Comparing the Expression of Olfaction-Related Genes in Gypsy Moth (Lymantria dispar) Adult Females and Larvae from One Flightless and Two Flight-Capable Populations

Andrea Clavijo McCormick1†, Ewald Grosse-Wilde2, David Wheeler3, Mark C. Mescher1, Bill S. Hansson2 and Consuelo M. De Moraes*1

1 Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland, 2 Department of Evolutionary Neuroethology, Max Planck Institute for Chemical Ecology, Jena, Germany, 3 Institute of Fundamental Sciences, College of Sciences, Massey University, Palmerston North, New Zealand

In insects, flight and sophisticated olfactory systems go hand in hand and are essential to survival and evolutionary success. Females of many Lepidopteran species have secondarily lost their flight ability, which may lead to changes in the olfactory capabilities of both larval and adult stages. The gypsy moth, Lymantria dispar, an important forest pest worldwide, is currently undergoing a diversification process with three recognized subspecies: the Asian gypsy moth (AGM), Lymantria dispar asiatica; the Japanese gypsy moth (JGM), Lymantria dispar japonica; and the European gypsy moth (EGM), Lymantria dispar dispar. Females of EGM populations from North America have lost their flight capacity whereas the JGM and AGM females are flight capable, making this an ideal system to investigate the relationship between flight and olfaction. We used next-generation sequencing to obtain female antennal and larval head capsule transcriptomes in order to (i) investigate the differences in expression of olfaction-related genes among populations; (ii) identify the most similar protein sequences reported for other organisms through a BLAST search, and (iii) establish the phylogenetic relationships of these sequences with respect to other insect species. Using this approach, we identified 115 putative chemosensory genes belonging to five families of olfaction-related genes. A principal component analysis (PCA) revealed that the gene-expression patterns of female antennal transcriptomes from different subspecies were more similar to one another than to the larval head capsules of their respective subspecies supporting strong chemosensory differences between the two developmental stages. An analysis of the shared and exclusively expressed genes for three populations shows no evidence that loss of flight affects the number or type of genes being expressed. These results indicate either (a) that loss of flight does not impact the olfactory gene repertoire or (b) that the secondary loss of flight in American EGM populations may be too recent to have caused major changes in the genes being expressed. However, we found higher expression values for
most olfaction-related genes in EGM females, suggesting that differences in transcription rates could be an adaptation of flightless females to their chemical environment. Differences in olfactory genes and their expression in the larvae appear to be unrelated to the flight ability of adult females and are likely adaptations to different ecological pressures.

Keywords: *Lymantria dispar*, transcriptome, odorant receptor, ionotropic receptor, gustatory receptor, odorant binding protein, chemosensory protein

INTRODUCTION

Flight is a leading factor contributing to the evolutionary success of insect species, enabling them to locate food and shelter, avoid predation and competition, and search for optimal oviposition sites for their offspring (Barbosa et al., 1989; Sattler, 1991; Hunter, 1995). Since host–plant location and oviposition in herbivorous insects are largely mediated by chemical cues (Bruce et al., 2005; Bruce and Pickett, 2011; Mescher and De Moraes, 2015), one would expect the evolution of flight to be accompanied by the development of sophisticated olfactory systems. New evidence even suggests that the odorant receptor family (OR), central to the olfactory systems of highly derived insects, emerged around the same time as flight (Missbach et al., 2014; Ioannidis et al., 2017). Furthermore, manipulation of OR-based odor detection in *Drosophila* also indicates that ORs play an important role in flight orientation (Getahun et al., 2016).

The females of many Lepidopteran species have secondarily lost their ability to fly, shifting the responsibility of host selection partly or entirely to the larvae (Barbosa et al., 1989; Sattler, 1991; Hunter, 1995). In this context, it is interesting to investigate whether the loss of flight has an impact on the olfaction of adults and larvae. The gypsy moth *Lymantria dispar* is one of the most important forest pest species worldwide, currently undergoing a diversification process involving the loss of flight by females of some populations (Schweitzer, 2004; Pogue and Scheckler, 2007). These features make *L. dispar* an ideal model to explore changes in expression of olfaction-related genes that are associated with flight ability.

The first chemosensory proteins (CSPs) from adult *L. dispar* were identified as early as 1989 (Vogt et al., 1989, 1991). Thereafter, Plettner and coworkers have made great contributions to our understanding of olfaction in this species, in particular concerning the structure and function of its pheromone binding proteins (Kowcun et al., 2001; Honson et al., 2003; Honson and Plettner, 2006; Plettner and Gries, 2010; Gong and Plettner, 2011; Yu and Plettner, 2013). Recently, the *L. dispar* olfactory co-receptor (ORCO), a crucial component of olfactory receptor complexes, has been identified (Vosshall and Hanßson, 2011; Lin et al., 2015). However, knowledge about olfaction-related proteins and the genes encoding them remains fragmentary for this species.

The gypsy moth is a highly polyphagous herbivore, capable of causing severe and widespread outbreaks in temperate Holarctic regions. At present, there are three recognized subspecies: the Asian Gypsy moth (AGM) *Lymantria dispar asiatica*, the Japanese Gypsy moth (JGM) *Lymantria dispar japonica*, and the European Gypsy moth (EGM) *Lymantria dispar dispar* (which encompasses both European and North American Gypsy moth populations). European Gypsy moth females from North American populations are flightless, possibly due to a founder effect associated with their introduction from Europe in the mid-nineteenth century. In contrast, the Asian and Japanese females can fly and disperse over extended distances (Barlow, 2004; NBII, 2011; APHIS, 2013).

The loss of flight in the EGM females restricts their ability to make host-plant choices, transferring the responsibility to the larvae, which disperse either passively through ballooning in the early instars or actively by crawling in the late instars (Capiónera and Barbosa, 1976; Lance and Barbosa, 1981, 1982). The extent to which flight capable females are involved in host-plant choices is not yet fully understood, but evidence suggests that both AGM and JGM actively disperse and display oviposition preferences under field conditions (Baranchikov, 1989; Sasaki et al., 2016).

Several efforts have been made to better understand the taxonomic and biogeographic distribution of female flight ability, as well as its heritability and phenotypic plasticity (Keena et al., 2001, 2007, 2008, 2010). However, no studies have yet documented variation in the odor perception systems of *L. dispers* subspecies, despite the likelihood that such differences may accompany the loss of female flight. Therefore, the aims of this study were to (i) Investigate the differences in expression of olfaction-related genes among populations, (ii) identify the most similar protein sequences reported for other organisms through a BLAST search, and (iii) establish the phylogenetic relationships of these sequences with respect to other model insect species, most of which have fully sequenced genomes.

To fulfill these aims we focused on five groups of chemosensory gene families: odorant receptors (ORs), odorant binding proteins (OBPs), CSPs, gustatory receptors (GRs), and ionotropic receptors (IRs). ORs are expressed in the cell membranes of olfactory sensory neurons (OSNs) and are responsible for the detection of odor molecules (Sanchez-Gracia et al., 2009). In general, OSNs will express either ORs or IRs, with the latter mostly tuned to compounds of lower molecular weight (Hallem et al., 2004, 2006; Benton et al., 2009; Silbering et al., 2011). All analyzed Lepidoptera species possess more OR than IR types (Crosset et al., 2010; Koenig et al., 2015; van Schooten et al., 2016), and these play a role in the detection of plant volatiles as well as pheromones (Nakagawa et al., 2005; Grosse-Wilde et al., 2006, 2007; Tanaka et al., 2009). In insects, OSNs associated with basiconic or trichoid sensilla express one OR gene, along with the co-receptor ORCO, which is highly conserved and broadly expressed (Krieger et al., 2003; Touhara and Vosshall, 2009).
Insect ORs are seven-transmembrane domain receptors with inverted membrane topology and are not phylogenetically related to vertebrate ORs (Benton et al., 2006).

OBPs contribute to the sensitivity of the olfactory system by binding, solubilizing and transporting odorants through the sensillar lymph (Leal, 2013). CSPs are likely to perform similar roles in chemical communication of insects as OBPs, but unlike these are also expressed in non-chemosensory tissues, and for this reason have been hypothesized to serve additional, as yet undiscovered, functions (Pelosi et al., 2005). Recent evidence suggests that OBPs are an adaptation to the detection of hydrophobic volatiles that became available as olfactory cues in the course of insect terrestrialization (Missbach et al., 2015); however, results in Drosophila suggest a different function for some OBPs (Larter et al., 2016). Structurally, insect OBPs and CSPs generally contain α-helical domains, but folded in two different patterns (Sandler et al., 2000; Lartigue et al., 2002; Tegoni et al., 2004).

GRs are typically expressed in gustatory receptor neurons (GRNs) within the taste sensilla in the mouthparts and are known to detect sugars, bitter compounds and non-volatile pheromones (Montell, 2013). However, some GR genes are also expressed in the antennae, suggesting that some members of this gene family may have an olfactory function (Hallem et al., 2006). This is further supported by the discovery of two GRs in Drosophila that act in the detection of CO$_2$ (Yao and Carlson, 2010). GR proteins are highly divergent in sequence, sharing as little as 8% amino acid identity across insect species, and it has been hypothesized that the GR gene family is an ancient chemoreceptor family from which insect OR genes subsequently evolved (Robertson et al., 2003; Hallem et al., 2006; Benton, 2015).

IRs are also involved in chemoreception and comprise a large and highly diverse gene family closely related to ionotropic glutamate receptors (iGluR), typically present in the OSNs associated with the coeloconic sensilla in the antennae (Rytz et al., 2013). Recent reports suggest there are multiple variant IRs with different ligand-binding domains that lack the characteristic glutamate-interacting residues (Benton et al., 2009). Unlike ORs, which are exclusively found in pterygote insects, IRs are present in all protostome species studied so far and may have evolved as long as 550–850 million years ago (Croset et al., 2010; Missbach et al., 2014). Similar iGluR-like genes are also present in plants, animals and prokaryotes, indicating that this is an important and ancient group of chemoreceptors (Benton et al., 2009; Rytz et al., 2013).

MATERIALS AND METHODS

Animals

Insects were provided as egg masses by Hannah Nadel, Supervisory Entomologist of the Animal and Plant Health Inspection Service (APHIS) of the United States Department of Agriculture (USDA). All egg masses came from laboratory cultures that had been maintained using carefully designed mating protocols to avoid the deleterious effects of inbreeding depression, details on the rearing system utilized for these colonies can be found in (Bell et al., 1981). Upon hatching larvae were fed ad libitum on artificial wheat germ diet prepared according to manufacturer’s instructions (MP Biomedicals LLC, Illkirch, France) and food was replaced twice per week. Caterpillars, pupae, and adult moths were maintained in a climate chamber at 20°C, 60% relative humidity and 16/8 h photoperiod.

The European gypsy moth (EGM) culture (Lymantria dispar dispar) originated from flightless L. dispar populations collected in New Jersey (US). The Japanese gypsy moth (JGM) culture (Lymantria dispar japonica), originated from flight-capable populations coming from the Northern Iwate district and Takizawa, Morika, Nishine (Japan). The AGM culture (Lymantria dispar asiatica) originated from flight-capable populations coming from the Primorskiy Krai ports (Vostochnyy, Slavyanka, Vladivostok, Nadzhdovka) in Russia.

RNA Extraction

RNA extraction was performed following the same procedure as in Koenig et al. (2015), with minor changes as outlined below. Antennae of 50 adult female moths (1–2 days old) from each population were excised from the base of the antennal scerite. Head capsules from 50 fifth instar larvae from each population were cut at the division point with the prothorax. Tissues were transferred to an Eppendorf tube, cooled with liquid nitrogen and stored at −86°C until extraction. For extraction, tissues were transferred into RL buffer (innuPREP RNA Mini Kit, Analytik Jen, Jena, Germany) and homogenized using a Tissuelyser (Qiagen, Hilden, Germany). The resultant homogenate was used with the innuPREP RNA Mini Kit (Analytik Jen, Jena, Germany) following the manufacturers protocol.

Sequencing, Assembly, and Annotation

Total RNA was sent to the Max Planck Genome Centre Cologne (Germany) for construction of TruSeq libraries and subsequent sequencing on an Illumina HiSeq3000. Read data was trimmed and cleaned by the Genome Centre using standard protocols. The resulting Illumina reads were assembled with CLC Genomics Workbench 8 (CLCbio), using the de novo algorithm and default parameters. Annotation was performed using Blast2GO 3 (Conesa et al., 2005; Götz et al., 2008). Additionally, assembled transcripts belonging to target chemosensory families (OR, OBP, IR, GR, and CSP) were identified by comparison against custom, manually curated databases created using the available literature on other Lepidopteran species (Wanner and Robertson, 2008; Grosse-Wilde et al., 2011; Heliconius-Genome-Consortium, 2012; Briscoe et al., 2013; Koenig et al., 2015).

Each of the predicted protein sequences was further compared to available sequences using the blastp algorithm and the nr database (NCBI)$^1$ to identify the most similar sequence, the organism expressing it and its putative function. We only report sequences yielding significant ($E < 0.05$) similarity values.

Alignments and Phylogenetic Trees

Protein sequences conceptually translated from the assembled transcripts were aligned with homologs from Bombyx mori, $^1$NCBI [https://www.ncbi.nlm.nih.gov/blast/ Accessed 02.02.2017.]
<table>
<thead>
<tr>
<th>Contig</th>
<th>ID</th>
<th>Female antennae</th>
<th>Larval head capsule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Normalized gene expression (RPKM)</td>
<td>Normalized gene expression (RPKM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JGM</td>
<td>EGM</td>
</tr>
<tr>
<td>CSPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>CSP1</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>282</td>
<td>CSP2</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>316</td>
<td>CSP3</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>529</td>
<td>CSP4</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>546</td>
<td>CSP5</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2133</td>
<td>CSP6</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>3155</td>
<td>CSP7</td>
<td>11.326</td>
<td>6.897</td>
</tr>
<tr>
<td>4694</td>
<td>CSP8</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4803</td>
<td>CSP9</td>
<td>3.140</td>
<td>1.954</td>
</tr>
<tr>
<td>4927</td>
<td>CSP10</td>
<td>33.255</td>
<td>29.942</td>
</tr>
<tr>
<td>5687</td>
<td>CSP11</td>
<td>1.279</td>
<td>0.239</td>
</tr>
<tr>
<td>6311</td>
<td>CSP12</td>
<td>1.216</td>
<td>6.056</td>
</tr>
<tr>
<td>7764</td>
<td>CSP13</td>
<td>11.247</td>
<td>0.778</td>
</tr>
<tr>
<td>9858</td>
<td>CSP14</td>
<td>1.279</td>
<td>0.239</td>
</tr>
<tr>
<td>10611</td>
<td>CSP15</td>
<td>1.279</td>
<td>0.239</td>
</tr>
<tr>
<td>11424</td>
<td>CSP16</td>
<td>1.279</td>
<td>0.239</td>
</tr>
<tr>
<td>14171</td>
<td>CSP17</td>
<td>299.658</td>
<td>89.839</td>
</tr>
<tr>
<td>20492</td>
<td>CSP18</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>21710</td>
<td>CSP19</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>21764</td>
<td>CSP20</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>24844</td>
<td>CSP21</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>28449</td>
<td>CSP22</td>
<td>−</td>
<td>0.438</td>
</tr>
<tr>
<td>OBPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15931</td>
<td>PBP_A</td>
<td>36.828</td>
<td>212.317</td>
</tr>
<tr>
<td>32039</td>
<td>PBP_B</td>
<td>34.345</td>
<td>294.223</td>
</tr>
<tr>
<td>32635</td>
<td>PBP_C</td>
<td>0.605</td>
<td>38.621</td>
</tr>
<tr>
<td>33887</td>
<td>PBP_D</td>
<td>2.019</td>
<td>14.140</td>
</tr>
<tr>
<td>34051</td>
<td>PBP_E</td>
<td>−</td>
<td>0.831</td>
</tr>
<tr>
<td>17785</td>
<td>GOBP_1</td>
<td>491.190</td>
<td>3,382.465</td>
</tr>
<tr>
<td>20294</td>
<td>GOBP_2</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>80</td>
<td>OBP1</td>
<td>1.705</td>
<td>0.995</td>
</tr>
<tr>
<td>81</td>
<td>OBP2</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>1985</td>
<td>OBP3</td>
<td>19.360</td>
<td>30.768</td>
</tr>
<tr>
<td>2548</td>
<td>OBP4</td>
<td>51.592</td>
<td>695.792</td>
</tr>
<tr>
<td>4026</td>
<td>OBP5</td>
<td>0.766</td>
<td>5.295</td>
</tr>
<tr>
<td>4999</td>
<td>OBP6</td>
<td>78.013</td>
<td>233.009</td>
</tr>
<tr>
<td>5449</td>
<td>OBP7</td>
<td>41.716</td>
<td>−</td>
</tr>
<tr>
<td>5666</td>
<td>OBP8</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>11867</td>
<td>OBP9</td>
<td>0.265</td>
<td>−</td>
</tr>
<tr>
<td>12090</td>
<td>OBP10</td>
<td>2.555</td>
<td>6.362</td>
</tr>
<tr>
<td>18226</td>
<td>OBP11</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>19950</td>
<td>OBP12</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>19951</td>
<td>OBP13</td>
<td>3.772</td>
<td>10.886</td>
</tr>
<tr>
<td>24041</td>
<td>OBP14</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>24520</td>
<td>OBP15</td>
<td>0.167</td>
<td>1.089</td>
</tr>
<tr>
<td>25516</td>
<td>OBP16</td>
<td>1.781</td>
<td>0.416</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Contig ID</th>
<th>Gene</th>
<th>Contig</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
<th>Normalized gene expression (RPKM)</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>26082</td>
<td>OBP17</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>1.317</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>26834</td>
<td>OBP18</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.554</td>
<td>−</td>
<td>−</td>
<td>0.570</td>
</tr>
<tr>
<td>26945</td>
<td>OBP19</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>3.192</td>
</tr>
<tr>
<td>31544</td>
<td>OBP20</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.978</td>
<td>2.344</td>
<td>6.919</td>
<td></td>
</tr>
<tr>
<td>33379</td>
<td>OBP21</td>
<td>48.640</td>
<td>587.915</td>
<td>112.383</td>
<td>0.074</td>
<td>0.570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33405</td>
<td>OBP22</td>
<td>21.469</td>
<td>132.303</td>
<td>42.160</td>
<td>0.593</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33456</td>
<td>OBP23</td>
<td>39.212</td>
<td>171.314</td>
<td>94.649</td>
<td>0.689</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34786</td>
<td>OBP24</td>
<td>18.142</td>
<td>65.449</td>
<td>31.083</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34788</td>
<td>OBP25</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>3.579</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GRs**

<table>
<thead>
<tr>
<th>Contig ID</th>
<th>Gene</th>
<th>Contig</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
<th>Normalized gene expression (RPKM)</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>587</td>
<td>GR1</td>
<td>0.606</td>
<td>5.087</td>
<td>3.135</td>
<td>140.325</td>
<td>98.163</td>
<td>64.391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23948</td>
<td>GR2</td>
<td>0.418</td>
<td>0.130</td>
<td>0.601</td>
<td>0.308</td>
<td>1.757</td>
<td>0.396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32417</td>
<td>GR3</td>
<td>0.374</td>
<td>0.349</td>
<td>0.323</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32355</td>
<td>GR4</td>
<td>0.471</td>
<td>0.436</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33472</td>
<td>GR5</td>
<td>0.372</td>
<td>1.505</td>
<td>1.070</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34141</td>
<td>GR6</td>
<td>0.760</td>
<td>2.130</td>
<td>1.313</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34172</td>
<td>GR7</td>
<td>0.337</td>
<td>0.321</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34277</td>
<td>GR8</td>
<td>0.471</td>
<td>0.436</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34291</td>
<td>GR9</td>
<td>0.372</td>
<td>1.505</td>
<td>1.070</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34451</td>
<td>GR10</td>
<td>−</td>
<td>0.347</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34464</td>
<td>GR11</td>
<td>−</td>
<td>0.408</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ORs**

<table>
<thead>
<tr>
<th>Contig ID</th>
<th>Gene</th>
<th>Contig</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
<th>Normalized gene expression (RPKM)</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>20670</td>
<td>ORCO</td>
<td>0.114</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.818</td>
<td>3.045</td>
<td>0.712</td>
<td></td>
</tr>
<tr>
<td>6380</td>
<td>OR1</td>
<td>0.197</td>
<td>2.116</td>
<td>0.766</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8989</td>
<td>OR2</td>
<td>3.058</td>
<td>12.081</td>
<td>5.635</td>
<td>22.206</td>
<td>18.520</td>
<td>13.248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15892</td>
<td>OR3</td>
<td>0.206</td>
<td>1.349</td>
<td>0.713</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17788</td>
<td>OR4</td>
<td>−</td>
<td>1.779</td>
<td>0.706</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27443</td>
<td>OR5</td>
<td>−</td>
<td>0.239</td>
<td>−</td>
<td>0.698</td>
<td>0.472</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31928</td>
<td>OR6</td>
<td>0.457</td>
<td>2.218</td>
<td>0.631</td>
<td>1.617</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32971</td>
<td>OR7</td>
<td>0.099</td>
<td>1.473</td>
<td>1.107</td>
<td>0.164</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33087</td>
<td>OR8</td>
<td>−</td>
<td>0.464</td>
<td>0.429</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33560</td>
<td>OR9</td>
<td>1.447</td>
<td>3.860</td>
<td>1.606</td>
<td>0.171</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33844</td>
<td>OR10</td>
<td>0.577</td>
<td>0.359</td>
<td>1.494</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33855</td>
<td>OR11</td>
<td>0.383</td>
<td>3.357</td>
<td>1.915</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33861</td>
<td>OR12</td>
<td>−</td>
<td>0.756</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33879</td>
<td>OR13</td>
<td>0.707</td>
<td>1.541</td>
<td>0.814</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33888</td>
<td>OR14</td>
<td>0.387</td>
<td>0.343</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33903</td>
<td>OR15</td>
<td>0.223</td>
<td>2.294</td>
<td>0.578</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33963</td>
<td>OR16</td>
<td>0.661</td>
<td>0.925</td>
<td>0.285</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33998</td>
<td>OR17</td>
<td>0.759</td>
<td>1.417</td>
<td>0.655</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34011</td>
<td>OR18</td>
<td>0.761</td>
<td>1.776</td>
<td>1.313</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34012</td>
<td>OR19</td>
<td>0.378</td>
<td>2.469</td>
<td>1.304</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34122</td>
<td>OR20</td>
<td>−</td>
<td>1.178</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34175</td>
<td>OR21</td>
<td>0.376</td>
<td>−</td>
<td>0.325</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34202</td>
<td>OR22</td>
<td>0.334</td>
<td>0.779</td>
<td>1.440</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34209</td>
<td>OR23</td>
<td>−</td>
<td>0.953</td>
<td>0.881</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34210</td>
<td>OR24</td>
<td>−</td>
<td>0.800</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34217</td>
<td>OR25</td>
<td>0.229</td>
<td>1.283</td>
<td>0.198</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 1 | Continued

<table>
<thead>
<tr>
<th>Contig</th>
<th>ID</th>
<th>Female antennae</th>
<th>Larval head capsule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Normalized gene expression (RPKM)</td>
<td>Normalized gene expression (RPKM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JGM</td>
<td>EGM</td>
</tr>
<tr>
<td>34226</td>
<td>OR26</td>
<td>0.307</td>
<td>0.859</td>
</tr>
<tr>
<td>34270</td>
<td>OR27</td>
<td>0.150</td>
<td>1.120</td>
</tr>
<tr>
<td>34280</td>
<td>OR28</td>
<td>–</td>
<td>0.291</td>
</tr>
<tr>
<td>34293</td>
<td>OR29</td>
<td>0.302</td>
<td>1.412</td>
</tr>
<tr>
<td>34376</td>
<td>OR30</td>
<td>–</td>
<td>1.006</td>
</tr>
<tr>
<td>34421</td>
<td>OR31</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>34819</td>
<td>OR32</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>34881</td>
<td>OR33</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**IRs**

<table>
<thead>
<tr>
<th>Contig</th>
<th>ID</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2720</td>
<td>IR1</td>
<td>–</td>
<td>–</td>
<td>6.302</td>
</tr>
<tr>
<td>5445</td>
<td>IR2</td>
<td>0.116</td>
<td>0.731</td>
<td>0.025</td>
</tr>
<tr>
<td>6481</td>
<td>IR3</td>
<td>–</td>
<td>–</td>
<td>6.003</td>
</tr>
<tr>
<td>8501</td>
<td>IR4</td>
<td>0.986</td>
<td>6.627</td>
<td>7.011</td>
</tr>
<tr>
<td>20445</td>
<td>IR5</td>
<td>1.028</td>
<td>6.673</td>
<td>2.796</td>
</tr>
<tr>
<td>30320</td>
<td>IR7</td>
<td>0.323</td>
<td>2.411</td>
<td>1.393</td>
</tr>
<tr>
<td>32236</td>
<td>IR8</td>
<td>0.423</td>
<td>1.500</td>
<td>0.803</td>
</tr>
<tr>
<td>32470</td>
<td>IR9</td>
<td>–</td>
<td>0.483</td>
<td>1.859</td>
</tr>
<tr>
<td>32618</td>
<td>IR10</td>
<td>–</td>
<td>0.720</td>
<td>0.333</td>
</tr>
<tr>
<td>33239</td>
<td>IR11</td>
<td>0.391</td>
<td>2.067</td>
<td>0.899</td>
</tr>
<tr>
<td>33881</td>
<td>IR12</td>
<td>1.419</td>
<td>4.142</td>
<td>1.021</td>
</tr>
<tr>
<td>33900</td>
<td>IR13</td>
<td>1.143</td>
<td>11.203</td>
<td>3.233</td>
</tr>
<tr>
<td>34101</td>
<td>IR14</td>
<td>0.280</td>
<td>2.789</td>
<td>1.611</td>
</tr>
<tr>
<td>34111</td>
<td>IR15</td>
<td>0.571</td>
<td>1.777</td>
<td>1.314</td>
</tr>
<tr>
<td>34374</td>
<td>IR16</td>
<td>–</td>
<td>0.591</td>
<td>0.546</td>
</tr>
</tbody>
</table>

**Glu-Rs**

<table>
<thead>
<tr>
<th>Contig</th>
<th>ID</th>
<th>JGM</th>
<th>EGM</th>
<th>AGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1873</td>
<td>Glu-RX.1</td>
<td>–</td>
<td>3.176</td>
<td>0.117</td>
</tr>
<tr>
<td>10442</td>
<td>Glu-RX.2</td>
<td>0.045</td>
<td>0.547</td>
<td>–</td>
</tr>
<tr>
<td>10516</td>
<td>Glu-RX.3</td>
<td>0.067</td>
<td>0.676</td>
<td>–</td>
</tr>
<tr>
<td>17885</td>
<td>Glu-RX.4</td>
<td>0.278</td>
<td>–</td>
<td>0.614</td>
</tr>
<tr>
<td>26117</td>
<td>Nmdar1</td>
<td>–</td>
<td>0.247</td>
<td>–</td>
</tr>
<tr>
<td>3548</td>
<td>Nmdar2</td>
<td>0.634</td>
<td>2.748</td>
<td>0.664</td>
</tr>
</tbody>
</table>

*Transcripts have been tentatively labeled following the naming code of closely related sequences ([Figures 3–6](#)).

CSP, Chemosensory protein; OBP, Odorant binding protein; GR, Gustatory receptor; OR, Odorant receptor; IR, Ionotropic receptor; Glu-R, Glutamate receptor; ORCO, Odorant receptor co-receptor; PBP, Pheromone binding protein; GOBP, General odorant binding protein; Nmdar, N-methyl-D-aspartate receptor; JGM, Japanese gypsy moth; EGM, European gypsy moth and AGM; Asian gypsy moth. Values in bold represent higher expression values for the EGM females in comparison to JGM and AGM populations.

---

**Danaus plexippus, Heliconius melpomene, and Manduca sexta** (Wanner and Robertson, 2008; Grosse-Wilde et al., 2011; Heliconius-Genome-Consortium, 2012; Briscoe et al., 2013; Koenig et al., 2015). In the case of GRs and ORs, sequences from the water flea *Daphnia pulex* were also included as an outgroup (Peñalva-Arana et al., 2009). For the CSPs and OBPs we included sequences from the Jumping Bristletail *Lepismachilis y-signata* and the Firebrat *Thermobia domestica* (Missbach et al., 2015). In the case of IRs (and Glu-Rs) sequences from *Drosophila melanogaster* have been included (Rytz et al., 2013).

For this purpose, we used MAFFT version 7 (Katoh et al., 2002; Katoh and Standley, 2013) with the “-auto” option. Phylogenetic trees were derived using the program FastTree-2, which uses the maximum likelihood method with a Shimodaira-Hasegawa test to estimate branch support values (Price et al., 2010). Figures were prepared for publication using the FigTree software 1.4.1 (Rambaut, 2007, 2012).

Some transcripts, corresponding to pheromone binding proteins (PBPs), general odorant binding proteins (GOBPs), glutamate receptors (Glu-Rs and Nmdars = N-methyl-D-aspartate receptors) and the ORCO were tentatively labeled following the naming code of closely related sequences.
FIGURE 1 | Shared and exclusively expressed genes for three populations of the Gypsy moth for different classes of olfaction-related gene families in both female antennae and larval head capsules.
**Quantification of Gene Expression**

For the quantification of gene expression levels in the respective tissues/subspecies, the annotated assemblies were used as a template, mapping the raw reads and performing RPKM analysis in CLC Genomics Workbench 8 using default settings. PCA plots were based on normalized count data that was transformed using the regularized log function implemented by the R package DESeq2 (doi: 10.1186/s13059-014-0550-8).

**RESULTS**

**Gene Identification and Expression Patterns for the Three *L. dispar* Populations**

We used next generation sequencing to obtain transcriptome assemblies of adult female antennae and larval head capsules from EGM, AGM, and JGM populations of *L. dispar*. The assemblies contained 28,004, 33,208, and 30,820 unique transcripts for EGM, AGM, and JGM populations, respectively. Blastx of the assembled transcripts to the NCBI refseq protein database revealed that that 46.3% (EGM), 52.6% (AGM), and 49.3% (JGM) had high homology (E < 1e-5) to previously characterized proteins at NCBI. To ascertain the transcript coverage of each assembly, we used Blastx to find the proportion of *B. mori* proteins that aligned in a high scoring alignment. We chose *B. mori* because it has one of the best characterized genomes of the Lepidoptera. This analysis showed that an average *L. dispar* transcript encodes just over half the expected protein sequence based on the best blastx hit to *B. mori*, possibly due to a high proportion of partial sequences (Supplementary Figure 1).

From the assembled transcripts we were able to identify 115 putative chemosensory transcripts belonging to the five families, 22 CSPs, 32 OBPs (including 2 GOBPs, and four pheromone binding proteins), 11 GRs, 33 ORs, and 16 IRs (Table 1). In addition we report 6 glutamate receptors (which are not chemosensory receptors) (Table 1). Our results show that 42 olfaction-related genes are found in at least one population in both female antennae and larval head capsules, 52 are exclusive to the female antennae, and 20 to the larval head capsules. A large contribution to the transcripts that are exclusive to the female antennae comes from the ORs (Table 1). Figure 1 depicts the differences and commonalities in gene expression (presence/absence) among the three populations for each chemosensory gene family.

A Principal Component Analysis (PCA) comparing the gene expression patterns (Normalized gene expression values—RPKM) for the antennal and head capsule transcriptomes revealed that female antennal transcriptomes were clustered, being more similar to one another than to the larval transcriptomes of the same population. In contrast, larval transcriptomes were not clustered, but separated along the second component axis (Figure 2).

**Best Match with Other Protein Sequences**

After performing Blast searches with the individual protein sequences, we found that most putative *L. dispar* CSPs have a high sequence homology with those already published for a number of Lepidopteran species, the majority of which are Noctuid moths belonging to the genera *Helicoverpa* or *Spodoptera* (Table 2).

**Phylogenetic Positioning of Putative Protein Sequences**

We constructed phylogenetic trees from alignments of the *L. dispar* CSPs with other published sequences from model insect species (*B. mori*, *H. melpomene*, *M. sexta*, and *D. plexippus*, *D. pulex*, *L. y-signata*, *T. domestica*, and *D. melanogaster*). Fasta sequences used to construct the trees can be found in Supplementary Files 1–4.

The phylogenetic trees showed that CSPs aligned well within the published sequences, but in a few cases formed clusters containing only *L. dispar* sequences (e.g., CSPs 3, 16, 13, 21; CSPs 2, 8, 14, and 10; CSPs 5, 4, 1) (Figure 3). For OBPs most sequences were closely related to those reported for the model species, except OBPs 10, 3, 1, 2, 6, 8, 13, and 14 forming a branch unique to *L. dispar* and a few others forming single nodes (e.g., OBP11) (Figure 4).

In the case of the GRs and ORs, sequences are remarkably well nested within those of model species. Of particular interest is GR2 making a single node, and branch containing GR1 and ORs 2 and 4 unique to *L. dispar* (Figure 5). For most IRs, we found that the candidate gene sequences were partially aligned with those of the model species, with a few sequences (e.g., IR2) forming single nodes. Most Glu-Rs formed a branch unique to *L. dispar* (Figure 6).

![Figure 2](image-url)
**TABLE 2** | List of *L. dispar* transcripts putatively involved in chemoreception, and characterization for the best hit after comparison with available protein sequences using the BlastP algorithm.

<table>
<thead>
<tr>
<th>ID</th>
<th>Best match/name</th>
<th>Species</th>
<th>Score</th>
<th>E-value</th>
<th>Acc. Nr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSP1</td>
<td>Chemosensory protein 1</td>
<td><em>Athetis dissimilis</em> (Lep:Noc)</td>
<td>156</td>
<td>1E-45</td>
<td>ALJ09810</td>
</tr>
<tr>
<td>CSP2</td>
<td>Chemosensory protein 12</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>159</td>
<td>1E-46</td>
<td>AKT26488</td>
</tr>
<tr>
<td>CSP3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CSP4</td>
<td>Chemosensory protein 7</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>71.2</td>
<td>8E-14</td>
<td>AKT26484</td>
</tr>
<tr>
<td>CSP5</td>
<td>Chemosensory protein 24</td>
<td><em>Cnaphalocrocis medinalis</em> (Lep:Cra)</td>
<td>131</td>
<td>2E-36</td>
<td>ALT31606</td>
</tr>
<tr>
<td>CSP6</td>
<td>Chemosensory protein</td>
<td><em>Helicoverpa assulta</em> (Lep:Noc)</td>
<td>150</td>
<td>1E-43</td>
<td>ABB91378</td>
</tr>
<tr>
<td>CSP7</td>
<td>chemosensory protein 10</td>
<td><em>Helicoverpa armigera</em> (Lep:Noc)</td>
<td>214</td>
<td>2E-68</td>
<td>AFR92094</td>
</tr>
<tr>
<td>CSP8</td>
<td>Chemosensory protein 12</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>132</td>
<td>2E-36</td>
<td>AKT26484</td>
</tr>
<tr>
<td>CSP9</td>
<td>Chemosensory protein 16</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>195</td>
<td>3E-61</td>
<td>AKT26491</td>
</tr>
<tr>
<td>CSP10</td>
<td>Chemosensory protein 12</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>144</td>
<td>4E-41</td>
<td>AKT26488</td>
</tr>
<tr>
<td>CSP11</td>
<td>Chemosensory protein 5</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>182</td>
<td>1E-55</td>
<td>AKT26482</td>
</tr>
<tr>
<td>CSP12</td>
<td>Chemosensory protein</td>
<td><em>Helicoverpa armigera</em> (Lep:Noc)</td>
<td>182</td>
<td>3E-56</td>
<td>ANW65100</td>
</tr>
<tr>
<td>CSP13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CSP14</td>
<td>Chemosensory protein 12</td>
<td><em>Spodoptera exigua</em> (Lep:Noc)</td>
<td>161</td>
<td>1E-47</td>
<td>AKT26488</td>
</tr>
<tr>
<td>CSP15</td>
<td>Chemosensory protein 25</td>
<td><em>Cnaphalocrocis medinalis</em> (Lep:Cra)</td>
<td>189</td>
<td>2E-58</td>
<td>ALT31607</td>
</tr>
<tr>
<td>CSP16</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CSP17</td>
<td>Chemosensory protein 27, partial</td>
<td><em>Cnaphalocrocis medinalis</em> (Lep:Cra)</td>
<td>182</td>
<td>7E-56</td>
<td>ALT31609</td>
</tr>
<tr>
<td>CSP18</td>
<td>Putative chemosensory protein</td>
<td><em>Sesamia inferens</em> (Lep:Noc)</td>
<td>97.4</td>
<td>5E-23</td>
<td>AGY49263</td>
</tr>
<tr>
<td>CSP19</td>
<td>Hypothetical protein KGM_11196</td>
<td><em>Spodoptera litura</em> (Lep:Noc)</td>
<td>152</td>
<td>1E-43</td>
<td>EJH76400</td>
</tr>
<tr>
<td>CSP20</td>
<td>Bulb-specific protein 3-like</td>
<td><em>Papilio machaon</em> (Lep:Pap)</td>
<td>82.8</td>
<td>2E-17</td>
<td>XP_014365701</td>
</tr>
<tr>
<td>CSP21</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CSP22</td>
<td>Chemosensory protein</td>
<td><em>Papilio xuthus</em> (Lep:Pap)</td>
<td>101</td>
<td>2E-24</td>
<td>BAF91714</td>
</tr>
</tbody>
</table>

**OBPs**

<table>
<thead>
<tr>
<th>OBPs</th>
<th>Pheromone binding protein 1 precursor</th>
<th>Sesamia nonagrioides (Lep:Noc)</th>
<th>165</th>
<th>3E-48</th>
<th>AAS49922</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBP_A</td>
<td>Pheromone binding protein</td>
<td><em>Heliothis virescens</em> (Lep:Noc)</td>
<td>189</td>
<td>1E-57</td>
<td>CAAS8504</td>
</tr>
<tr>
<td>PBP_B</td>
<td>Pheromone-binding protein 3</td>
<td><em>Spodoptera litura</em> (Lep:Noc)</td>
<td>208</td>
<td>4E-65</td>
<td>AIS72934</td>
</tr>
<tr>
<td>PBP_C</td>
<td>Pheromone binding protein 2</td>
<td><em>Epiphyas postvittana</em> (Lep:Tot)</td>
<td>88.6</td>
<td>5E-20</td>
<td>AAL05868</td>
</tr>
<tr>
<td>PBP_D</td>
<td>Pheromone binding protein 3</td>
<td><em>Sesamia inferens</em> (Lep:Noc)</td>
<td>49.3</td>
<td>–006</td>
<td>AEQ30020</td>
</tr>
<tr>
<td>GOBP1</td>
<td>General odorant-binding protein 2</td>
<td><em>Heliothis virescens</em> (Lep:Noc)</td>
<td>424</td>
<td>3E-134</td>
<td>Q27288</td>
</tr>
<tr>
<td>OBPs</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>OBPs</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>OBPs</td>
<td>Odorant binding protein 26</td>
<td><em>Spodoptera frugiperda</em> (Lep:Noc)</td>
<td>139</td>
<td>2E-38</td>
<td>AAR28762</td>
</tr>
<tr>
<td>OBPs</td>
<td>Odorant binding protein 26</td>
<td><em>Spodoptera frugiperda</em> (Lep:Noc)</td>
<td>133</td>
<td>4E-36</td>
<td>AAR28762</td>
</tr>
<tr>
<td>OBPs</td>
<td>Odorant binding protein 26</td>
<td><em>Spodoptera frugiperda</em> (Lep:Noc)</td>
<td>141</td>
<td>3E-39</td>
<td>AAR28762</td>
</tr>
<tr>
<td>OBPs</td>
<td>Odorant binding protein 26</td>
<td><em>Galleria mellonella</em> (Lep:Pyr)</td>
<td>248</td>
<td>7E-81</td>
<td>AAA85090</td>
</tr>
<tr>
<td>OBPs</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>OBPs</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>OBPs</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>ID</th>
<th>Best match/name</th>
<th>Species</th>
<th>Score</th>
<th>E-value</th>
<th>Acc. Nr</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBP17</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>AGH70102</td>
</tr>
<tr>
<td>OBP18</td>
<td>Odorant binding protein 9</td>
<td>Spodoptera exigua (Lep:Noc)</td>
<td>75.1</td>
<td>8E-15</td>
<td>AGP03455</td>
</tr>
<tr>
<td>OBP19</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>OBP20</td>
<td>Odorant binding protein 44a, isoform A</td>
<td>Drosophila melanogaster (Dipt:Dro)</td>
<td>456</td>
<td>3E-148</td>
<td>NP_610358</td>
</tr>
<tr>
<td>OBP21</td>
<td>Odorant binding protein 3</td>
<td>Spodoptera litura (Lep:Noc)</td>
<td>275</td>
<td>1E-80</td>
<td>AK87964</td>
</tr>
<tr>
<td>OBP22</td>
<td>General odorant binding protein 72-like</td>
<td>Papilio machaon (Lep:Pap)</td>
<td>176</td>
<td>4E-53</td>
<td>XP_014369849</td>
</tr>
<tr>
<td>OBP23</td>
<td>Odorant binding protein 4</td>
<td>Spodoptera litura (Lep:Noc)</td>
<td>184</td>
<td>7E-56</td>
<td>AK87965</td>
</tr>
<tr>
<td>OBP24</td>
<td>Odorant binding protein 13</td>
<td>Helicoverpa armigera (Lep:Noc)</td>
<td>253</td>
<td>4E-72</td>
<td>EBS4588</td>
</tr>
<tr>
<td>OBP25</td>
<td>Odorant binding protein 1</td>
<td>Cnaphalocrocis medinalis (Lep:Cra)</td>
<td>126</td>
<td>2E-34</td>
<td>AFG72998</td>
</tr>
<tr>
<td>GR1</td>
<td>Uncharacterized protein LOC106133470 (pred)</td>
<td>Amyelois transitella (Lep:Pyr)</td>
<td>349</td>
<td>8E-117</td>
<td>XP_013188565</td>
</tr>
<tr>
<td>GR2</td>
<td>Ecdysis triggering hormone receptor subtype-A</td>
<td>Manduca sexta (Lep:Sph)</td>
<td>535</td>
<td>0</td>
<td>AAX19163</td>
</tr>
<tr>
<td>GR3</td>
<td>Gustatory receptor 3, partial</td>
<td>Athetis dissimilis (Lep:Noc)</td>
<td>195</td>
<td>5E-60</td>
<td>ALM26253</td>
</tr>
<tr>
<td>GR4</td>
<td>Olfactory receptor 1</td>
<td>Daphnia indica (Lep:Cra)</td>
<td>84.3</td>
<td>8E-17</td>
<td>BAG71147</td>
</tr>
<tr>
<td>GR5</td>
<td>Olfactory receptor 4, partial</td>
<td>Helicoverpa armigera (Lep:Noc)</td>
<td>160</td>
<td>6E-45</td>
<td>ACF32962</td>
</tr>
<tr>
<td>GR6</td>
<td>Odorant receptor 47, partial</td>
<td>Athetis dissimilis (Lep:Noc)</td>
<td>317</td>
<td>8E-102</td>
<td>ALM26237</td>
</tr>
<tr>
<td>GR7</td>
<td>Odorant receptor</td>
<td>Dendrolimus kikuchi (Lep:Las)</td>
<td>490</td>
<td>3E-169</td>
<td>AIO1083</td>
</tr>
<tr>
<td>GR8</td>
<td>Gustatory and odorant receptor 24-like (pred)</td>
<td>Plutella xylostella (Lep:Plu)</td>
<td>192</td>
<td>2E-58</td>
<td>XP_011558384</td>
</tr>
<tr>
<td>GR9</td>
<td>Odorant receptor</td>
<td>Dendrolimus kikuchi (Lep:Las)</td>
<td>310</td>
<td>5E-98</td>
<td>AIO1083</td>
</tr>
<tr>
<td>GR10</td>
<td>Odorant receptor</td>
<td>Dendrolimus kikuchi (Lep:Las)</td>
<td>136</td>
<td>9E-36</td>
<td>AIO1090</td>
</tr>
<tr>
<td>GR11</td>
<td>Odorant receptor, partial</td>
<td>Helicoverpa armigera (Lep:Noc)</td>
<td>116</td>
<td>2E-28</td>
<td>AIG51896</td>
</tr>
</tbody>
</table>

**GRs**

- **ORCs**
  - Protein trapped in endoderm-1 isof orm X2 (pred)
  - Amyelois transitella (Lep:Pyr)

- **OR1**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 426
  - E-value: 4E-143
  - Accession: AIG51860

- **OR2**
  - Ecdysis triggering hormone receptor subtype-A
  - Manduca sexta (Lep:Sph)
  - Score: 535
  - E-value: 0
  - Accession: AAX19163

- **OR3**
  - Putative odorant receptor
  - Sesamia inferens (Lep:Noc)
  - Score: 509
  - E-value: 6E-175
  - Accession: AYG14579

- **OR4**
  - Odorant receptor, partial
  - Helicoverpa armigera (Lep:Noc)
  - Score: 59.7
  - E-value: 8E-6
  - Accession: AIG51896

- **OR5**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 86.7
  - E-value: 2E-17
  - Accession: AIG51875

- **OR6**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 335
  - E-value: 2E-109
  - Accession: AIG51896

- **OR7**
  - Putative odorant-binding protein
  - Helicoverpa armigera (Lep:Noc)
  - Score: 169
  - E-value: 3E-51
  - Accession: AEGJ0553

- **OR8**
  - Putative odorant receptor, partial
  - Sesamia inferens (Lep:Noc)
  - Score: 286
  - E-value: 2E-92
  - Accession: AYG14577

- **OR9**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 620
  - E-value: 0
  - Accession: AIG51879

- **OR10**
  - Putative olfactory receptor 12
  - Spodoptera litura (Lep:Noc)
  - Score: 274
  - E-value: 9E-86
  - Accession: AGG08878

- **OR11**
  - Olfactory receptor 10
  - Helicoverpa armigera (Lep:Noc)
  - Score: 468
  - E-value: 3E-160
  - Accession: AJG42376

- **OR12**
  - Putative odorant receptor, partial
  - Sesamia inferens (Lep:Noc)
  - Score: 135
  - E-value: 5E-39
  - Accession: AYG14575

- **OR13**
  - Odorant receptor 28
  - Athetis dissimilis (Lep:Noc)
  - Score: 254
  - E-value: 9E-80
  - Accession: ALM26217

- **OR14**
  - Putative olfactory receptor 21, partial
  - Ostrinia furnacalis (Lep:Cra)
  - Score: 147
  - E-value: 8E-40
  - Accession: BAR43463

- **OR15**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 188
  - E-value: 9E-55
  - Accession: AIG51873

- **OR16**
  - Odorant receptor 21
  - Athetis dissimilis (Lep:Noc)
  - Score: 149
  - E-value: 6E-41
  - Accession: ALM26210

- **OR17**
  - Odorant receptor 30a-like (predicted)
  - Papilio machaon (Lep:Pap)
  - Score: 98.6
  - E-value: 6E-22
  - Accession: XP_014367947

- **OR18**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 345
  - E-value: 2E-114
  - Accession: AIG51887

- **OR19**
  - Odorant receptor
  - Helicoverpa armigera (Lep:Noc)
  - Score: 268
  - E-value: 9E-85
  - Accession: AIG51887

- **OR20**
  - Olfactory receptor 12, partial
  - Helicoverpa assulta (Lep:Noc)
  - Score: 106
  - E-value: 6E-27
  - Accession: AJD61550

- **OR21**
  - Putative odorant receptor, partial
  - Sesamia inferens (Lep:Noc)
  - Score: 104
  - E-value: 4E-26
  - Accession: AYG14570

- **OR22**
  - Odorant receptor 35
  - Athetis dissimilis (Lep:Noc)
  - Score: 377
  - E-value: 1E-127
  - Accession: ALM26225

- **OR23**
  - Uncharacterized protein LOC106129649 (predicted)
  - Amyelois transitella (Lep:Pyr)
  - Score: 185
  - E-value: 1E-53
  - Accession: XP_013183708

- **OR24**
  - Olfactory receptor 29
  - Manduca sexta (Lep:Sph)
  - Score: 140
  - E-value: 3E-37
  - Accession: CUQ99410

(Continued)


**DISCUSSION**

In recent years, considerable progress has been made in our understanding of insect olfaction. Antennal transcriptomes are available for insect species belonging to several orders, including Diptera, Coleoptera, and Lepidoptera (Grosse-Wilde et al., 2011; Andersson et al., 2013; Rinker et al., 2013; Leitch et al., 2015; Zhang et al., 2015). Within the Lepidoptera, the transcriptomes of model species such as *B. mori, D. plexippus, H. melpomene, H. virescens,* and *M. sexta* have been thoroughly investigated (Krieger et al., 2003, 2004; Nakagawa et al., 2005; Wanner et al., 2007; Wanner and Robertson, 2008; Tanaka et al., 2009; Briscoe et al., 2013; Koenig et al., 2015; van Schooten et al., 2016).

This knowledge is rapidly expanding to other economically important species like *Helicoverpa armigera, Cydia pomonella,* and *Spodoptera littoralis,* where it could greatly aid in improving already existing and developing new semiochemical-based management strategies (Bengtsson et al., 2012; Jacquin-Joly et al., 2012; Liu et al., 2012). This report represents an expansive characterization of the chemosensory transcripts and their encoded proteins of *L. dispar,* and increases the number of available olfactory-related sequences for Lepidopteran species of...
We identified a total of 115 putative olfactory transcripts for *L. dispar*. This number is similar to the one reported on a previous study comparing *S. littoralis* adult antennae and larval head capsules (127) (Poivet et al., 2013), and another study investigating the adult antennal transcriptome of *H. armigera*.
FIGURE 4 | Maximum likelihood dendrogram based on protein sequences of candidate odorant binding proteins (OBPs). Included are the putative sequences for Lymantria dispar (Ldis) plus those available for model Lepidoptera species Bombyx mori (Bmor), Danaus plexippus (Dple), Heliconius melpomene (Hmel), and Manduca sexta (Msex), we also included sequences from the Jumping Bristletail Lepismachila y-signata (Lsig) and the Firebrat Thermobia domestica (Tdom).

(131), and H. assulta (129) (the latter did not include GRs and we excluded sensory neuron membrane proteins from the total count) (Zhang et al., 2014). The conserved number of olfaction-related genes suggests a core group of genes control olfaction in moth species belonging to the superfamily Noctuoidea (Kristensen et al., 2007; Zahiri et al., 2011). Given the similar number of genes across these species, we could speculate that olfactory differences emerge as a product of functional
diversification, while the genes themselves are products of duplication. However, at this stage we can’t rule out specific expansions of certain gene clusters balanced by contraction in others, and the evaluation of this possibility must await a more detailed understanding of olfactory differences in the Noctuoidea.

A study investigating expression patterns between adults and larvae of *S. littoralis* found that adults and larvae express similar
numbers of OBPs and CSPs, while the caterpillar OR and IR repertoires were much smaller than the adult ones, and some GRs were found to be adult-specific (Poivet et al., 2013). We also encountered a similar number of CSPs being expressed in both stages and reduced IR, OR, and GR repertoires in the larval stages. However, in contrast to the previous study, we found that larvae had a higher CSP repertoire than adults including eight larval-specific genes. This pattern could reflect species-specific adaptations since in S. littoralis host-plant selection is mainly accomplished by adult females, who make suitable choices for the
larvae as eclosion occurs rapidly after egg laying (Anderson and Alborn, 1999; Profitt et al., 2015). In contrast, *L. dispar* eggs of all populations undergo an overwintering process accompanied by changes in the distribution and quality of the resources from oviposition until larval hatching (Barbosa et al., 1989; Sattler, 1991; Hunter, 1995). Therefore, larval stages need to make host-choices to a greater or lesser extent, which may explain the observed differences in the number of CSP genes being expressed in the larval stages.

The strong reduction in the ORs in larvae vs. adults seems to be commonplace in insects and has been reported for a number of species, including *D. melanogaster*, *Aedes aegypti*, *M. sexta*, and *B. mori* (Hallem et al., 2004; Kreher et al., 2005; Bobbot et al., 2007; Tanaka et al., 2009; Koenig et al., 2015). In *L. dispar*, this reduction is quite dramatic, with only six ORs being expressed in the larvae vs. 35 in the female antennae. Results from the PCA analysis indicate that gene expression patterns of female antennal transcriptomes from different subspecies are more similar to one another than to the larval head capsules of their respective subspecies, further supporting strong differences in chemosensory perception between adult and larval stages.

After exploring the amount of shared and exclusively expressed genes for three populations (Figure 1), we observed that AGM and EGM populations share a high number of commonly expressed genes, whereas the JGM population appears to be more divergent, having a high number (14) of uniquely expressed genes. These results suggest that the observed differences are unrelated to flight capacity, indicating either that (a) loss of flight does not impact the olfactory gene repertoire or (b) the secondary loss of flight in the American EGM populations may be too recent to have caused major changes in the genes being expressed.

Interestingly, females from the flightless EGM population display higher gene expression values (RPKM) when compared with JGM and AGM females for most olfaction-related genes except CSPs (Table 1). This could indicate that changes in transcription rates could play an important role in the adaptation of flightless females to their chemical environment. The high variability in olfactory genes and their expression in the larvae suggest that these patterns are unrelated to loss of flight, and we speculate that they are rather adaptations to different ecological pressures.

A detailed comparison of the protein sequences with those reported for other Lepidopteran species through Blast searches and phylogenetic trees supports the common ancestry and high degree of conservation for most olfaction-related gene families within the Lepidoptera, and reveals a high sequence similarity between *L. dispar* and other members of the Noctuidae clade, particularly for ORs and GRs. A recent study investigating the evolution of these chemoreceptors in the Lepidoptera suggests that the common ancestor of this clade harbored only few OR and GR genes, and that while the number of genes increased greatly during the evolution of the clade, it remained relatively low in comparison to other insect groups. This high degree of conservation possibly occurred because olfaction-related gene expression in the Lepidoptera is under strict regulatory control, limiting the establishment of newly emerged genes (Engsontia et al., 2014).

Although most of our candidate sequences had close alignments with those reported for other model species, a few cases remain where *L. dispar* sequences were observed to form clusters or single nodes. Further studies are required to confirm the identity of these sequences and establish whether lineage-specific gene expansion occurs in the Lymantridiinae clade (including closely related species such as the Douglas-fir tussock moth *Orgyia pseudotsugata* and the nun moth *Lymantria monachus*) or the superfamily Noctuoidea (including more distantly related species such as *Spodoptera* spp. and *Helicoverpa* spp.).

**CONCLUSIONS**

This work represents the most complete description of chemosensory genes and proteins for *L. dispar* to date. Our results reveal differential gene expression between adult and larval stages characterized by fewer IR, OR, and GR genes being expressed in the larvae, but more CSP genes in comparison to the adults. Comparisons of protein sequences with those from other Lepidopteran species and organisms from different taxa support the common ancestry and high degree of conservation for most olfaction-related gene families. The gene expression patterns in female antennae are more similar to one another than they are to their respective larval stages, whereas larval gene expression patterns are highly divergent across populations. After exploring the number of unique and commonly expressed genes, AGM and EGM populations were found to share a high number of commonly expressed genes, whereas the JGM population appeared to be more divergent. These results indicate that either (a) loss of flight does not impact the olfactory gene repertoire or (b) the secondary loss of flight in American EGM populations may be too recent to cause major changes in the genes being expressed. Nevertheless, higher expression values for GRs, IRs, OBPs, and ORs in EGM females suggest that differences in transcription rates could be an adaptation of flightless females to their chemical environment. Differences in the larval olfactory-related gene expression, on the other hand, are likely responses to unique ecological pressures rather than to female flight ability. Further studies are required to understand the deeper evolutionary and ecological significance of these findings.

**AUTHOR CONTRIBUTIONS**

AC, EG, CD, MM, and BH designed research; AC and EG collected data; AC, EG, and DW analyzed data; all authors contributed to the writing process. All authors read and approved the submitted version.

**FUNDING**

This project was possible thanks to the funding provided to AC by the International Postdoctoral Fellowship.
ACKNOWLEDGMENTS

We would like to express our gratitude to Hannah Nadel, Supervisory Entomologist of the Animal and Plant Health Inspection Service (APHIS) of the United States Department of Agriculture (USDA) for providing the insects used in this study. We would also like to acknowledge the reviewers for their constructive criticism and valuable suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/nevo.2017.00115/full#supplementary-material


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Clavijo McCormick, Grose-Wilde, Wheeler, Mescher, Hansson and De Moraes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.