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 Koutroumpa[*,**] , Emma Persyn[*,††], Christelle Monsempès[*], Marie-Christine François[*], Emmanuelle Jacquin-Joly[*]

* 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.

[†] INRAE, UMR Institut de Génétique, Environnement et Protection des Plantes (IGEPP), BioInformatics Platform for Agroecosystems Arthropods (BIPAA), Campus Beaulieu, 35042 Rennes, France.

[‡] INRIA, IRISA, GenOuest Core Facility, Campus de Beaulieu, Rennes, 35042, France.

[§] Department of Entomology, Texas A&M, University, College Station, TX 77843, USA.

** Current affiliation: INRAE, Université Tours, Infectiologie et Santé Publique (ISP), F37380 Nouzilly, France.

[††] CIRAD, UMR PVBMT, F-97410, St Pierre, Réunion, France.

Running head: *Spodoptera littoralis* genome

Key words: *Spodoptera littoralis*, gustatory receptor, transposable elements

Abstract

The bitter taste, triggered via gustatory receptors, serves as an important natural defense against the ingestion of poisonous foods in animals, and the diversity of food diet 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 receptors expansion has not been considered yet.

In this work, we hypothesized that transposable elements and possibly positive selection would be involved in the active dynamic of gustatory receptor evolution in *Spodoptera spp*. We first sequenced *de novo* the full 465Mb genome of *S. littoralis*, and manually annotated all 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 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.

Introduction

Animals rely heavily on their sense of taste to discriminate between harmful poisonous foods, usually through the detection of bitter taste, 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[1], and in insects such as the body louse[2] - an obligate ectoparasite of human -, the fig wasp *Ceratosolen solmsi*[3]– specialized on *Ficus* – and many Lepidoptera specialist feeders, although mammals and insect GRs are unrelated. Reversely, the diversity of food diet 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*[4] and herbivorous species such as noctuid species[5–7].

In polyphagous noctuids, the sequencing of the genomes of *Spodoptera frugiperda* and *S. litura* revealed GR repertoires of 231 and 237 genes[5,8], respectively, more than twice as much compared to other monophagous and oligophagous Lepidoptera species (*Bombyx mori*: 69 genes, *Heliconius melpomene*: 73 genes)[9–12], 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[10,13–15]. 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 non-random way [16]. Similar or identical TEs can induce chromosomal rearrangements such as deletions,

68 insertions and even duplications [17–19], features that are frequent in multigene families such as
69 GRs. Surprisingly, their potential role in insect GR expansion has not been considered yet.
70 In order to study in more details GRs evolution and the potential role of TEs in GRs expansion,
71 we sequenced an additional genome of a Spodoptera species: *Spodoptera littoralis.* So far, only
72 38 GRs identified[20–22] in *S. littoralis* whereas several hundreds of GRs were annotated in its
73 counterparts *S. litura* and *S. frugiperda*. To investigate this singularity, we report here the
74 sequencing of the *S. littoralis* genome, its full assembly, functional automatic annotation and
75 expert annotation of all chemosensory gene families, namely soluble carrier proteins (odorant-
76 binding proteins: OBPs, and chemosensory proteins: CSPs)[23] and the three major families of
77 insect chemosensory receptors (odorant receptors: ORs, ionotropic receptors: IRs and GRs)[24].
78 With a particular focus on gustation, we also reannotated GRs in *S. litura* and *S. frugiperda*.
79 Then, we analyzed the evolutionary history of GRs, by looking at the enrichment for
80 transposable elements in the vicinity of GRs and by analyzing selective pressures acting on the
81 different GR clades.

Methods and Materials

Estimation of *Spodoptera littoralis* genome size

The genome size of *Spodoptera littoralis* was estimated using flow cytometry. Genome size estimates were produced as described before[25]. In brief, the head of a *S. littoralis* along with the head of a female *Drosophila virilis* standard (1C = 328 Mbp) 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 hours in the cold and dark. 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) X 328 Mbp.

Spodoptera littoralis genome sequencing and assembly

Biological material and genomic DNA extraction

Whole genomic DNA was extracted from two male larvae obtained after two inbred generations resulting from a single pair of *S. littoralis* originating from a laboratory colony maintained in INRAE Versailles since 2000s on suitable laboratory diet (Poitout and Bues 1974). 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 was obtained.

Sequencing

Different types of libraries were generated for two sequencing technologies: Illumina and PacBio. For Illumina sequencing, five libraries were prepared and constructed according to the Illumina manufacturer's protocol (one library of 170, one of 250 and three of 500 bp). Illumina sequencing was performed at the BGI-tech facilities (Shenzen, China) on a HiSeq2500 machine. Around 68 Gb were obtained, representing 144X of the estimated genome size (470 Mb) (Supp 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 34X of the estimated genome size (Supp data 1). High quality sequences were obtained by generating circular consensus sequencing (CCS).

Genome assembly

A first assembly was done using Platanus (v1.2.1)[26] with Illumina data. A second assembly was obtained by doing scaffolding with SSPACE-LR (modified)[27] using PacBio data and gap filling using GapCloser[28]. These second assembly was evaluated using Benchmarking Universal Single-Copy Orthologue (BUSCO v3.0.2)[29] with a reference set of 1,658 genes conserved in Insecta.

Structural and functional genome annotation

Structural automatic genome annotation was done with BRAKER (v1.11) [30] using all RNAseq data described in Supp 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 (Supp data 1)[20–22]. Reads were trimmed using Trimmomatic (v0.36)[31] 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)[32] with the default parameters except for the following parameters : outFilterMultimapNmax = 5, outFilterMismatchNmax = 3, alignIntronMin = 10, alignIntronMax = 50000 and alignMatesGapMax = 50000. As done for the genome assembly, gene annotation was evaluated using Benchmarking Universal Single-Copy Orthologue (BUSCO v3.0.2) [29] with a reference set of 1,658 proteins conserved in Insecta. Putative

functions of predicted proteins were assigned using blastp (v2.6.0) against GenBank NR (non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF) release 09/2017, 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), odorant receptors (ORs) and ionotropic receptors (IRs) was performed using known sequences from other species with their genome sequenced (*S. frugiperda*, *S. litura*, *B. mori*, *H. melpomene* and *Danaus plexippus*)[8,10,11,33,34]. 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[35]. 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[36] and Exonerate[37] to define intron/exon boundaries and to create gene models. Outputs from Scipio and Exonerate were then visualized on a Apollo browser[38] 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*[20,22,39]. The classification of deduced proteins and their integrity were verified using blastp against the non-redundant (NR) GenBank database. When genes were suspected to be split on different scaffolds, protein sequences were merged for further analyses. OBPs were also annotated in the *S. litura* genome, using a similar procedure. For OBPs and CSPs, SignalP-5.0[40] was used to determine the presence or absence of a signal peptide. Hereafter, the

157 abbreviations Slit, Slitu and Sfru (for *S. littoralis S. litura* and *S. frugiperda*, respectively) are
158 used before gene names to clarify the species.
159
160 **Iterative annotation and re-annotation of gustatory receptors (GRs)**
161 The initial annotation of gustatory receptor genes was carried out the same way as for the other
162 genes involved in chemosensation. Subsequent steps were then added to annotate the full
163 repertoire of GRs. At the end of the manual curation, all the newly identified amino acid GR
164 sequences were added to the set of known GR sequences to perform a new cycle of annotation.
165 This iterative strategy was used for *S. littoralis* as well as for *S. litura* and *S. frugiperda* and
166 was performed until no new GR sequence was identified.
167 At the end of the annotation, all GR amino acid sequences were aligned for each species
168 individually using MAFFT v7.0[41] in order to identify and filter allelic sequences. Between
169 alleles, only the longest sequence was retained for further analysis. Pseudogenes were identified
170 as partial sequences containing one or multiple stop codons. Genes were considered complete
171 when both following conditions were met: 1) a start and a stop codon were identified and 2) a
172 sequence length >350 amino acids. *S. littoralis* gene names were attributed based on orthology
173 relationships with *S. frugiperda* when possible. *S. frugiperda* newly identified genes compared
174 to the previous publications were numbered starting from SfruGR232. *S. litura* newly identified
175 gene names were numbered starting from SlituGR240.
176
177 **Annotation and enrichment analysis of transposable elements around chemosensory**
178 **receptor genes in *Spodoptera* species**
179 The annotation of transposable elements (TEs) in *S. littoralis* genome was performed using
180 REPET (Galaxy Lite v2.5). The TEdenovo pipeline[42] was used to identify consensus sequences
181 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[43] was then used to annotate all repetitive elements in the genome using the library of TE consensus and to build a non-redundant library in which redundant consensus were eliminated (length >=98%, identity >=95%). The non-redundant library of TEs was finally used to perform the *S. littoralis* genome annotation with the TEannot pipeline.

The tool LOLA (Locus Overlap Analysis) within the R package Bioconductor[44] was used to test for enrichment of TEs within the genomic regions containing chemosensory receptor genes (ORs and GRs) in both *S. littoralis* and *frugiperda*. To run LOLA with data from *S. littoralis*, 3 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, this dataset of the genome leaded to the creation of 114 chemosensory regions for the GRs and 63 regions for the ORs. The second dataset, the region universe, contained 1000 random regions of similar sizes selected from the genome. 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 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[5], as well as chemosensory receptor re-annotations from the present work and leaded 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 data set contained GRs from polyphagous (*S. littoralis*, *S. litura*, *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 was performed with ClustalO[45] and the phylogeny was reconstructed using PhyML 3.0[46] (http://www.atgc-montpellier.fr/phyml/) with the automatic selection of the best substitution model by SMS[47]. The resulting phylogenetic tree was edited using FigTree v1.4.2 (https://github.com/rambaut/figtree) and Inkscape 0.92 (https://inkscape.org/fr/). Branch supports were estimated using the approximate likelihood-ratio test (aLRT)[48] 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[41]. The reconstruction of the phylogenetic trees was carried out the same way as for the GRs.

Tree reconciliation

Estimates of gains and losses of GR genes across the Noctuidae were inferred using the reconciliation methods implemented in Notung v2.6[49,50]. The species tree was generated using TimeTree.org[51] 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[52]. 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

232 substitutions. For each subtree, a codon alignment was performed using protein sequence
233 alignments performed using MAFFT and PAL2NAL[53] in order to convert the amino acid
234 alignment to a codon alignment, and a phylogenetic tree was reconstructed based on this
235 alignment. Sequences introducing large gaps in the alignment were removed in order to
236 compute codeml on the largest alignment possible. To estimate the selective pressures acting
237 on the evolution of the lepidopteran GR genes, the "m0 model" from codeml of the PAML
238 package was computed on the 13 subtrees to estimate the global ω (ratio of non-synonymous
239 substitutions dN/ratio of synonymous substitutions dS)[54]. The ω value reflects the mode of
240 evolution, with ω>1 indicating positive selection, ω<1 indicating purifying selection and ω=1
241 indicating neutral evolution. To further infer positive selection, two comparisons between
242 evolutionary models were conducted. First, the comparison between M8 and M8a models can
243 detect positive selection acting on sites, i.e. columns of the alignment[55,56]. This comparison was
244 conducted only when the global ω calculated from the m0 model was > 0.3. The second
245 comparison between branch-site model A and its neutral counterpart can detect positive
246 selection acting on particular sites on a specific lineage[57]. Here, we tested all the terminal
247 branches of the trees for which both the global ω was elevated and the comparison between
248 models M8 and M8a statistically significant. Since many branches were tested for each tree, a
249 correction for multiple testing to control for false discovery rate was applied: the q-value
250 (Storey and al, R package version 2.22.0)[58]. In the case of a statistically significant q-value
251 (<0.05), positively selected sites were inspected for possible artifacts due to partial sequences
252 or misalignment.
253
254 ***Putative functional assignation***
255 In order to assign putative functions to several candidate SlitGRs, both their phylogenetic
256 position and theoretical 3D structure were analyzed. For the theoretical structures, the

257 AlphaFold algorithm[59] was used to model candidate SlitGRs as well as their *B. mori* ortholog
258 GRs with known function: BmorGR9 and BmorGR66. Structures were then compared between
259 orthologs using the MatchMaker tool of Chimera and the RMSD (Root Mean Square Deviation)
260 computed using the same tool[60].

Results and discussion

Genome assembly and automatic annotation of the *Spodoptera 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 95% of complete BUSCO genes, with almost 94% 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 BUSCO analysis indicated that almost 97% of BUSCO proteins were complete, with more than 88% being present in single-copy (Table 2). These data showed the good quality of the *S. littoralis* genome assembly, 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 equipment, 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[23], and the membrane bound receptors (ORs: seven transmembrane proteins expressed in the membrane of olfactory sensory neurons, GRs: seven transmembrane proteins hosted by taste neurons, and IRs: three transmembrane proteins sensing acids and amines [61–63]).

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.*

286 *frugiperda*[33] and the 23 CSP genes annotated in *S. litura*[8]. Among all these sequences, 16 CSP
287 genes are 1:1 orthologs between the three *Spodoptera* species included in the tree while 11 CSP
288 genes are 1:1 orthologs with BmorCSPs (from *B. mori*), showing the high level of conservation
289 in this gene family (Supp data 2 Figure S1).
290 We also annotated 53 OBP genes in *S. littoralis.* Among these genes, 49 were complete and 48
291 possessed a signal peptide (Supp data 2). The phylogenetic tree revealed a clade enriched in
292 *Spodoptera* OBPs (9 SlitOBPs, 9 SlituOBPs and 10 SfruOBPs) (Figure S2). This expansion
293 probably arose from recent tandem duplications as most of the genes of the expansion are
294 organized in synteny in the three species (Figure S3).
295 We annotated 44 IR genes in the *S. littoralis* genome, 43 of which encoding a full-length
296 sequence with various sizes containing 547 to 948 amino acids (AAs) (Supp data 2). In addition
297 to the two conserved co-receptors IR8a and IR25a[64], we identified 18 candidate antennal IRs
298 putatively involved in odorant detection, 23 divergent IRs putatively involved in taste and 12
299 ionotropic glutamate receptors (iGluRs). The total IR number was similar to the 44 IR genes
300 annotated in *S. litura*[65] and the 43 IR genes annotated in *S. frugiperda*[5]. Among all these
301 sequences, 43 IR genes are 1:1 orthologs between the three *Spodoptera* species (IR100g was
302 missing in *S. frugiperda*). The phylogenetic tree revealed a clade containing divergent IRs and
303 two lineage expansions were observed (IR7d and IR100), likely attributed to gene
304 duplications[65]. The number of divergent IRs was much higher in *Spodoptera* species (*S.*
305 *littoralis*: 26, *S. litura*: 26, *S. frugiperda*: 25) than in *H. melpomene* (16) and *B. mori* (6). By
306 contrast, phylogenetic analysis (Figure S4) reveals that *S. littoralis* antennal IRs retained a
307 single copy within each orthologous group.
308 We annotated 73 OR genes in the *S. littoralis* genome scattered among 61 scaffolds (Supp data
309 2), including the obligatory co-receptor ORco (ref Jones et al 2005 Curr Biol). The number of
310 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 Figure S5.

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*[66], *S. litura*[6], *S. frugiperda*[5] and *Agrotis ipsilon*[67] revealed an important expansion of gustatory receptors 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* (Supp data 3). Our GR analysis not only revealed that the full repertoire of *S. littoralis* GRs is in fact much more important 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 to 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. Indeed, the highest number of alleles was reached in *S. frugiperda*, a genome with a high level of heterozygosity[33], while alleles were less frequent in the two other *Spodoptera* genomes considered. When omitting pseudogenes and alleles, the final repertoires of GRs are composed of 325 genes in *S.*

littoralis, 274 GRs in *S. frugiperda* and 280 GRs in *S. litura*. As previously shown, multiple clusters of GRs were also found in the *S. littoralis* genome. The two main clusters were found on scaffolds 1414 and 878 that contained each 27 GR genes. The phylogeny reconstructed using the GR sequences from the three *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, eleven clades were enriched in *Spodoptera* genes (numbered from A to K in Figure 1) and encompassed the majority of the three *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 one exon were identified, clustered on scaffold 67 and belonging to the same phylogenetic clade (Figure 1). Interestingly, this clade was very conserved with a 1:1 orthology relationship between the three *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[11]. BmorGR53 is able to detect the bitter tastant and feeding deterrent coumarine [68]. It is then likely 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 (Figure 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

361 genes such as *B. mori* (70 GRs) and *H. melpomene* (73 GRs), the turnover of genes compared
362 to the ancestors is important (33 and 41 gene gains, 25 and 26 gene losses, respectively).
363
364 **Annotation of transposable elements, enrichment analysis and selection pressure**
365 To get more insights about the mechanisms that led to the formation of massive genomic
366 clusters of GR genes, we looked at 1) whether TEs could be involved and 2) the selective
367 pressures acting on GR genes.
368 TEs have been shown to be involved in countless mechanisms of evolution in insects, such as
369 insecticide resistance, the evolution of regulatory networks, immunity, climate adaptation[16,69–
370 74] and some of them have even been domesticated as genes[75]. Gene families involved in these
371 traits have been shown to be enriched in TEs and gene family expansions have been correlated
372 with TE content, for instance in termites[76]. Interestingly, enrichment in TEs has not been
373 reported for insect GR gene clusters so far. While annotating GRs in the *S. littoralis* genome,
374 we noticed the frequent co-occurrence of TEs on the same scaffolds. We thus annotated TEs in
375 the *S. littoralis* genome and calculated their enrichment in the vicinity of GR genes. We also
376 carried out the same enrichment analysis in *S. frugiperda* genomes, as TE annotation in this last
377 species has been done using the same REPET pipeline as in our study. The *de novo* constructed
378 library contained 1089 consensus sequences of TEs and was used to annotate the *S. littoralis*
379 genome. The repeat coverage for the *S. littoralis* genome was 30.22%, representing 140 Mb,
380 which is similar to that of *S. frugiperda* (29.10%), *S. litura* (31.8%) and *S. exigua* (33.12%)[5,6,77].
381 The relative contribution of the different classes of repetitive elements revealed that Class I
382 elements were more represented than Class II elements (66.96% vs 20.83%), a classical feature
383 of insect genomes[75] (Figure 3, Table 5). However, the repartition and proportions between the
384 different classes differed between these species. The Class I SINE was the most represented in
385 *S. frugiperda* (12.52%)[33] and one of them was found to be enriched in the vicinity of the GRs

386 in both *S. littoralis* and *S. frugiperda* while the class I LINE elements were the most represented
387 in both *S. litura* and *S. exigua*, although with a lower proportion of all repeated elements
388 (27.73% and 14.81%, respectively). Remarkably, the proportion of LINE elements identified
389 in the *S. littoralis* genome was the highest reported so far in arthropods[78], accounting for 52.18%
390 of all repetitive elements. In two subspecies of the Asian gypsy moth *Lymantria dispar*, the
391 accumulation of this particular class of transposable elements was found to be responsible for
392 their large genome size[79], a phenomenon also observed in other insect species[75]. The
393 accumulation of the same elements in the *S. littoralis* genome could explain its larger size
394 compared to its *Spodoptera* counterparts (470Mb vs ~400Mb for *S. frugiperda*, 438Mb for *S.*
395 *litura*, 408-448Mb for *S. exigua*). The second most represented was DNA transposons, Class II
396 TIR elements, representing 11.04% of all TEs (Figure 3, Table 5).
397 The enrichment of TEs in the vicinity of GR gene clusters was tested in both *Spodoptera*
398 *littoralis* and *frugiperda*, and we found that a particular category of class I retrotransposons, a
399 SINE transposon, was significantly enriched (q-value < 0.005) in the vicinity of the GRs in both
400 species (q-value = 0.0043 and 0.0078, respectively), suggesting a transposon-mediated
401 mechanism for the formation of GR clusters (Supp data 4). SINE elements are typically small
402 (80-500 bp) and originate from accidental retrotransposition of various polymerase III
403 transcripts. These elements are non-autonomous, therefore their involvement in the dynamic of
404 the GR multigene family may be related to their potential to induce genome rearrangements via
405 unequal crossing over, hence potential drivers of duplication, as previously shown in other
406 insect species [80–83]. Given their prevalence in the *S. littoralis* genome, the potential role of these
407 TEs in the GR family dynamic is probably just one of their numerous functions. The same
408 enrichment analysis performed for the OR loci showed no significant enrichment in *S. littoralis*
409 but did show enrichment in an uncharacterized class of transposons in the vicinity of SfruORs
410 (Supp data 4).

411

412 Several studies have shown the importance of positive selection in the evolution of multigene
413 families, especially in chemosensory genes such as ORs and GRs[84,85]. Positively selected
414 chemoreceptors may be linked to adaptation in *Drosophila* species[86,87]. In the pea aphid,
415 signatures of selection have been identified in chemosensory genes, including GRs and ORs,
416 which may be implicated in the divergence of pea aphid host races [88–90]. We thus analyzed
417 selective pressures focusing on 13 clades of interest in the *Spodoptera* GR phylogeny: the
418 potential clade of CO_2 receptors, the potential clade of sugar receptors and the 11 expended
419 lineages within the so-called bitter receptor clades. For all 13 clades, we observed low global
420 ω values ranging from 0.01 to 0.42, with the highest observed for candidate bitter receptor
421 clades. The comparison between models M8 and M8a was statistically significant for clades C,
422 F and J, indicating a signal of positive selection. Branch-site models on terminal branches of
423 the associated trees were then tested on these clades. For clade J, no GR was revealed as
424 evolving under positive selection. However, for clades C and F, 2 and 5 GRs were identified as
425 positively selected, respectively (Table 6). Within these GR sequences, very few positively
426 selected sites were identified for each gene (between 0 and 3, Supp data 5). This finding is
427 coherent with previous studies showing the same pattern of evolutionary rates[91,92], especially
428 in *S. frugiperda*[5] (3 GRs under positive selection when comparing two host strains). Taken
429 together, all positive selection analyses indicate that positive selection within the GR gene
430 family is cryptic and may not play an important role in shaping the evolution of *Spodoptera*
431 GRs. Anyhow, the few positively selected GRs may be interesting candidates for further
432 functional studies.

433

434 **Putative functional assignation of candidate SlitGRs**

435 The complexity of the evolution of the bitter GRs is reflected by their complex functioning.
436 Indeed, in contrast with the relatively simple OR/Orco association that is the basis for olfaction,
437 the molecular basis for gustation is marked by several characteristics that were recently
438 identified in *D. melanogaster*. First, some GRs have to be co-expressed within the same neuron
439 in order to be able to respond to a stimulus. Second, it seems that GR-GR inhibition can
440 modulate neuron responses [93]. The challenge in the next few years will be to characterize both
441 the response spectra and precise expression patterns of GRs of interest. However, those GRs of
442 interest need to be selected. The present work provides us with some valuable candidates such
443 as the single exon GRs for which the function is known in *B. mori*. Also, it seems that individual
444 GRs can play an important role in the ecology of a species. Among examples are BmorGR9,
445 which binds D-fructose without the need of any other GR[94,95], and BmorGR66, whose silencing
446 confers to the monophagous *B. mori* larva the ability to feed from different food sources[96]. We
447 identified their *S. littoralis* orthologs as SlitGR9 and SlitGR15, respectively. We predicted their
448 3D structures using AlphaFold and compared them to the AlphaFold predicted structures of *B.*
449 *mori* orthologues. The RMSD computed between the whole 3D structures of BmorGR9 and
450 SlitGR9 was 10.855Å (Figure 4A), but when the N-terminal end, as well as the loop between
451 the transmembrane domains 4 and 5, were removed (regions that were disordered and difficult
452 to predict), the RMSD improved to 1.170Å (Figure 4B). Both structures were strikingly similar
453 on their extracellular side, suggesting that SlitGR9 is likely a D-fructose receptor in *S. littoralis*.
454 The ligand of BmorGR66 is not known, however, this receptor is responsible for the feeding
455 difference of *B. mori* for mulberry leaves[96]. Its ortholog SlitGR15 is a key candidate for
456 functional studies to test if this GR has an impact on the feeding preference in *S. littoralis* as
457 well. When comparing both full structures, the RMSD was 6.518Å (Figure 4C) while it
458 decreased to 3.599Å when the N-terminal of both structures were removed from the analysis

(Figure 4D). Interestingly, differences were visible between both structures 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*[8,20,21]. 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.

Tables

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

	Slit genome v1.0	Slit genome v2.0
Number of scaffolds	123499	28891
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)	1584 (95.5%)	1603 (96.68%)
Complete and single-copy BUSCOs (S)	1558 (94%)	1467 (88.48%)
Complete and duplicated BUSCOs (D)	26 (1.6%)	136 (8.2%)
Fragmented BUSCOs (F)	45 (2.7%)	48 (2.89%)
Missing BUSCOs (M)	29 (1.7%)	7 (0.42%)
Total BUSCO groups searched	1658	1658

Table 3. GR repertoires of *Spodoptera* species.

	S. littoralis	*S. frugiperda*	*S. 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

Table 4. Number of *Spodoptera* GRs by expansion clade. The percentages represent the proportion of *Spodoptera* genes to the total number of GRs annotated in the three *Spodoptera* species (complete + partial genes indicated in Table 3).

Clade	***S. littoralis***	***S. frugiperda***	***S. 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%)

484 **Table 5. Repartition of repetitive elements in *S. littoralis* genome based on the**
485 **classification established by Wicker et al.** [97]. noCat means repetitive elements that could not
486 be classified into the existing categories.

	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%

487

488 **Table 6. Selective pressure analysis.** NS: non significant, /: not calculated.

Clade	**# 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	/	/
CO2	11	0.01074	/	/

489

Figure legends

Figure 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 blue and yellow, respectively. All the other GRs are part of the bitter receptor clades. The star indicates the clade of single-exon GRs. Midpoint rooting was used. Circles indicate nodes strongly supported by the likelihood-ratio test (aLRT>0.9). The scale bar represents 0.5 amino acid substitutions per site.

Figure 2 GR gain and loss estimates across lepidopterans

The gene-tree of GRs generated using PhyML was reconciled with the species-tree using Notung[50] 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.

Figure 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.

Figure 4 3D structures of BmorGR9 and BmorGR66 and their respective orthologs, SlitGR9 and SlitGR15

3D structures were predicted using AlphaFold2[59]. **A** and **C**. Alignment of full 3D structures. In **B** and **D**, the disordered regions such as the large extracellular loops (**B**) and the N-terminal ends (**B** and **D**) were removed for the comparison of the orthologs' structures.

Supplementary figures

Figure S1 Phylogeny of lepidopteran CSPs

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 MAFFT and the phylogenetic tree was reconstructed using PhyML. Midpoint rooting was used. Circles indicate nodes strongly supported by the likelihood-ratio test (aLRT>0.9). The scale bar represents 0.5 amino acid substitutions per site.

Figure S2 Phylogeny of lepidopteran OBPs

The dataset included 53 amino acid sequences from *S. littoralis* (Noctuoidea, red), 53 sequences from *S. litura* (Noctuoidea, green), 53 sequences from *S. frugiperda* (Noctuoidea, orange), 44 sequences from *B. mori* (Bombycoidea, blue) and 43 sequences from *H. melpomene* (Papilionoidea, cyan). Sequences were aligned using MAFFT and the phylogenetic tree was reconstructed using PhyML. Subfamilies are indicated with different colors (yellow: Minus C subfamily, green: Plus-C subfamily, blue: PBP-GOBP subfamily). Midpoint rooting was used. Circles indicate nodes strongly supported by the likelihood-ratio test (aLRT>0.9). The red star indicates expansion in *Spodoptera*. The scale bar represents 0.5 amino acid substitutions per site.

Figure S3 Genomic organization of the *Spodoptera* OBP genes. Scaffolds/chromosomes are represented in gray, with their numbers in italic. Gene names are indicated and their orientations are represented by the arrows.

Figure S4 Phylogeny of lepidopteran IRs

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 MAFFT and the phylogenetic tree was reconstructed using PHYML. Colors indicate different categories of IRs (yellow: divergent IRs, grey: ionotropic glutamate receptors, orange: IR25a coreceptorsThe tree was rooted using the iGluR clade. Circles indicate basal nodes strongly supported by the likelihood-ratio test (aLRT>0.9). The scale bar represents 0.5 amino acid substitutions per site.

Figure S5 Phylogeny of lepidopteran ORs

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 MAFFT and the phylogenetic tree was reconstructed using PHYML. The tree was rooted using the Orco clade (purple). Circles indicate basal nodes strongly supported by the likelihood-ratio test (aLRT>0.9). The scale bar represents 0.5 amino acid substitutions per site.

Data Availability Statement

The assembled genome of *Spodoptera littoralis* as well as the genomic data of *Spodoptera litura* (v1.0)[8] and *Spodoptera frugiperda* (Corn variant, v3.1)[33] are all publicly available on the BIPAA platform (https://bipaa.genouest.org) and on NCBI (XXXX).

Acknowledgments

The *S. littoralis* genome has been sequenced in the framework of the i5K initiative (http://i5k.github.io/) and the InsectGenomes projectprojet at INRAE leaded by Denis Tagu (IGEPP).

568

569 ***Conflict of Interest***

570 The authors declare that there is no conflict of interest

571

572 ***Funder Information***

573 This work has been funded by INRAE (France), the French National Research Agency (ANR-
574 16-CE21-0002-01, and ANR-16-CE02-0003) and by the European Union's Horizon 2020
575 research and innovation program under the Marie Skłodowska-Curie grant agreement no.
576 764840 for the ITN IGNITE project.

577

578 **Literature Cited**

579 1. Hong W, Zhao H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes
580 common to other bats. Proceedings of the Royal Society B: Biological Sciences. 2014
581 [accessed 2021 Nov 19];281(1788).
582 https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2014.1079.
583 doi:10.1098/RSPB.2014.1079
584 2. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM,
585 Kennedy RC, Elhaik E, et al. Genome sequences of the human body louse and its primary
586 endosymbiont provide insights into the permanent parasitic lifestyle. Proceedings of the
587 National Academy of Sciences of the United States of America. 2010 [accessed 2018 Jul
588 20];107(27):12168–73. http://www.ncbi.nlm.nih.gov/pubmed/20566863.
589 doi:10.1073/pnas.1003379107
590 3. Xiao JH, Yue Z, Jia LY, Yang XH, Niu LH, Wang Z, Zhang P, Sun BF, He SM, Li Z, et al.
591 Obligate mutualism within a host drives the extreme specialization of a fig wasp genome.
592 Genome biology. 2013 [accessed 2021 Nov 19];14(12).

593 https://pubmed.ncbi.nlm.nih.gov/24359812/. doi:10.1186/GB-2013-14-12-R141

594 4. Li S, Zhu S, Jia Q, Yuan D, Ren C, Li K, Liu S, Cui Y, Zhao H, Cao Y, et al. The genomic
595 and functional landscapes of developmental plasticity in the American cockroach. Nature
596 Communications. 2018 [accessed 2018 Mar 27];9(1):1008.
597 http://www.nature.com/articles/s41467-018-03281-1. doi:10.1038/s41467-018-03281-1

598 5. Gouin A, Bretaudeau A, Nam K, Gimenez S, Aury J-M, Duvic B, Hilliou F, Durand N,
599 Montagné N, Darboux I, et al. Two genomes of highly polyphagous lepidopteran pests
600 (Spodoptera frugiperda, Noctuidae) with different host-plant ranges. Scientific Reports. 2017
601 [accessed 2017 Oct 2];7(1):11816. http://www.nature.com/articles/s41598-017-10461-4.
602 doi:10.1038/s41598-017-10461-4

603 6. Cheng T, Wu J, Wu Y, Chilukuri R V., Huang L, Yamamoto K, Feng L, Li W, Chen Z,
604 Guo H, et al. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid
605 pest. Nature Ecology and Evolution. 2017;1(11):1747–1756.
606 http://dx.doi.org/10.1038/s41559-017-0314-4. doi:10.1038/s41559-017-0314-4

607 7. Xu W, Papanicolaou A, Zhang H-J, Anderson A. Expansion of a bitter taste receptor family
608 in a polyphagous insect herbivore. Scientific Reports. 2016;6:23666.
609 http://www.nature.com/articles/srep23666. doi:10.1038/srep23666

610 8. Cheng T, Wu J, Wu Y, Chilukuri R V., Huang L, Yamamoto K, Feng L, Li W, Chen Z,
611 Guo H, et al. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid
612 pest. Nature Ecology & Evolution. 2017;1(November).
613 http://www.nature.com/articles/s41559-017-0314-4. doi:10.1038/s41559-017-0314-4

614 9. The International Silkworm Genome Consortium. The genome of a lepidopteran model
615 insect, the silkworm Bombyx mori. Insect Biochemistry and Molecular Biology.
616 2008;38(12):1036–1045.
617 https://www.sciencedirect.com/science/article/pii/S0965174808002099.

618 doi:https://doi.org/10.1016/j.ibmb.2008.11.004

619 10. Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, Jamie GA, Martin SH,

620 Dasmahapatra KK, Ferguson LC, Mallet J, et al. Female Behaviour Drives Expression and

621 Evolution of Gustatory Receptors in Butterflies. PLoS Genetics. 2013;9(7).

622 doi:10.1371/journal.pgen.1003620

623 11. Guo H, Cheng T, Chen Z, Jiang L, Guo Y, Liu J, Li S, Taniai K, Asaoka K, Kadono-

624 Okuda K, et al. Expression map of a complete set of gustatory receptor genes in

625 chemosensory organs of Bombyx mori. Insect Biochemistry and Molecular Biology. 2017.

626 doi:10.1016/j.ibmb.2017.02.001

627 12. Consortium THG. Butterfly genome reveals promiscuous exchange of mimicry

628 adaptations among species. Nature. 2012;487(7405):94–98. internal-

629 pdf://228.60.152.105/2012-Butterfly genome reveals promiscuous exch.pdf.

630 doi:10.1038/nature11041

631 13. Suzuki HC, Ozaki K, Makino T, Uchiyama H, Yajima S, Kawata M. Evolution of

632 gustatory receptor gene family provides insights into adaptation to diverse host plants in

633 nymphalid butterflies. Genome Biology and Evolution. 2018 May 18 [accessed 2018 May

634 18]. https://academic.oup.com/gbe/advance-article/doi/10.1093/gbe/evy093/4999382.

635 doi:10.1093/gbe/evy093

636 14. Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. Family Size Evolution in

637 Drosophila Chemosensory Gene Families: A Comparative Analysis with a Critical Appraisal

638 of Methods. Genome Biology and Evolution. 2014 [accessed 2018 Jan 8];6(7):1669–1682.

639 https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evu130.

640 doi:10.1093/gbe/evu130

641 15. Nei M, Rooney AP. Concerted and Birth-and-Death Evolution of Multigene Families.

642 Annual Review of Genetics. 2005;39(1):121–152.

643 doi:10.1146/annurev.genet.39.073003.112240

644 16. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M,
645 Izsvák Z, Levin HL, Macfarlan TS, et al. Ten things you should know about transposable
646 elements. Genome biology. 2018;19(1):199. doi:10.1186/s13059-018-1577-z

647 17. Fiston-Lavier AS, Anxolabehere D, Quesneville H. A model of segmental duplication
648 formation in Drosophila melanogaster. Genome Research. 2007;17(10):1458–1470.
649 doi:10.1101/gr.6208307

650 18. Gray YHM. It takes two transposons to tango: Transposable-element-mediated
651 chromosomal rearrangements. Trends in Genetics. 2000;16(10):461–468. doi:10.1016/S0168-
652 9525(00)02104-1

653 19. Lim JK. Intrachromosomal rearrangements mediated by hobo transposons in Drosophila
654 melanogaster. Proceedings of the National Academy of Sciences of the United States of
655 America. 1988;85(23):9153–9157. doi:10.1073/pnas.85.23.9153

656 20. Koutroumpa FA, Monsempes C, François M-C, Severac D, Montagné N, Meslin C,
657 Jacquin-Joly E. Description of Chemosensory Genes in Unexplored Tissues of the Moth
658 Spodoptera littoralis. Frontiers in Ecology and Evolution. 2021;0:391.
659 doi:10.3389/FEVO.2021.678277

660 21. Walker WB, Roy A, Anderson P, Schlyter F, Hansson BS, Larsson MC. Transcriptome
661 Analysis of Gene Families Involved in Chemosensory Function in Spodoptera littoralis
662 (Lepidoptera: Noctuidae). BMC Genomics. 2019 [accessed 2019 Dec 5];20(1):428.
663 https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-5815-x.
664 doi:10.1186/s12864-019-5815-x

665 22. Poivet E, Gallot A, Montagné N, Glaser N, Legeai F, Jacquin-Joly E. A Comparison of
666 the Olfactory Gene Repertoires of Adults and Larvae in the Noctuid Moth Spodoptera
667 littoralis. PLoS ONE. 2013;8(4). doi:10.1371/journal.pone.0060263

668 23. Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: diverse tasks of
669 soluble olfactory proteins in insects. Biological Reviews. 2018 [accessed 2017 May
670 15];93(1):184–200. http://doi.wiley.com/10.1111/brv.12339. doi:10.1111/brv.12339
671 24. Robertson HM. Molecular Evolution of the Major Arthropod Chemoreceptor Gene
672 Families. Annual Review of Entomology. 2019 [accessed 2018 Oct 16];64(1):annurev-ento-
673 020117-043322. https://www.annualreviews.org/doi/10.1146/annurev-ento-020117-043322.
674 doi:10.1146/annurev-ento-020117-043322
675 25. Johnston JS, Bernardini A, Hjelmen CE. Genome Size Estimation and Quantitative
676 Cytogenetics in Insects. Methods in molecular biology (Clifton, N.J.). 2019 [accessed 2021
677 Nov 19];1858:15–26. https://pubmed.ncbi.nlm.nih.gov/30414107/. doi:10.1007/978-1-4939-
678 8775-7_2
679 26. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada
680 M, Nagayasu E, Maruyama H, et al. Efficient de novo assembly of highly heterozygous
681 genomes from whole-genome shotgun short reads. Genome Research. 2014 [accessed 2019
682 Jul 10];24(8):1384–1395. https://genome.cshlp.org/content/24/8/1384.abstract.
683 doi:10.1101/GR.170720.113
684 27. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using
685 long read sequence information. BMC Bioinformatics. 2014 [accessed 2019 Jun
686 26];15(1):211. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-
687 211. doi:10.1186/1471-2105-15-211
688 28. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome
689 Biology. 2012 [accessed 2019 Jul 10];13(6):R56.
690 http://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-6-r56. doi:10.1186/gb-
691 2012-13-6-r56
692 29. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO:

693 assessing genome assembly and annotation completeness with single-copy orthologs.
694 Bioinformatics. 2015 [accessed 2017 Dec 5];31(19):3210–3212.
695 https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv351.
696 doi:10.1093/bioinformatics/btv351
697 30. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised
698 RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS: Table 1.
699 Bioinformatics. 2016 [accessed 2019 Jul 10];32(5):767–769.
700 http://www.ncbi.nlm.nih.gov/pubmed/26559507. doi:10.1093/bioinformatics/btv661
701 31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
702 data. Bioinformatics. 2014 [accessed 2019 Jul 10];30(15):2114–2120.
703 http://www.ncbi.nlm.nih.gov/pubmed/24695404. doi:10.1093/bioinformatics/btu170
704 32. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
705 Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 [accessed
706 2019 Jul 10];29(1):15–21. http://www.ncbi.nlm.nih.gov/pubmed/23104886.
707 doi:10.1093/bioinformatics/bts635
708 33. Gouin A, Bretaudeau A, Nam K, Gimenez S, Aury JM, Duvic B, Hilliou F, Durand N,
709 Montagné N, Darboux I, et al. Two genomes of highly polyphagous lepidopteran pests
710 (Spodoptera frugiperda, Noctuidae) with different host-plant ranges. Scientific Reports.
711 2017;7(1):1–12. doi:10.1038/s41598-017-10461-4
712 34. Zhan S, Merlin C, Boore JL, Reppert SM. The monarch butterfly genome yields insights
713 into long-distance migration. Cell. 2011;147(5):1171–1185. internal-
714 pdf://153.165.37.102/Zhan-2011-The monarch butterfly genome yields.pdf.
715 doi:10.1016/j.cell.2011.09.052
716 35. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
717 BLAST+: architecture and applications. BMC Bioinformatics. 2009 [accessed 2019 Jun

718 26];10(1):421. http://www.biomedcentral.com/1471-2105/10/421. doi:10.1186/1471-2105-
719 10-421
720 36. Keller O, Odronitz F, Stanke M, Kollmar M, Waack S. Scipio: Using protein sequences to
721 determine the precise exon/intron structures of genes and their orthologs in closely related
722 species. BMC Bioinformatics. 2008 [accessed 2018 Jan 16];9(1):278.
723 http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-278.
724 doi:10.1186/1471-2105-9-278
725 37. Slater G, Birney E, Box G, Smith T, Waterman M, Altschul S, Gish W, Miller W, Myers
726 E, Lipman D, et al. Automated generation of heuristics for biological sequence comparison.
727 BMC Bioinformatics. 2005;6(1):31.
728 http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-6-31.
729 doi:10.1186/1471-2105-6-31
730 38. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes
731 IH, Elsik CG, Lewis SE. Web Apollo: a web-based genomic annotation editing platform.
732 Genome Biology. 2013 [accessed 2018 Jan 16];14(8):R93.
733 http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-8-r93. doi:10.1186/gb-
734 2013-14-8-r93
735 39. Walker WB, Roy A, Anderson P, Schlyter F, Hansson BS, Larsson MC. Transcriptome
736 Analysis of Gene Families Involved in Chemosensory Function in Spodoptera littoralis
737 (Lepidoptera: Noctuidae). BMC Genomics. 2019 [accessed 2019 Jun 4];20(1):428.
738 https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-5815-x.
739 doi:10.1186/s12864-019-5815-x
740 40. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S,
741 von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural
742 networks. Nature Biotechnology. 2019;37(4):420–423. doi:10.1038/s41587-019-0036-z

743 41. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
744 Improvements in Performance and Usability. Molecular Biology and Evolution. 2013
745 [accessed 2018 Jan 16];30(4):772–780. https://academic.oup.com/mbe/article-
746 lookup/doi/10.1093/molbev/mst010. doi:10.1093/molbev/mst010
747 42. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering Transposable Element
748 Diversification in De Novo Annotation Approaches Xu Y, editor. PLoS ONE. 2011 [accessed
749 2019 Jun 26];6(1):e16526. https://dx.plos.org/10.1371/journal.pone.0016526.
750 doi:10.1371/journal.pone.0016526
751 43. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M,
752 Anxolabehere D. Combined Evidence Annotation of Transposable Elements in Genome
753 Sequences. PLoS Computational Biology. 2005 [accessed 2019 Jun 26];1(2):e22.
754 https://dx.plos.org/10.1371/journal.pcbi.0010022. doi:10.1371/journal.pcbi.0010022
755 44. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory
756 elements in R and Bioconductor. Bioinformatics. 2016 [accessed 2020 Apr 27];32(4):587–
757 589. https://academic.oup.com/bioinformatics/article-abstract/32/4/587/1743969.
758 doi:10.1093/bioinformatics/btv612
759 45. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H,
760 Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple
761 sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539. internal-
762 pdf://105.231.243.229/Sievers-2011-Fast, scalable generation of high.pdf.
763 doi:10.1038/msb.2011.75
764 46. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large
765 phylogenies by maximum likelihood. Syst Biol. 2003;52(5):696–704.
766 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&
767 list_uids=14530136

768 47. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. Molecular
769 Biology and Evolution. 2017 [accessed 2018 Jan 16];34(9):2422–2424.
770 http://www.ncbi.nlm.nih.gov/pubmed/28472384. doi:10.1093/molbev/msx149
771 48. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast,
772 accurate, and powerful alternative. Systematic biology. 2006;55(4):539–552. internal-
773 pdf://254.68.141.233/Anisimova-2006-Approximate likelihood-ratio te.pdf
774 49. Darby CA, Stolzer M, Ropp PJ, Barker D, Durand D. Xenolog classification.
775 Bioinformatics. 2016 [accessed 2019 Jun 26];33(5):btw686.
776 https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw686.
777 doi:10.1093/bioinformatics/btw686
778 50. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses,
779 transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics. 2012
780 [accessed 2019 Jun 26];28(18):i409–i415. https://academic.oup.com/bioinformatics/article-
781 lookup/doi/10.1093/bioinformatics/bts386. doi:10.1093/bioinformatics/bts386
782 51. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines,
783 Timetrees, and Divergence Times. Molecular Biology and Evolution. 2017 [accessed 2019
784 Jun 26];34(7):1812–1819. http://www.ncbi.nlm.nih.gov/pubmed/28387841.
785 doi:10.1093/molbev/msx116
786 52. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol.
787 2007;24:1586–1591. internal-pdf://120.223.106.85/Yang-2007-PAML 4_ Phylogenetic
788 analysis by max.pdf
789 53. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence
790 alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(Web Server
791 issue):W609-12.
792 54. Yang Z, Swanson WJ, Vacquier VD. Maximum-likelihood analysis of molecular

793 adaptation in abalone sperm lysin reveals variable selective pressures among lineages and
794 sites. Mol Biol Evol. 2000;17(10):1446–1455. internal-pdf://118.68.164.92/Yang-2000-
795 Maximum-likelihood analysis of molec.pdf
796 55. Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization
797 proteins. Mol Biol Evol. 2003;20(1):18–20.
798 56. Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods
799 for detecting adaptive evolution in protein coding sequences and for identifying positively
800 selected sites. Genetics. 2004;168(2):1041–1051.
801 57. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for
802 detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–2479.
803 internal-pdf://240.85.165.245/Zhang-2005-Evaluation of an improved branch-si.pdf
804 58. Storey J, Bass A, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery
805 rate control. 2021 [accessed 2021 Dec 6]:R package version 2.22.0.
806 https://www.bioconductor.org/packages/release/bioc/html/qvalue.html
807 59. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K,
808 Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with
809 AlphaFold. Nature. 2021 [accessed 2021 Aug 3];596(7873):583–589.
810 https://doi.org/10.1038/s41586-021-03819-2. doi:10.1038/s41586-021-03819-2
811 60. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE.
812 UCSF Chimera—a visualization system for exploratory research and analysis. Journal of
813 computational chemistry. 2004;25(13):1605–1612. internal-pdf://3.231.87.78/Pettersen-2004-
814 UCSF Chimera—a visualization sy.pdf
815 61. Rytz R, Croset V, Benton R. Ionotropic Receptors (IRs): Chemosensory ionotropic
816 glutamate receptors in Drosophila and beyond. Insect Biochemistry and Molecular Biology.
817 2013;43(9):888–897. doi:10.1016/j.ibmb.2013.02.007

818 62. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. Variant Ionotropic Glutamate
819 Receptors as Chemosensory Receptors in Drosophila. Cell. 2009;136(1):149–162.
820 doi:10.1016/j.cell.2008.12.001
821 63. Croset V, Schleyer M, Arguello JR, Gerber B, Benton R. A molecular and neuronal basis
822 for amino acid sensing in the Drosophila larva. Scientific Reports. 2016;6.
823 doi:10.1038/srep34871
824 64. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton
825 R. Ancient Protostome Origin of Chemosensory Ionotropic Glutamate Receptors and the
826 Evolution of Insect Taste and Olfaction Stern DL, editor. PLoS Genetics. 2010 [accessed
827 2018 Jan 8];6(8):e1001064. http://dx.plos.org/10.1371/journal.pgen.1001064.
828 doi:10.1371/journal.pgen.1001064
829 65. Zhu JY, Xu ZW, Zhang XM, Liu NY. Genome-based identification and analysis of
830 ionotropic receptors in Spodoptera litura. Science of Nature. 2018;105(5–6).
831 doi:10.1007/s00114-018-1563-z
832 66. Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, Jermiin LS, Mcgaughran A,
833 Oakeshott JG, Papanikolaou A, Perera OP, et al. Genomic innovations, transcriptional
834 plasticity and gene loss underlying the evolution and divergence of two highly polyphagous
835 and invasive Helicoverpa pest species. [accessed 2017 Aug 8].
836 https://bmcbiol.biomedcentral.com/track/pdf/10.1186/s12915-017-0402-
837 6?site=bmcbiol.biomedcentral.com. doi:10.1186/s12915-017-0402-6
838 67. Wang Y, Fang G, Chen X, Cao Y, Wu N, Cui Q, Zhu C, Qian L, Huang Y, Zhan S. The
839 genome of the black cutworm Agrotis ipsilon. Insect Biochemistry and Molecular Biology.
840 2021;139:103665. doi:10.1016/J.IBMB.2021.103665
841 68. Kasubuchi M, Shii F, Tsuneto K, Yamagishi T, Adegawa S, Endo H, Sato R. Insect taste
842 receptors relevant to host identification by recognition of secondary metabolite patterns of

843 non-host plants. Biochemical and Biophysical Research Communications. 2018:2–7.
844 http://linkinghub.elsevier.com/retrieve/pii/S0006291X18307770.
845 doi:10.1016/j.bbrc.2018.04.014
846 69. Parisot N, Vargas-Chávez C, Goubert C, Baa-Puyoulet P, Balmand S, Beranger L, Blanc
847 C, Bonnamour A, Boulesteix M, Burlet N, et al. The transposable element-rich genome of the
848 cereal pest Sitophilus oryzae. BMC Biology 2021 19:1. 2021 [accessed 2021 Nov
849 19];19(1):1–28. https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-021-01158-2.
850 doi:10.1186/S12915-021-01158-2
851 70. Rebollo R, Romanish MT, Mager DL. Transposable Elements: An Abundant and Natural
852 Source of Regulatory Sequences for Host Genes. http://dx.doi.org/10.1146/annurev-genet-
853 110711-155621. 2012 [accessed 2021 Nov 19];46:21–42.
854 https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-110711-155621.
855 doi:10.1146/ANNUREV-GENET-110711-155621
856 71. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from
857 conflicts to benefits. Nature reviews. Genetics. 2016;advance on(2):71–86.
858 http://dx.doi.org/10.1038/nrg.2016.139. doi:10.1038/nrg.2016.139
859 72. Chen S, Li X. Transposable elements are enriched within or in close proximity to
860 xenobiotic-metabolizing cytochrome P450 genes. BMC Evolutionary Biology. 2007
861 [accessed 2021 Nov 19];7(1):1–13.
862 https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-7-46. doi:10.1186/1471-
863 2148-7-46/TABLES/2
864 73. You M, Yue Z, He W, Yang X, Yang G, Xie M, Zhan D, Baxter SW, Vasseur L, Gurr
865 GM, et al. A heterozygous moth genome provides insights into herbivory and detoxification.
866 Nature Genetics 2013 45:2. 2013 [accessed 2021 Nov 19];45(2):220–225.
867 https://www.nature.com/articles/ng.2524. doi:10.1038/ng.2524

868 74. Singh KS, Troczka BJ, Duarte A, Balabanidou V, Trissi N, Carabajal Paladino LZ,
869 Nguyen P, Zimmer CT, Papapostolou KM, Randall E, et al. The genetic architecture of a host
870 shift: An adaptive walk protected an aphid and its endosymbiont from plant chemical
871 defenses. Science Advances. 2020 [accessed 2021 Nov 19];6(19).
872 https://www.science.org/doi/abs/10.1126/sciadv.aba1070.
873 doi:10.1126/SCIADV.ABA1070/SUPPL_FILE/ABA1070_SM.PDF
874 75. Maumus F, Fiston-Lavier AS, Quesneville H. Impact of transposable elements on insect
875 genomes and biology. Current Opinion in Insect Science. 2015;7:30–36.
876 http://dx.doi.org/10.1016/j.cois.2015.01.001. doi:10.1016/j.cois.2015.01.001
877 76. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers
878 CP, Dinh H, Doddapaneni H, Dugan S, et al. Hemimetabolous genomes reveal molecular
879 basis of termite eusociality. Nature Ecology & Evolution 2017 2:3. 2018 [accessed 2021 Nov
880 19];2(3):557–566. https://www.nature.com/articles/s41559-017-0459-1. doi:10.1038/s41559-
881 017-0459-1
882 77. Zhang F, Zhang J, Yang Y, Wu Y. A chromosome-level genome assembly for the beet
883 armyworm (Spodoptera exigua) using PacBio and Hi-C sequencing. bioRxiv. 2019 Jan
884 1:2019.12.26.889121.
885 http://biorxiv.org/content/early/2019/12/28/2019.12.26.889121.abstract.
886 doi:10.1101/2019.12.26.889121
887 78. Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O,
888 Misof B. Diversity and evolution of the transposable element repertoire in arthropods with
889 particular reference to insects. BMC Evolutionary Biology. 2019;19(1). doi:10.1186/s12862-
890 018-1324-9
891 79. Hebert FO, Freschi L, Blackburn G, Béliveau C, Dewar K, Boyle B, Gundersen-Rindal
892 DE, Sparks ME, Cusson M, Hamelin RC, et al. Expansion of LINEs and species-specific

893 DNA repeats drives genome expansion in Asian Gypsy Moths. Scientific Reports.
894 2019;9(1):1–10. doi:10.1038/s41598-019-52840-z
895 80. Fiston-Lavier AS, Anxolabehere D, Quesneville H. A model of segmental duplication
896 formation in Drosophila melanogaster. Genome Research. 2007 [accessed 2020 Dec
897 17];17(10):1458–1470. www.genome.org/cgi/doi/10.1101/gr.6208307.
898 doi:10.1101/gr.6208307
899 81. Remnant EJ, Good RT, Schmidt JM, Lumb C, Robin C, Daborn PJ, Batterham P. Gene
900 duplication in the major insecticide target site, Rdl, in Drosophila melanogaster. Proceedings
901 of the National Academy of Sciences of the United States of America. 2013 [accessed 2020
902 Dec 17];110(36):14706–14710. www.pnas.org/cgi/doi/10.1073/pnas.1311341110.
903 doi:10.1073/pnas.1311341110
904 82. Mathiopoulos KD, Della Torre A, Predazzi V, Petrarca V, Coluzzi M. Cloning of
905 inversion breakpoints in the Anopheles gambiae complex traces a transposable element at the
906 inversion junction. Proceedings of the National Academy of Sciences of the United States of
907 America. 1998 [accessed 2020 Dec 17];95(21):12444–12449.
908 /pmc/articles/PMC22850/?report=abstract. doi:10.1073/pnas.95.21.12444
909 83. McKenzie SK, Kronauer DJC. The genomic architecture and molecular evolution of ant
910 odorant receptors. Genome research. 2018 Sep 24 [accessed 2018 Oct 1]:gr.237123.118.
911 http://www.ncbi.nlm.nih.gov/pubmed/30249741. doi:10.1101/gr.237123.118
912 84. McBride CS, Arguello JR. Five Drosophila genomes reveal nonneutral evolution and the
913 signature of host specialization in the chemoreceptor superfamily. Genetics. 2007 [accessed
914 2021 Jul 2];177(3):1395–1416. https://pubmed.ncbi.nlm.nih.gov/18039874/.
915 doi:10.1534/genetics.107.078683
916 85. Smadja C, Shi P, Butlin RK, Robertson HM. Large gene family expansions and adaptive
917 evolution for odorant and gustatory receptors in the pea aphid, Acyrthosiphon pisum.

Molecular Biology and Evolution. 2009;26(9):2073–2086. doi:10.1093/molbev/msp116

86. Hickner P V., 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 [accessed 2018 Jan 8];17(1):648. http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2983-9. doi:10.1186/s12864-016-2983-9

87. 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 Evolutionary Biology. 2018 [accessed 2018 Sep 25];18(1):144. https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-018-1250-x. doi:10.1186/s12862-018-1250-x

88. Smadja CM, Canbäck B, Vitalis R, Gautier M, Ferrari J, Zhou J-J, Butlin RK. LARGE-SCALE CANDIDATE GENE SCAN REVEALS THE ROLE OF CHEMORECEPTOR GENES IN HOST PLANT SPECIALIZATION AND SPECIATION IN THE PEA APHID. Evolution. 2012 [accessed 2016 Oct 6];66(9):2723–2738. http://doi.wiley.com/10.1111/j.1558-5646.2012.01612.x. doi:10.1111/j.1558-5646.2012.01612.x

89. Eyres I, Duvaux L, Gharbi K, Tucker R, Hopkins D, Simon JC, Ferrari J, Smadja CM, Butlin RK. Targeted re-sequencing confirms the importance of chemosensory genes in aphid host race differentiation. Molecular Ecology. 2017;26(1):43–58. doi:10.1111/mec.13818

90. Nouhaud P, Gautier M, Gouin A, Jaquiéry J, Peccoud J, Legeai F, Mieuzet L, Smadja CM, Lemaitre C, Vitalis R, et al. Identifying Genomic Hotspots Of Differentiation And Candidate Genes Involved In The Adaptive Divergence Of Pea Aphid Host Races. Molecular Ecology. 2018:0–3. http://doi.wiley.com/10.1111/mec.14799. doi:10.1111/mec.14799

91. Engsontia P, Sangket U, Chotigeat W, Satasook C. Molecular evolution of the odorant and

943 gustatory receptor genes in lepidopteran insects: Implications for their adaptation and
944 speciation. Journal of Molecular Evolution. 2014;79(1–2):21–39. doi:10.1007/s00239-014-
945 9633-0
946 92. Suzuki HC, Ozaki K, Makino T, Uchiyama H, Yajima S, Kawata M. Evolution of
947 gustatory receptor gene family provides insights into adaptation to diverse host plants in
948 nymphalid butterflies. Genome Biology and Evolution. 2018;(May).
949 https://academic.oup.com/gbe/advance-article/doi/10.1093/gbe/evy093/4999382.
950 doi:10.1093/gbe/evy093
951 93. Dweck HKM, Carlson JR. Molecular Logic and Evolution of Bitter Taste in Drosophila.
952 Current Biology. 2020 [accessed 2019 Dec 18];30(1):17-30.e3.
953 https://doi.org/10.1016/j.cub.2019.11.005. doi:10.1016/j.cub.2019.11.005
954 94. Mang D, Shu M, Tanaka S, Nagata S, Takada T, Endo H, Kikuta S, Tabunoki H,
955 Iwabuchi K, Sato R. Expression of the fructose receptor BmGr9 and its involvement in the
956 promotion of feeding, suggested by its co-expression with neuropeptide F1 in Bombyx mori.
957 Insect Biochemistry and Molecular Biology. 2016;75:58–69.
958 doi:10.1016/J.IBMB.2016.06.001
959 95. Sato K, Tanaka K, Touhara K. Sugar-regulated cation channel formed by an insect
960 gustatory receptor. Proceedings of the National Academy of Sciences. 2011;108(28):11680–
961 11685. http://www.pnas.org/cgi/doi/10.1073/pnas.1019622108.
962 doi:10.1073/pnas.1019622108
963 96. 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,
964 Tan A-J. A determining factor for insect feeding preference in the silkworm, Bombyx mori
965 Agrawal AA, editor. PLOS Biology. 2019 [accessed 2019 Mar 13];17(2):e3000162.
966 http://dx.plos.org/10.1371/journal.pbio.3000162. doi:10.1371/journal.pbio.3000162
967 97. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P,

968 Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable

969 elements. Nature Reviews Genetics. 2007 [accessed 2017 Jun 6];8(12):973–982.

970 http://www.nature.com/doifinder/10.1038/nrg2165. doi:10.1038/nrg2165

971