on homology with other lepidopteran sequences and on RNAseq data available for *S. littoralis* (Poivet *et al.* 2013; Walker *et al.* 2019; Koutroumpa *et al.* 2021), matching halves were joined when located on different scaffolds. The classification of deduced proteins and their integrity were verified using blastp against the nonredundant (NR) GenBank database. When genes were suspected to be split on different scaffolds, protein sequences were merged for further analyses. A previous transcriptomic work identified 38 OBP genes in *S. litura* (Gu *et al.* 2015). This number being low compared with the repertoire of other lepidopteran species, *S. litura* OBPs were also annotated in the recent genome (Cheng *et al.* 2017). For OBPs and CSPs, SignalP-5.0 (Almagro Armenteros *et al.* 2019) was used to determine the presence or absence of a signal peptide. Hereafter, the abbreviations Slit, Slitu, and Sfru (for *S. littoralis*, *S. litura*, and *S. frugiperda*, respectively) are used before gene names to clarify the species.

Iterative annotation and reannotation of GRs
The initial annotation of GR genes was carried out the same way as for the other genes involved in chemoreception with 1 modification: at the end of the manual curation, all the newly identified amino acid GR sequences were added to the query set of known GR sequences to perform a new cycle of annotation. This iterative strategy was used for *S. littoralis* as well as for *S. litura* and *S. frugiperda* and was performed until no new GR sequence was identified.

At the end of the annotation, all GR amino acid sequences were aligned for each species individually using MAFFT v7.0 (Katoh and Standley 2013) in order to identify and filter allelic sequences and to verify the presence of the conserved GR domain (TYhhhhhQF in the transmembrane domain 7) (Brand *et al.* 2018) (Supplementary Data 2). Alleles were considered as such when they shared at least 90% identity with other annotated sequences. Between alleles, only the longest sequence was retained for further analysis. Pseudogenes were identified as partial sequences containing one or multiple stop codons. Genes were considered complete when both following conditions were met: (1) a start and a stop codon were identified and (2) a sequence length >350 amino acids. *Spodoptera littoralis* gene names were attributed based on orthology relationships with *S. frugiperda* when possible. *Spodoptera frugiperda* newly identified genes compared with the previous publications were numbered starting from SfruGR232. *Spodoptera litura* newly identified gene names were numbered starting from SlituGR240.

Annotation and enrichment analysis of TEs around chemosensory receptor genes in Spodoptera species
The annotation of TEs in *S. littoralis* genome was performed using REPET (Galaxy Lite v2.5). The TEdenovo pipeline (Flutre *et al.* 2011) was used to identify consensus sequences representative of each type of repetitive elements. Only contigs of a length >10 kb were used as input for the pipeline. Consensus sequences were built only if at least 3 similar copies were detected in the genome. The TEannot pipeline (Quesneville *et al.* 2005) was then used to annotate all repetitive elements in the genome using the library of TE consensus and to build an NR library in which redundant consensus were eliminated (length ≥98%, identity ≥95%). The NR library of TEs was finally used to perform the *S. littoralis* genome annotation with the TEannot pipeline.

The tool Locus Overlap Analysis (LOLA) within the R package Bioconductor (Sheffield and Bock 2016) was used to test for enrichment of TEs within the genomic regions containing chemosensory receptor genes (ORs and GRs) in both *S. littoralis* and *S. frugiperda*. To run LOLA with data from *S. littoralis*, 3 types of datasets were created. The first dataset, the query set, contained genomic regions of 10 kb around each chemosensory receptor gene. Since these genes were mostly organized in clusters within the genome, regions with overlap were combined leading to the creation of 114 chemosensory regions for the GRs and 63 regions for the ORs. The second dataset, the region universe, contained all genic regions from the genome. The region universe was created by retrieving the gene coordinates from the Official Gene Set from both genome (OGS3.0 for *S. littoralis*, OGS2.2 for *S. frugiperda*), and expanding to 10 kb around each gene. Similarly to the GR regions, regions with overlap were combined. This led to the creation of 14,072 genic regions for *S. littoralis* and 11,053 genic regions for *S. frugiperda*. The last dataset, the reference dataset, contained the coordinates of TEs previously identified by the REPET analysis. The enrichment in TE content within the chemosensory regions and the control regions were then compared using LOLA using a Fisher's Exact Test with false discovery rate correction (*q*-value) to assess the significance of overlap in each pairwise comparison. The same method was used using *S. frugiperda* TEs, previously annotated using the same tool REPET (Gouin *et al.* 2017), as well as chemosensory receptor reannotations from the present work and led to the creation of 191 chemosensory regions for the GRs and 88 regions for the ORs.

Evolutionary analyses
Phylogenetic tree reconstructions
Chemosensory-related protein trees were constructed for OBPs, CSPs, ORs, IRs, and GRs. For GRs, the phylogeny was built using GR amino acid sequences from different Lepidoptera species with various diets. In order to take into account the whole repertoire of GRs in our analysis, only species in which the GRs were annotated following whole-genome sequencing were considered. The dataset contained GRs from polyphagous (*S. littoralis*, *S. litura*, and *S. frugiperda*), oligophagous (*H. melpomene*—73 GRs, *Manduca sexta*—45 GRs) and monophagous species (*B. mori*—72 GRs). The multiple sequence alignment of all GR amino acid sequences (expect short partial sequences) was performed with ClustalO (Sievers *et al.* 2011) and the phylogeny was reconstructed using PhyML 3.0 (Guindon and Gascuel 2003) (http://www.atgc-montpellier.fr/phyml/) with the automatic selection of the best substitution model by SMS (Lefort *et al.* 2017). The CO_2 and sugar receptor clades were used to root the tree. The resulting phylogenetic tree was edited using FigTree v1.4.2 (https://github.com/rambaut/fig

tree) and Inkscape 0.92 (https://inkscape.org/fr/). Branch supports were estimated using the approximate likelihood-ratio test (Anisimova and Gascuel 2006) implemented at http://www.atgc-montpellier.fr/phyml/. For other gene families, sequences from various Lepidoptera species were retrieved and aligned with *S. littoralis* sequences using MAFFT (Katoh and Standley 2013). The reconstruction of the phylogenetic trees was carried out the same way as for the GRs, and the OR tree was rooted with the Orco clade.

Tree reconciliation

Estimates of gains and losses of GR genes across the Noctuidae were inferred using the reconciliation methods implemented in Notung v2.6 (Stolzer *et al.* 2012; Darby *et al.* 2016). The species tree was generated using TimeTree.org (Kumar *et al.* 2017) and the gene tree was the reconstructed phylogeny of the GRs generated by PhyML.

Evolutionary pressures

The codeml software of the package PAML was used to infer selective pressures (Yang 2007). Because of the high divergence between GRs across the phylogeny, selective pressures were inferred on 13 subtrees extracted from the GR phylogeny in order to minimize the ratio of synonymous substitutions. For each subtree, an alignment of the protein sequences was performed using MAFFT, converted to codon alignment using PAL2NAL (Suyama *et al.* 2006), and a phylogenetic tree was reconstructed based on the protein sequence alignment. Sequences introducing large gaps in the alignment were removed in order to compute codeml on the largest alignment possible. To estimate the selective pressures acting on the evolution of the lepidopteran GR genes, the "m0 model" from codeml of the PAML package was computed on the 13 subtrees to estimate the global ω (ratio of nonsynonymous substitutions dN/ratio of synonymous substitutions dS) (Yang *et al.* 2000). The ω value reflects the mode of evolution, with $\omega > 1$ indicating positive selection, $\omega < 1$ indicating purifying selection, and $\omega = 1$ indicating neutral evolution. To further infer positive selection, 2 comparisons between evolutionary models were conducted. First, the comparison between M8 and M8a models can detect positive selection acting on sites, i.e. columns of the alignment (Swanson *et al.* 2003; Wong *et al.* 2004). This comparison was conducted only when the global ω calculated from the m0 model was >0.3. The second comparison between branch-site model A and its neutral counterpart can detect positive selection acting on particular sites on a specific lineage (Zhang *et al.* 2005), a method reported to be more powerful than other comparisons implemented in PAML to detect episodic positive selection (Yang and Dos Reis 2011). Here, we tested all the terminal branches of the trees for which both the global ω was elevated and the comparison between models M8 and M8a statistically significant. Since many branches were tested for each tree, a correction for multiple testing to control for false discovery rate was applied: the q-value (q-value R package version 2.22.0; Storey *et al.* 2021). In the case of a statistically significant q-value (<0.05), positively selected sites were inspected for possible artifacts due to partial sequences or misalignment.

Putative functional assignation

In order to assign putative functions to several candidate SlitGRs, both their phylogenetic position and theoretical 3D structure were analyzed. For the theoretical structures, the AlphaFold algorithm (Jumper *et al.* 2021) was used to model candidate SlitGRs as well as their *B. mori* ortholog GRs with known function: BmorGR9

and BmorGR66. Structures were then compared between orthologs using the MatchMaker tool of Chimera and the root mean square deviation (RMSD) computed using the same tool (Pettersen *et al.* 2004). Docking of D-fructose was performed on both BmorGR9 and SlitGR9 using the Webina webserver (https://durrantlab.pitt.edu/webina/) (Kochnev *et al.* 2020).

Results and discussion
Genome assembly and automatic annotation of the *S. littoralis* genome

The first assembly of *S. littoralis* (v1.0), obtained with short Illumina reads, contained 123,499 scaffolds with a N50 of 18 kb and an assembly total size around 470 Mb. The second assembly (v2.0), obtained with a combination of both short Illumina and long PacBio reads, contained 28,891 scaffolds, with a N50 of 64 kb and an assembly total size around 465 Mb (Table 1). The genome size of *S. littoralis* was in good correlation with flow cytometry evaluation (470 Mb). The BUSCO analysis revealed that the second assembly contained more than 97% of complete BUSCO genes, with more than 95% of them being present in single-copy (Table 2). This second assembly was then used as the final assembly in all the following analyses. A total of 35,801 genes were predicted using BRAKER (OGS3.0_20171108). The number of genes annotated in lepidopteran genomes is usually much lower (around 15,000 genes in general). This unusual number is probably due to duplicates, as revealed by the BUSCO analysis (10% duplicated annotated genes). However, almost 97% of BUSCO proteins were complete, with more than 86% being present in single-copy (Table 2). These data show that the *S. littoralis* genome assembly is of good quality, thus allowing for accurate comparison with other *Spodoptera* genomes.

OBP, CSP, OR, and IR chemosensory gene repertoires were of comparable size among *Spodoptera* spp.

To have a full view of the *S. littoralis* chemosensory gene repertoires, we manually curated all the major chemosensory-related gene families, including soluble carrier proteins (OBPs and CSPs), proposed to facilitate chemical transfer to chemosensory receptors (Pelosi *et al.* 2018), and the membrane bound receptors (ORs: 7 transmembrane receptors expressed in the membrane of olfactory sensory neurons; GRs: 7 transmembrane receptors hosted by taste neurons; and IRs: 3 transmembrane proteins sensing acids and amines).

The genome of *S. littoralis* contained 23 CSP genes, all of them encoding full-length sequences with a signal peptide. This number of genes is similar to the 22 CSP genes annotated in *S. frugiperda* (Gouin *et al.* 2017) and the 23 CSP genes annotated in *S. litura* (Cheng *et al.* 2017). Among all these sequences, 16 CSP genes are 1:1 orthologs between the 3 *Spodoptera* species included in the tree while 11 CSP genes are 1:1 orthologs with BmorCSPs

Table 1. Statistics of the *S. littoralis* genome assemblies.

	Slit genome v1.0	Slit genome v2.0
Number of scaffolds	123,499	28,891
Total size of scaffolds	470 Mb	465 Mb
Longest scaffold	236 kb	816 kb
N50 scaffold length	18 kb	64 kb
scaffold %N	0.41	0.92

Table 2. BUSCO statistics on *S. littoralis* genome and annotation.

	Slit genome v2.0	Annotation BRAKER OGS3.0
Complete BUSCOs (C)	5,139 (97.2%)	5,111 (96.7%)
Complete and single-copy BUSCOs (S)	5,049 (95.5%)	4,563 (86.3%)
Complete and duplicated BUSCOs (D)	90 (1.7%)	548 (10.4%)
Fragmented BUSCOs (F)	98 (1.9%)	131 (2.5%)
Missing BUSCOs (M)	49 (0.9%)	44 (0.8%)
Total BUSCO groups searched	5,286	5,286

(from *B. mori*), showing the high level of conservation in this gene family (Supplementary Fig. 1 and Data 3).

We also annotated 51 OBP genes in *S. littoralis*. Among these genes, 48 were complete and 47 possessed a signal peptide (Supplementary Data 3). The phylogenetic tree revealed a clade enriched in *Spodoptera* OBPs (9 SlitOBPs, 9 SlituOBPs, and 10 SfruOBPs) (Supplementary Fig. 2). This expansion probably arose from recent tandem duplications at the base of this group as most of the genes of the expansion are organized in synteny in the 3 species (Supplementary Fig. 3).

We annotated 44 IR genes in the *S. littoralis* genome, 43 of which encoding a full-length sequence with various sizes containing 547–948 amino acids (Supplementary Data 3). In addition to the 2 conserved coreceptors IR8a and IR25a (Croset *et al.* 2010), we identified 18 candidate antennal IRs putatively involved in odorant detection, 23 divergent IRs putatively involved in taste, and 12 ionotropic glutamate receptors (iGluRs). The total IR number was similar to the 44 IR genes annotated in *S. litura* (Zhu *et al.* 2018) and the 43 IR genes annotated in *S. frugiperda* (Gouin *et al.* 2017). Among all these sequences, 43 IR genes are 1:1 orthologs between the 3 *Spodoptera* species (IR100g was missing in *S. frugiperda*). The phylogenetic tree revealed a clade containing divergent IRs and 2 lineage expansions were observed (IR7d and IR100), likely attributed to gene duplications (Zhu *et al.* 2018). The number of divergent IRs was much higher in *Spodoptera* species (*S. littoralis*: 26, *S. litura*: 26, *S. frugiperda*: 25) than in *H. melpomene* (16) and *B. mori* (6). By contrast, phylogenetic analysis (Supplementary Fig. 4) reveals that *S. littoralis* antennal IRs retained a single copy within each orthologous group.

We annotated 73 OR genes in the *S. littoralis* genome scattered among 61 scaffolds (Supplementary Data 2), including the obligatory coreceptor ORco. The number of OR genes in the *S. littoralis* repertoire was similar to the repertoire of closely related species (69 in *S. frugiperda*, 73 in *S. litura*) and other Lepidoptera (64 in *D. plexippus*, 73 in *M. sexta*). The phylogenetic tree of ORs is presented in Supplementary Fig. 5.

Altogether, our annotations revealed that OBP, CSP, OR, and IR repertoires were of comparable size among the *Spodoptera* spp. investigated.

A highly dynamic evolution of the GR multigene family in *Spodoptera* species

Newly obtained genomes of polyphagous noctuidae species such as *H. armigera* (Pearce et al. 2017), *S. litura* (Cheng *et al.* 2017), *S. frugiperda* (Gouin *et al.* 2017), and *Agrotis ipsilon* (Wang *et al.* 2021) revealed an important expansion of GRs in these species, suggesting an adaptation mechanism to polyphagy. Here, using these known GR protein sequences and an iterative annotation process, we annotated an even larger repertoire of GRs in the *S. littoralis* genome. In view of these data, we searched for possible missing GRs in the *S. frugiperda* and *S. litura* genomes to complete their GR repertoires (Table 3). We annotated a total of 376 GR genes scattered on 110 scaffolds in the genome of *S. littoralis*, and

reannotated 417 GRs on 196 scaffolds in the *S. frugiperda* genome and 293 GRs on 30 scaffolds in *S. litura* (Supplementary Data 4). When omitting pseudogenes and alleles, the final repertoires of GRs are composed of 325 genes in *S. littoralis*, 278 GRs in *S. frugiperda*, and 280 GRs in *S. litura*. Our GR analysis not only revealed that the full repertoire of *S. littoralis* GRs is in fact much bigger than previously reported, but also that the GR numbers in *S. litura* and *S. frugiperda* have been under evaluated (although the presence of some alleles may over evaluate these numbers). Among these sequences, several were indeed allelic version of previously annotated genes but several new genes were also identified (Table 3). Among these genes, the percentage of complete genes varied between species, from only 41% in *S. frugiperda* compared with 79% in *S. litura* while the percentage of complete GRs in *S. littoralis* was intermediate (73%). The percentage of allelic sequences were also highly variable, probably depending on the heterozygosity level of each considered genome and quality of the assembly (Supplementary Data 5). Indeed, the highest number of alleles was reached in *S. frugiperda*, a genome with a high level of heterozygosity (Gouin *et al.* 2017), while alleles were less frequent in the 2 other *Spodoptera* genomes considered. Multiple partial genes were also part of these repertoires, many of them being located at the boundaries of scaffolds. For the partial genes annotated in the middle of scaffolds, several explanations are possible: either because of misassembly in the region, or because the tools used to annotate failed to identify some exons or even because these partial sequences could be in fact pseudogenes. As previously shown, multiple clusters of GRs were also found in the *S. littoralis* genome. The 2 main clusters were found on scaffolds 1,414 and 878 that contained each 27 GR genes. The phylogeny reconstructed using the GR sequences from the 3 *Spodoptera* species as well as those from *B. mori* (BmorGRs) and *H. melpomene* (HmelGRs) showed that a few *Spodoptera* GRs clustered with candidate CO_2, sucrose and fructose receptors, while the majority of the *Spodoptera* GRs were part of the so-called bitter receptor clades. Among the candidate bitter receptor clades, 11 clades were enriched in *Spodoptera* genes (numbered from A to K in Fig. 1) and encompassed the majority of the 3 *Spodoptera* GR repertoires (Table 4). When belonging to the same phylogenetic clade, GRs from the same species tend to be located on the same scaffold, supporting the theory of the expansion of these genes by tandem duplications and few gene losses. For the subsequent analysis, only complete and partial genes were considered while pseudogenes were discarded. Four *S. littoralis* GRs with only 1 exon were identified, clustered on scaffold 67 and belonging to the same phylogenetic clade (Fig. 1). Interestingly, this clade was very conserved with a 1:1 orthology relationship between the 3 *Spodoptera* species, the SlituGRs and SfruGRs being also monoexonic. All these monoexonic genes are orthologs with BmorGR53, a single exon gene that is highly expressed at the larval stage but not in the adult (Guo *et al.* 2017). BmorGR53 is able to detect the bitter tastant and feeding deterrent coumarine. It is then likely

Table 3. GR repertoires of *Spodoptera* species.

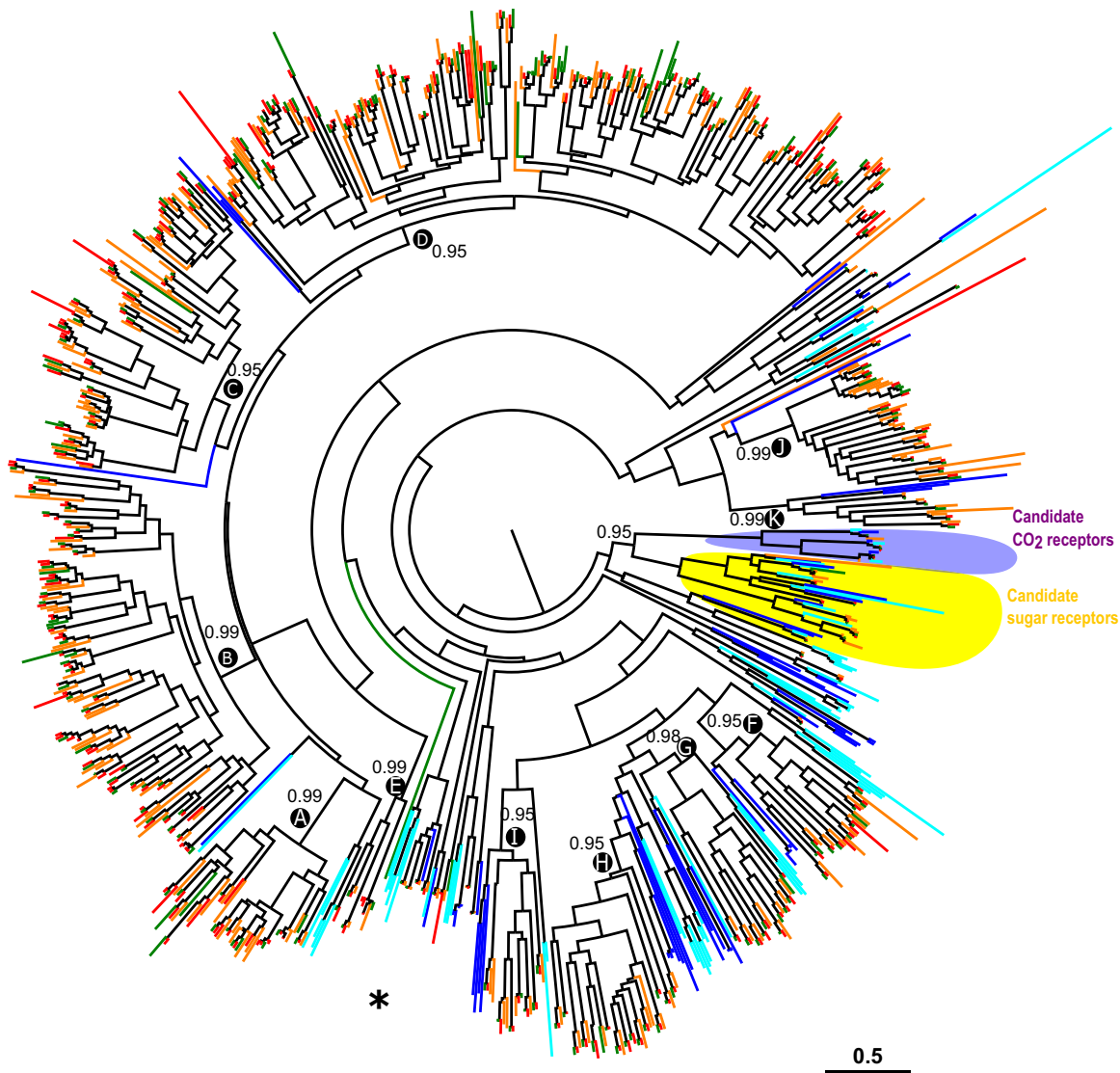|	*Spodoptera littoralis*	*Spodoptera frugiperda*	*Spodoptera litura*
Number of GR previously annotated	38	231	237
Complete genes	275 (73%)	172 (41%)	231 (79%)
Partial genes	50 (13%)	106 (25%)	49 (17%)
Pseudogenes	19 (5%)	22 (5%)	7 (2%)
Alleles	29 (8%)	117 (28%)	6 (2%)
Total in this work	373	417	293

Fig. 1. Phylogeny of lepidopteran GRs. The dataset included amino acid sequences from *S. littoralis* (Noctuoidea, red), *S. litura* (Noctuoidea, green), *S. frugiperda* (Noctuoidea, orange), *B. mori* (Bombycoidea, blue), and *H. melpomene* (Papilionoidea, cyan). Sequences were aligned using ClustalO and the phylogenetic tree was reconstructed using PHYML. CO_2 receptor candidates as well as sugar receptor candidates are indicated in purple and yellow, respectively. All the other GRs are part of the bitter receptor clades. The star indicates the clade of single-exon GRs. The clade containing putative CO_2 and sugar receptors was used to root the tree. Bootstrap values are indicated for the main clades. The scale bar represents 0.5 amino acid substitutions per site.

that these 4 single exon GRs play an important role in host–plant recognition in *Spodoptera* species as well.

The GR phylogeny served as a basis for the reconciliation of the gene- and species-tree in order to estimate gene gains and losses. The Notung analysis revealed that the ancestral repertoire of GRs of Noctuidae species contained 58 genes (Fig. 2). Given the

numbers of GRs annotated in *Spodoptera* species, it is not surprising that the highest gene gains occurred in the ancestor of *Spodoptera* species (296 gene gains). However, even for species with a smaller repertoire of genes such as *B. mori* (70 GRs) and *H. melpomene* (73 GRs), the turnover of genes compared with the ancestors is high (33 and 41 gene gains, 25 and 26 gene losses, respectively).

Table 4. Number of *Spodoptera* putative bitter receptors by expansion clade.

Clade	*Spodoptera littoralis*	*Spodoptera frugiperda*	*Spodoptera litura*
A	20	15	14
B	65	42	44
C	40	40	33
D	97	74	89
E	7	3	4
F	12	10	11
G	10	8	10
H	16	11	16
I	7	8	7
J	16	21	18
K	4	6	5
Total	294 (90.5%)	238 (86.9%)	251 (89,6%)

The percentages represent the proportion of *Spodoptera* genes to the total number of GRs annotated in the 3 *Spodoptera* species (complete + partial genes indicated in Table 3).
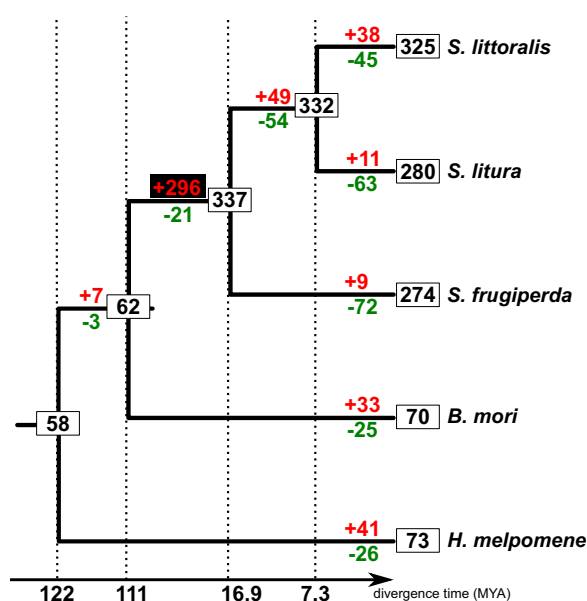

Fig. 2. GR gain and loss estimates across lepidopterans. The gene tree of GRs generated using PhyML was reconciled with the species-tree using Notung (Stolzer *et al.* 2012) to estimate gene gains and losses. Numbers in boxes represent the size of GR repertoire for extant species as well as ancestors at the nodes of the species tree. Gene gains are indicated in red while gene losses are indicated in green. The expansion that occurred in the ancestor of *Spodoptera* species is indicated in red on a black background.

Annotation of TEs, enrichment analysis, and selection pressure

To get more insights about the mechanisms that led to the formation of massive genomic clusters of GR genes, we looked at (1) whether TEs could be involved and (2) the selective pressures acting on GR genes.

TEs have been shown to be involved in countless mechanisms of evolution in insects, such as insecticide resistance, the evolution of regulatory networks, immunity, climate adaptation (Chen and Li 2007; Rebollo *et al.* 2012; You *et al.* 2013; Chuong *et al.* 2017; Bourque *et al.* 2018; Singh *et al.* 2020; Parisot *et al.* 2021), and some of them have even been domesticated as genes (Maumus *et al.* 2015). Gene families involved in these traits have been shown to be enriched in TEs and gene family expansions have been correlated with TE content, for instance in termites (Harrison *et al.*

2018). Interestingly, enrichment in TEs has not been reported for insect GR gene clusters so far. While annotating GRs in the *S. littoralis* genome, we noticed the frequent co-occurrence of TEs on the same scaffolds. We thus annotated TEs in the *S. littoralis* genome and calculated their enrichment in the vicinity of GR genes. We also carried out the same enrichment analysis in *S. frugiperda* genomes, as TE annotation in this last species has been done using the same REPET pipeline as in our study. The de novo constructed library contained 1,089 consensus sequences of TEs and was used to annotate the *S. littoralis* genome. The repeat coverage for the *S. littoralis* genome was 30.22%, representing 140 Mb, which is similar to that of *S. frugiperda* (29.10%), *S. litura* (31.8%), and *S. exigua* (33.12%) (Cheng *et al.* 2017; Gouin *et al.* 2017; Zhang, Zhang, Yang, *et al.* 2019). The relative contribution of the different classes of repetitive elements revealed that class I elements were more represented than class II elements (66.96% vs 20.83%), a classical feature of insect genomes (Maumus *et al.* 2015) (Fig. 3 and Table 5). However, the repartition and proportions between the different classes differed between these species. The class I SINE was the most represented in *S. frugiperda* (12.52%) (Gouin *et al.* 2017) while the class I LINE elements were the most represented in both *S. litura* and *S. exigua*, although with a lower proportion of all repeated elements (27.73% and 14.81%, respectively). Remarkably, the proportion of LINE elements identified in the *S. littoralis* genome was the highest reported so far in arthropods (Petersen *et al.* 2019), accounting for 52.18% of all repetitive elements. In 2 subspecies of the Asian gypsy moth *Lymantria dispar*, the accumulation of this particular class of TEs was found to be responsible for their large genome size (Hebert *et al.* 2019), a phenomenon also observed in other insect species (Maumus *et al.* 2015). The accumulation of the same elements in the *S. littoralis* genome could explain its larger size compared with its *Spodoptera* counterparts (465 Mb vs ∼400Mb for *S. frugiperda*, 438 Mb for *S. litura*, 408–448 Mb for *S. exigua*). The second most represented was DNA transposons, class II TIR elements, representing 11.04% of all TEs (Fig. 3 and Table 5).

The enrichment of TEs in the vicinity of GR gene clusters was tested in both *S. littoralis* and *S. frugiperda*, and we found that a particular category of class I retrotransposons, a SINE transposon, was significantly enriched in the vicinity of the GRs in the *S. littoralis* genome (*q*-value = 0.038), suggesting a transposon-mediated mechanism for the formation of GR clusters (Supplementary Data 6). SINE elements are typically small (80–500 bp) and originate from accidental retrotransposition of various polymerase III transcripts. These elements are nonautonomous, therefore their involvement in the dynamic of the GR multigene family may be related to their potential to induce genome rearrangements via unequal crossing over, hence potential drivers of duplication, as previously shown in other insect species. Given their prevalence in the *S. littoralis* genome, the potential role of these TEs in the GR family dynamic is probably just one of their numerous functions. The same enrichment analysis performed for the OR loci showed no significant enrichment in both *S. littoralis* and *S. frugiperda* (Supplementary Data 6). The involvement of TEs in the expansion of chemosensory gene families is not restricted to *S. littoralis* and GRs, as similar results were also found in ant genomes for ORs (Schrader *et al.* 2014; McKenzie and Kronauer 2018), suggesting that TEs are major players in the evolution of insect genomes and in species adaptation (Maumus *et al.* 2015; Gilbert *et al.* 2021).

Several studies have shown the importance of positive selection in the evolution of multigene families, especially in chemosensory genes such as ORs and GRs (McBride *et al.* 2007; Smadja
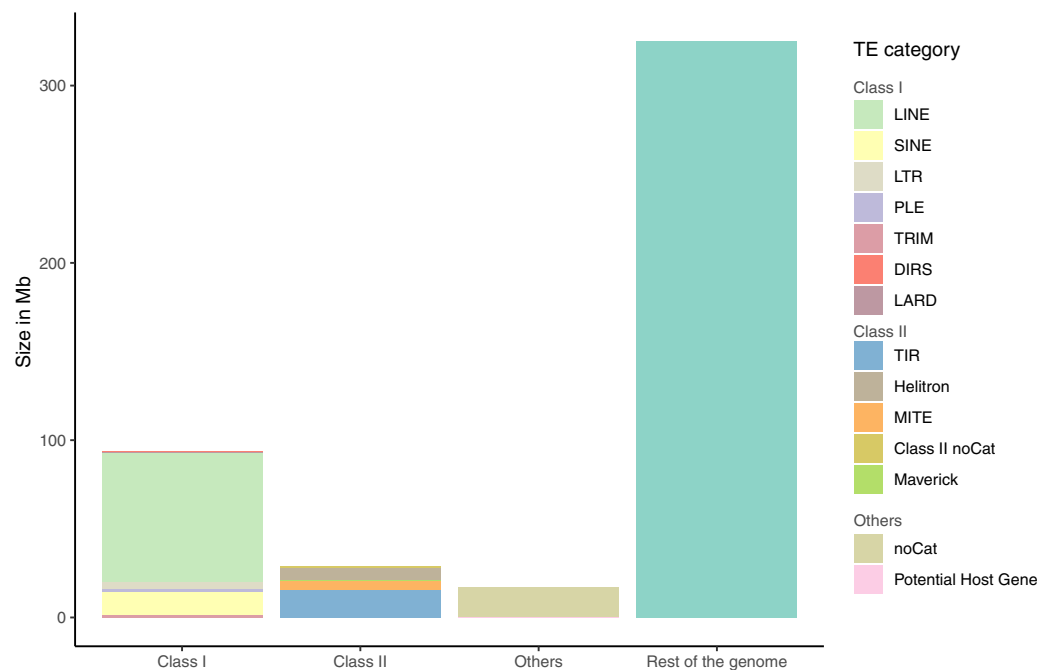
Fig. 3. Repartition and size of repeat content in *S. littoralis* genome. Repetitive elements account for 30.22% of *S. littoralis* genome. Class I elements are more abundant than class II. The class I LINE elements represent more than half of all repetitive elements.

Table 5. Repartition of repetitive elements in *S. littoralis* genome based on the classification established by Wicker *et al.* (2007).

	TE category	% of coverage of all repetitive elements
Class I retrotransposons	DIRS	0.20%
	LARD	0.19%
	LINE	52.18%
	LTR	3.02%
	PLE	1.12%
	SINE	9.33%
	TRIM	0.92%
Class II DNA transposons	Helitron	5.17%
	MITE	3.90%
	Maverick	0.01%
	TIR	11.04%
	Class II noCat	0.71%
Others	noCat	11.87%
	Potential host gene	0.35%

noCat means repetitive elements that could not be classified into the existing categories.

et al. 2009). Positively selected chemoreceptors may be linked to adaptation in *Drosophila* species (Hickner *et al.* 2016; Diaz *et al.* 2018). In the pea aphid, signatures of selection have been identified in chemosensory genes, including GRs and ORs, which may be implicated in the divergence of pea aphid host races. We thus analyzed selective pressures focusing on 13 clades of interest in the *Spodoptera* GR phylogeny: the potential clade of CO_2 receptors, the potential clade of sugar receptors and the 11 expended lineages within the so-called bitter receptor clades. For all 13 clades, we observed low global ω values ranging from 0.01 to 0.42, with the highest observed for candidate bitter receptor clades. The comparison between models M8 and M8a was statistically significant for clades C, F, and J, indicating a signal of positive selection. Branch-site models on terminal branches of the associated trees were then tested on these clades. For clade J, no GR was revealed as evolving under positive selection. However, for clades C and F, 2 and 5 GRs were identified as positively selected, respectively (Table 6). Within these GR sequences, very few positively selected

sites were identified for each gene (between 0 and 3; Supplementary Data 7). This finding is coherent with previous studies showing the same pattern of evolutionary rates (Engsontia *et al.* 2014; Suzuki *et al.* 2018), especially in *S. frugiperda* (Gouin *et al.* 2017) (3 GRs under positive selection when comparing 2 host strains). Taken together, all positive selection analyses indicate that positive selection within the GR gene family is cryptic and may not play an important role in shaping the evolution of *Spodoptera* GRs. Anyhow, the few positively selected GRs may be interesting candidates for further functional studies.

Putative functional assignation of candidate SlitGRs

The complexity of the evolution of the bitter GRs is reflected by their complex functioning. Indeed, in contrast with the relatively simple OR/Orco association that is the basis for olfaction, the molecular basis for gustation is marked by several characteristics that were recently identified in *D. melanogaster*. First, some GRs

Table 6. Selective pressure analysis.

Clade	No. of sequences	ω M0 (d_N/d_S)	P-value (M8 vs M8a)	Branch-site
A	45	0.34109	NS	/
B	133	0.34146	NS	/
C	86	0.34386	0.044804*	Slit_GR217, Slitu_GR155
D	218	0.29174	/	
E	14	0.18639	/	/
F	33	0.41616	0.000504**	Sfru_GR44, Sfru_GR49, Slit_GR44
G	28	0.36571	NS	/
H	38	0.31933	NS	/
I	16	0.22375	/	/
J	38	0.42257	0.005319**	NS
K	11	0.17393	/	/
Sugar	27	0.05662	/	/
CO_2	11	0.01074	/	/

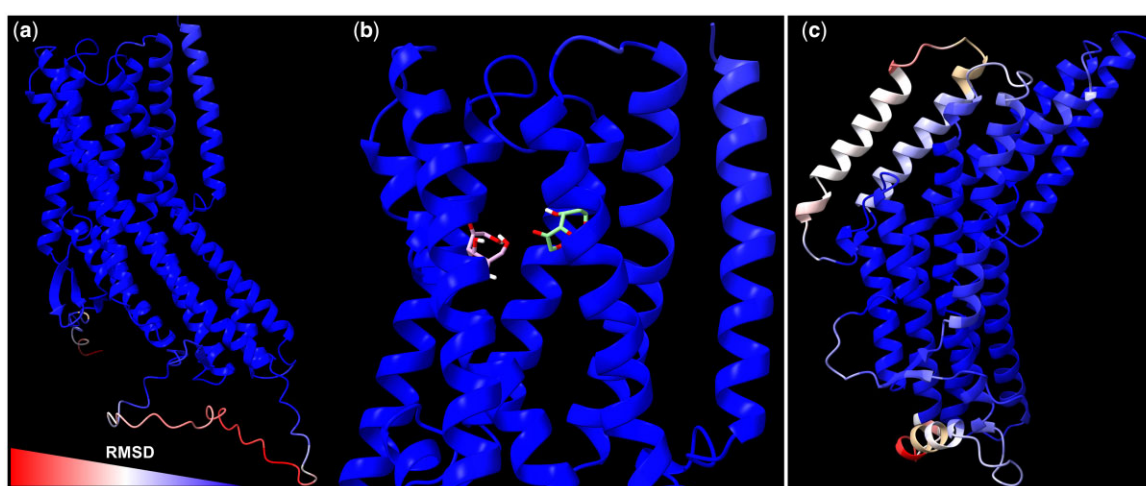NS, non-significant; /, not calculated.
* *p*-value <0.05.
** *p*-value <0.01.

Fig. 4. Comparison of 3D structures of BmorGR9 and BmorGR66 with their respective orthologs, SlitGR9 and SlitGR15. Three-dimensional structures were predicted using AlphaFold2 (Jumper *et al.* 2021). a, c) BmorGR9/SlitGR9 and BmorGR66/SlitGR15 were aligned using matchmaker. The RMSD value computed between both structures is represented here on the BmorGR9 and BmorGR66 structures, respectively, in a red (high RMSD) to blue palette (low RMSD). b) Docking of D-fructose in BmorGR9 (green) and in SlitGR9 (pink). In (a–c), structures are represented oriented with their extracellular domain at the top and the intracellular domain at the bottom.

have to be coexpressed within the same neuron in order to be able to respond to a stimulus. Second, it seems that GR–GR inhibition can modulate neuron responses. The challenge in the next few years will be to characterize both the response spectra and precise expression patterns of GRs of interest. However, those GRs of interest need to be selected. The present work provides us with some valuable candidates such as the single exon GRs for which the function is known in *B. mori*. Also, it seems that individual GRs can play an important role in the ecology of a species. Among examples are BmorGR9, which binds D-fructose without the need of any other GR (Sato *et al.* 2011; Mang *et al.* 2016), and BmorGR66, whose silencing confers to the monophagous *B. mori* larva the ability to feed from different food sources (Zhang, Zhang, Niu, *et al.* 2019). We identified their *S. littoralis* orthologs as SlitGR9 and SlitGR15, respectively. We predicted their 3D structures using AlphaFold and compared them with the AlphaFold predicted structures of *B. mori* orthologues. Globally, all the predicted structures obtained reasonable pLDDT scores, especially in the extracellular domain where the putative binding pocket may be located, with higher confidence for SlitGR9 and BmorGR9 than for BmorGR66 and SlitGR15 (Supplementary Data 8). The

RMSD computed between the atoms of the 3D structures of BmorGR9 and SlitGR9 ranged from 0.0427 to 26.5 Å (global RMSD: 7.804) (Fig. 4a). While disordered regions, that are difficult to predict, such as the N-terminal end or the loop between transmembrane domains 4 and 5 differ, both structures are strikingly similar, suggesting that SlitGR9 is likely a D-fructose receptor in *S. littoralis*. Additionally, docking experiments on each structure revealed that their binding pockets are located in the same region of the receptor (Fig. 4b). The ligand of BmorGR66 is not known, however, this receptor is responsible for the feeding difference of *B. mori* for mulberry leaves (Zhang, Zhang, Niu, *et al.* 2019). Its ortholog SlitGR15 is a key candidate for functional studies to test if this GR has an impact on the feeding preference in *S. littoralis* as well. When comparing both full structures, the RMSD was 3.431 Å (Fig. 4c). Interestingly, the main differences between both structures were visible in the extracellular domains of the proteins, suggesting that the binding pockets may differ as well.

Apart from these particular GRs, the neuronal coding of taste via more than 200 genes in species like *Spodoptera* moths is not known. But are all these GRs at play in effective taste sense? In fact, comparison of GR gene repertoires with transcript

repertoires showed that a small proportion of the gene repertoire is actually expressed in the canonical gustatory tissues of *Spodoptera* spp., as can be seen in *S. littoralis* and *S. litura* (Cheng *et al.* 2017; Walker *et al.* 2019; Koutroumpa *et al.* 2021). In addition, GR expression levels—especially that of bitter receptors—are rather low. Whether the genome acts as a "reservoir" for a multitude of GR genes to be selectively expressed in accordance with the evolution of food preference remains to be investigated. In that view, the identification of regulatory genomic regions and transcription factors in the vicinity of GR regions that may be at play in GR expression choice would help understanding if and how GRs evolved according to polyphagy.

Data availability

The assembled genome of *S. littoralis* as well as the genomic data of *Spodoptera litura* (v1.0) (Cheng *et al.* 2017) and *S. frugiperda* (Corn variant, v3.1) (Gouin *et al.* 2017) are all publicly available on the BIPAA platform (https://bipaa.genouest.org). In addition, *S. litura* (bioProject PRJNA344815) and *S. littoralis* genome are available on NCBI (JALJOW000000000). Files and code for the LOLA analysis are available at dataINRAE (https://data.inrae.fr/dataverse/slit-toralis_genome).

Supplemental material is available at G3 online.

Acknowledgments

Funding

Conflicts of interest

None declared.

Literature cited

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37(4):420–423.

Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. Family size evolution in *Drosophila* chemosensory gene families: a comparative analysis with a critical appraisal of methods. Genome Biol Evol. 2014;6(7):1669–1682.

Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 2006;55(4):539–552.

Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012;13(6):R56.

Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics. 2014;15:211.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, *et al.* Ten things you should know about transposable elements. Genome Biol. 2018;19(1):199.

Brand P, Robertson HM, Lin W, Pothula R, Klingeman WE, Jurat-Fuentes JL, Johnson BR. The origin of the odorant receptor gene family in insects eLife 2018;7:e38340.

Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, Jamie GA, Martin SH, Dasmahapatra KK, Ferguson LC, Mallet J, *et al.* Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genet. 2013;9(7):e1003620.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

Chen S, Li X. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. BMC Evol Biol. 2007;7:1–13.

Cheng T, Wu J, Wu Y, Chilukuri RV, Huang L, Yamamoto K, Feng L, Li W, Chen Z, Guo H, *et al.* Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. Nat Ecol Evol. 2017;1(11):1747–1756.

Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017; 18(2):71–86.

Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 2010;6(8):e1001064.

Croset V, Schleyer M, Arguello JR, Gerber B, Benton R. A molecular and neuronal basis for amino acid sensing in the *Drosophila* larva. Sci Rep. 2016;6. doi:10.1038/srep34871.

Darby CA, Stolzer M, Ropp PJ, Barker D, Durand D. Xenolog classification. Bioinformatics. 2016;33:btw686.

Diaz F, Allan CW, Matzkin LM. Positive selection at sites of chemosensory genes is associated with the recent divergence and local ecological adaptation in cactophilic *Drosophila*. BMC Evol Biol. 2018;18:144.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

Engsontia P, Sangket U, Chotigeat W, Satasook C. Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. J Mol Evol. 2014;79(1–2):21–39.

Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. PLoS One. 2011;6(1):e16526.

Gilbert C, Peccoud J, Cordaux R. Transposable elements and the evolution of insects. Annu Rev Entomol. 2021;66(1):355–372.

Gouin A, Bretaudeau A, Nam K, Gimenez S, Aury J-M, Duvic B, Hilliou F, Durand N, Montagné N, Darboux I, *et al.* Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. Sci Rep. 2017;7(1):11816.

Greilhuber J, Doležel J, Lysák MA, Bennett MD. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. Ann Bot. 2005;95(1):255–260.

Gu SH, Zhou JJ, Gao S, Wang DH, Li XC, *et al.* Identification and comparative expression analysis of odorant binding protein genes in the tobacco cutworm *Spodoptera litura*. Sci Rep. 2015;515:1–17.

Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52(5):696–704.

Guo H, Cheng T, Chen Z, Jiang L, Guo Y, Liu J, Li S, Taniai K, Asaoka K, Kadono-Okuda K, *et al.* Expression map of a complete set of gustatory receptor genes in chemosensory organs of *Bombyx mori*. Insect Biochem Mol Biol. 2017;10(6):1351–1362.

Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, *et al.* Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat Ecol Evol. 2018;2(3):557–566.

Hebert FO, Freschi L, Blackburn G, Béliveau C, Dewar K, *et al.* Expansion of LINEs and species-specific DNA repeats drives genome expansion in Asian Gypsy Moths. Sci Rep. 2019;9:1–10.

Hickner PV, Rivaldi CL, Johnson CM, Siddappaji M, Raster GJ, Syed Z. The making of a pest: insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. BMC Genomics. 2016;17:648.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS: table 1. Bioinformatics. 2016; 32(5):767–769.

Hong W, Zhao H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. Proc R Soc B Biol Sci. 2014;281:20141079.

Johnston JS, Bernardini A, Hjelmen CE. Genome size estimation and quantitative cytogenetics in insects. Methods Mol Biol. 2019;1858: 15–26.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.* Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, *et al.* Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014;24(8):1384–1395.

Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780.

Keller O, Odronitz F, Stanke M, Kollmar M, Waack S. Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. BMC Bioinformatics. 2008;9:278.

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, *et al.* Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A. 2010;107(27):12168–12173.

Kochnev Y, Hellemann E, Cassidy KC, Durrant JD. Webina: an open-source library and web app that runs AutoDock Vina entirely in the web browser. Bioinformatics. 2020;36(16):4513–4515.

Koutroumpa FA, Monsempes C, François M-C, Severac D, Montagné N, Meslin C, Jacquin-Joly E. Description of chemosensory genes in unexplored tissues of the moth *Spodoptera littoralis*. Front Ecol Evol. 2021;9. doi:10.3389/fevo.2021.678277.

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017; 34(7):1812–1819.

Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE, *et al.* Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 2013; 14(8):R93.

Lefort V, Longueville J-E, Gascuel O. SMS: smart model selection in PhyML. Mol Biol Evol. 2017;34(9):2422–2424.

Li S, Zhu S, Jia Q, Yuan D, Ren C, Li K, Liu S, Cui Y, Zhao H, Cao Y, *et al.* The genomic and functional landscapes of developmental plasticity in the American cockroach. Nat Commun. 2018;9(1):1008.

Mang D, Shu M, Tanaka S, Nagata S, Takada T, Endo H, Kikuta S, Tabunoki H, Iwabuchi K, Sato R, *et al.* Expression of the fructose receptor BmGr9 and its involvement in the promotion of feeding, suggested by its co-expression with neuropeptide F1 in Bombyx mori. Insect Biochem Mol Biol. 2016;75:58–69.

Maumus F, Fiston-Lavier AS, Quesneville H. Impact of transposable elements on insect genomes and biology. Curr Opin Insect Sci. 2015;7:30–36.

McBride CS, Arguello JR, O'Meara BC. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics. 2007;177(3):1395–1416.

McKenzie SK, Kronauer DJC. The genomic architecture and molecular evolution of ant odorant receptors. Genome Res. 2018;28(11): 1757–1765.

Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005;39:121–152.

Parisot N, Vargas-Chávez C, Goubert C, Baa-Puyoulet P, Balmand S, Beranger L, Blanc C, Bonnamour A, Boulesteix M, Burlet N, *et al.* The transposable element-rich genome of the cereal pest *Sitophilus oryzae*. BMC Biol. 2021;19(1):241.

Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, Jermiin LS, McGaughran A, Oakeshott JG, Papanikolaou A, Perera OP, *et al.* Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. BMC Biol. 2017;15. doi: 10.1186/s12915-017-0402-6.

Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. Biol Rev Camb Philos Soc. 2018;93(1):184–200.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Stephen R, Niehuis O, Misof B. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol. 2019;19:11.doi: 10.1186/s12862-018-1324-9.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13): 1605–1612.

Poitout S, Bues R. Rearing of the larvae of twenty-eight species of Noctuidae and two species of Arctiidae, Lepidoptera, on a simple artificial diet differences in rearing according to the species. Ann Zool Ecol Anim. 1974;6:431–441.

Poivet E, Gallot A, Montagné N, Glaser N, Legeai F, Jacquin-Joly E. A comparison of the olfactory gene repertoires of adults and Larvae in the Noctuid Moth *Spodoptera littoralis*. PLoS One. 2013;8(4):e60263.

Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. Combined evidence annotation of transposable elements in genome sequences. PLoS Comp Biol. 2005;1(2):e22.

Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21–42. doi:10.1146/annurev-genet-110711–155621.

Robertson HM. Molecular evolution of the major arthropod chemoreceptor gene families. Annu Rev Entomol. 2019;64(1):227–242.

Sato K, Tanaka K, Touhara K. Sugar-regulated cation channel formed by an insect gustatory receptor. Proc Natl Acad Sci U S A. 2011;108(28):11680–11685.

Schrader L, Kim JW, Ence D, Zimin A, Klein A, *et al*. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun. 2014;5:1–10.

Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 2016;32(4):587–589.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19): 3210–3212.

Singh KS, Troczka BJ, Duarte A, Balabanidou V, Trissi N, Carabajal Paladino LZ, Nguyen P, Zimmer CT, Papapostolou KM, Randall E, *et al*. The genetic architecture of a host shift: an adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. Sci Adv. 2020;6(19):eaba1070.

Slater G, Birney E, Box G, Smith T, Waterman M, *et al*. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31.

Smadja C, Shi P, Butlin RK, Robertson HM. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrthosiphon pisum*. Mol Biol Evol. 2009;26(9): 2073–2086.

Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics. 2012;28(18): i409–i415.

Storey J, Bass A, Dabney A, Robinson D. qvalue: q-Value Estimation for False Discovery Rate Control. R package version 2.22.0. 2021.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(Web Server Issue):W609–W612.

Suzuki HC, Ozaki K, Makino T, Uchiyama H, Yajima S, Kawata M. Evolution of gustatory receptor gene family provides insights into adaptation to diverse host plants in nymphalid butterflies. Genome Biol Evol. 2018;10(6):1351–1362.

Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol. 2003;20(1): 18–20.

The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012;487:94–98.

The International Silkworm Genome Consortium. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochem Mol Biol. 2008;38:1036–1045.

Walker WB, Roy A, Anderson P, Schlyter F, Hansson BS, Larsson MC. Transcriptome analysis of gene families involved in chemosensory function in *Spodoptera littoralis* (Lepidoptera: Noctuidae). BMC Genomics. 2019;20(1):428.

Wang Y, Fang G, Chen X, Cao Y, Wu N, Cui Q, Zhu C, Qian L, Huang Y, Zhan S, *et al*. The genome of the black cutworm *Agrotis ipsilon*. Insect Biochem Mol Biol. 2021;139:103665.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al*. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8(12):973–982.

Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 2004;168(2):1041–1051.

Xiao J-H, Yue Z, Jia L-Y, Yang X-H, Niu L-H, Wang Z, Zhang P, Sun B-F, He S-M, Li Z, *et al*. Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. Genome Biol. 2013; 14(12):R141.

Xu W, Papanicolaou A, Zhang H-J, Anderson A. Expansion of a bitter taste receptor family in a polyphagous insect herbivore. Sci Rep. 2016;6:23666.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–1591.

Yang Z, Dos Reis M. Statistical properties of the branch-site test of positive selection. Mol Biol Evol. 2011;28(3):1217–1228.

Yang Z, Swanson WJ, Vacquier VD. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol. 2000; 17(10):1446–1455.

You M, Yue Z, He W, Yang X, Yang G, Xie M, Zhan D, Baxter SW, Vasseur L, Gurr GM, *et al*. A heterozygous moth genome provides insights into herbivory and detoxification. Nat Genet. 2013;45(2): 220–225.

Zhan S, Merlin C, Boore JL, Reppert SM. The monarch butterfly genome yields insights into long-distance migration. Cell. 2011; 147(5):1171–1185.

Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–2479.

Zhang Z-J, Zhang S-S, Niu B-L, Ji D-F, Liu X-J, Li M-W, Bai H, Palli SR, Wang C-Z, Tan A-J, *et al*. A determining factor for insect feeding preference in the silkworm, *Bombyx mori*. PLoS Biol. 2019;17(2): e3000162.

Zhang F, Zhang J, Yang Y, Wu Y. A chromosome-level genome assembly for the beet armyworm (*Spodoptera exigua*) using PacBio and Hi-C sequencing. bioRxiv 2019.12.26.889121; 2019. doi: 10.1101/2019.12.26.889121.

Zhu JY, Xu ZW, Zhang XM, Liu NY. Genome-based identification and analysis of ionotropic receptors in *Spodoptera litura*. Sci Nat. 2018;105.

Communicating editor: A. Sethuraman