



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2013

Evolution and functional characterisation of the IR chemosensory receptors in the *Drosophila* larval gustatory system

Croset Vincent

Croset Vincent, 2013, Evolution and functional characterisation of the IR chemosensory receptors in the *Drosophila* larval gustatory system

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive.

<http://serval.unil.ch>

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Centre Intégréatif de Génomique

Evolution and functional characterisation of the IR chemosensory receptors in the *Drosophila* larval gustatory system

Thèse de doctorat ès sciences de la vie (PhD)

Présentée à la Faculté de Biologie et Médecine
de l'Université de Lausanne

par

Vincent CROSET

Master en Génomique et Biologie Expérimentale
de l'Université de Lausanne

Jury

Prof. Antoine Guisan, Président
Prof. Richard Benton, Directeur de thèse
Prof. Bertram Gerber, Expert
Prof. Bruno Lemaitre, Expert

Lausanne, February 2013

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

<i>Président</i>	Monsieur Prof. Antoine Guisan
<i>Directeur de thèse</i>	Monsieur Prof. Richard Benton
<i>Experts</i>	Monsieur Prof. Bruno Lemaitre
	Monsieur Prof. Bertram Gerber

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Vincent Croset

Master of Science de l'Université de Lausanne


intitulée

**Evolution and functional characterisation
of the IR Chemosensory receptors
in the *Drosophila* larval gustatory system**

Lausanne, le 12 avril 2013

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Antoine Guisan



"Time flies like an arrow; fruit flies like a banana."

Groucho Marx

Abstract

Chemosensory receptor gene families encode divergent proteins capable of detecting a huge diversity of environmental stimuli that are constantly changing over evolutionary time as organisms adapt to distinct ecological niches. While olfaction is dedicated to the detection of volatile compounds, taste is key to assess food quality for nutritional value and presence of toxic substances. The sense of taste also provides initial signals to mediate endocrine regulation of appetite and food metabolism and plays a role in kin recognition.

The fruit fly *Drosophila melanogaster* is a very good model for studying smell and taste because these senses are very important in insects and because a broad variety of genetic tools are available in *Drosophila*. Recently, a family of 66 chemosensory receptors, the Ionotropic Receptors (IRs) was described in fruit flies. IRs are distantly related to ionotropic glutamate receptors (iGluRs), but their evolutionary origin from these synaptic receptors is unclear. While 16 IRs are expressed in the olfactory system, nothing is known about the other members of this repertoire.

In this thesis, I describe bioinformatic, expression and functional analyses of the IRs aimed at understanding how these receptors have evolved, and at characterising the role of the non-olfactory IRs. I show that these have emerged at the basis of the protostome lineage and probably have acquired their sensory function very early. Moreover, although several IRs are conserved across insects, there are rapid and dramatic changes in the size and divergence of IR repertoires across species. I then performed a

comprehensive analysis of IR expression in the larva of *Drosophila melanogaster*, which is a good model to study taste and feeding mechanisms as it spends most of its time eating or foraging. I found that most of the divergent members of the IR repertoire are expressed in both peripheral and internal gustatory neurons, suggesting that these are involved in taste perception. Finally, through the establishment of a new neurophysiological assay in larvae, I identified for the first time subsets of IR neurons that preferentially detect sugars and amino acids, indicating that IRs might be involved in sensing these compounds.

Together, my results indicate that IRs are an evolutionarily dynamic and functionally versatile family of receptors. In contrast to the olfactory IRs that are well-conserved, gustatory IRs are rapidly evolving species-specific receptors that are likely to be involved in detecting a wide variety of tastants.

Résumé en français

La plupart des animaux possèdent de grandes familles de récepteurs chimiosensoriels dont la fonction est de détecter l'immense diversité de composés chimiques présents dans l'environnement. Ces récepteurs évoluent en même temps que les organismes s'adaptent à leur écosystème. Il existe deux manières de percevoir ces signaux chimiques : l'olfaction et le goût. Alors que le système olfactif perçoit les composés volatiles, le sens du goût permet d'évaluer, par contact, la qualité de la nourriture, de détecter des substances toxiques et de réguler l'appétit et le métabolisme. L'un des organismes modèles les plus pertinents pour étudier le sens du goût est le stade larvaire de la mouche du vinaigre *Drosophila melanogaster*. En effet, la principale fonction du stade larvaire est de trouver de la nourriture et de manger. De plus, il est possible d'utiliser tous les outils génétiques développés chez la drosophile.

Récemment, une nouvelle famille de 66 récepteurs chimiosensoriels appelés Récepteurs Ionotropiques (IRs) a été découverte chez la drosophile. Bien que leur origine soit peu claire, ces récepteurs sont similaires aux récepteurs ionotropiques glutamatergiques impliqués dans la transmission synaptique. 16 IRs sont exprimés dans le système olfactif de la mouche adulte, mais pour l'instant on ne connaît rien des autres membres de cette famille.

Durant ma thèse, j'ai effectué des recherches sur l'évolution de ces récepteurs ainsi que sur l'expression et la fonction des IRs non olfactifs. Je démontre que les IRs sont apparus chez l'ancêtre commun des

protostomiens et ont probablement acquis leur fonction sensorielle très rapidement. De plus, bien qu'un certain nombre d'IRs olfactifs soient conservés chez les insectes, d'importantes variations dans la taille et la divergence des répertoires d'IRs entre les espèces ont été constatées. J'ai également découvert qu'un grand nombre d'IRs non olfactifs sont exprimés dans différents organes gustatifs, ce qui leur confère probablement une fonction dans la perception des goûts. Finalement, pour la première fois, des neurones exprimant des IRs ont été identifiés pour leur fonction dans la perception de sucres et d'acides aminés chez la larve.

Mes résultats présentent les IRs comme une famille très dynamique, aux fonctions très variées, qui joue un rôle tant dans l'odorat que dans le goût, et dont la fonction est restée importante tout au long de l'évolution. De plus, l'identification de neurones spécialisés dans la perception de certains composés permettra l'étude des circuits neuronaux impliqués dans le traitement de ces informations.

Résumé en français pour large public

Tous les animaux sont capables de percevoir des signaux chimiques provenant de leur environnement. Grâce à l'odorat, ils peuvent détecter des molécules présentes dans l'air ; et grâce au goût, ils évaluent la qualité de leur nourriture. Pour être capable de percevoir et de distinguer une multitude de molécules différentes, les individus de chaque espèce possèdent un grand nombre de récepteurs dans leurs organes olfactifs et gustatifs. Ces récepteurs diffèrent en fonction des différents composés que chaque espèce rencontre dans son quotidien. Une abeille n'aura donc pas les mêmes récepteurs qu'un poisson ou qu'un ver de terre.

La mouche du vinaigre, ou drosophile, est un modèle de choix pour l'étude de la perception des goûts et des odeurs. En effet, en tant qu'organisme modèle, de nombreux outils permettant de manipuler aisément ses gènes et ses cellules ont été développés. De plus, la larve de drosophile passe presque tout son temps à manger, et la manière dont elle perçoit le goût est donc particulièrement intéressante à étudier.

Durant ma thèse, je me suis penché sur l'une de ces familles de récepteurs, appelés IRs, chez la drosophile. Bien qu'ils aient été identifiés chez cet animal, la question se posait de savoir s'ils étaient également présents chez d'autres espèces. De plus, bien que l'on sache qu'une partie de ces récepteurs fonctionne comme récepteurs olfactifs, la fonction des autres membres de cette famille reste encore inconnue.

J'ai tout d'abord recherché les IRs chez un grand nombre d'espèces, ce qui a permis de montrer que ceux-ci sont présents chez la majorité des

invertébrés, et qu'ils sont apparus très tôt lors de l'évolution, il y a environ 550 millions d'années. Deuxièmement, une grande partie de ces IRs se trouvent dans les organes gustatifs de la mouche et sont donc certainement impliqués dans la perception des goûts. Finalement, certains de ces IRs sont présents dans des cellules détectant principalement des sucres ou des acides aminés. Cela permet donc de supposer que les IRs sont impliqués dans la perception de ces composés. Mes résultats ont donc permis de saisir l'importance des IRs chez les invertébrés, ainsi que de mieux comprendre leur fonction.

Table of Contents

Abstract	3
Résumé en français	5
Résumé en français pour large public	7
Table of Contents	9
Chapter 1: Introduction	11
Smell and taste as two modalities for sensing chemicals in the environment.....	12
Chemosensation in mammals.....	13
Chemosensation in nematodes.....	15
Olfaction in <i>Drosophila</i>	16
<i>Olfactory organs and receptors</i>	16
<i>Organisation of the Drosophila olfactory system</i>	16
Taste perception in <i>Drosophila</i>	18
<i>Taste organs in Drosophila</i>	18
<i>Drosophila gustatory receptors</i>	20
<i>The suboesophageal ganglion: the primary gustatory centre</i>	21
The IRs: a novel family of chemosensory receptors	22
<i>Expression and function of olfactory IRs</i>	22
<i>Functional architecture of the IRs</i>	23
<i>Behavioural function of IRs</i>	26
Chapter 2: The Evolution of IRs	28
Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction	28
<i>Summary of results</i>	28
<i>My contribution to this work</i>	30
IRs in insect genomes.....	31
<i>Summaries of results</i>	31
<i>My contribution to this work</i>	31
Chapter 3: A map of IR expression in the larval chemosensory system 32	
Introduction	32
<i>The larva of Drosophila melanogaster: a model for the study of taste perception</i>	32
<i>Olfaction in Drosophila larvae</i>	33
<i>Taste perception in Drosophila larvae</i>	34
Production of IR-Gal4 lines	35
Expression of IRs in larval chemosensory organs	36
Central projections of IR neurons.....	40
Receptor co-expression in gustatory neurons.....	43
Discussion.....	47
<i>Non-olfactory IRs are candidate gustatory receptors</i>	47
<i>Perspectives</i>	49
Material and Methods.....	49
<i>IR Gal4 transgenes</i>	49
<i>Histology</i>	50

Chapter 4: Behavioural characterisation of IR25a mutants	52
Introduction	52
<i>A behavioural assay for testing gustatory preference</i>	<i>53</i>
<i>IR25a mutant: a mutant lacking most IR functions?</i>	<i>53</i>
IR25a is not involved in the detection of sugars and bitter compounds	54
Reduced preference for acidic media in IR25a mutants	56
IR25a plays a minor role in modulating amino acid sensing	58
Discussion	61
Material and methods	62
<i>Fly strains</i>	<i>62</i>
<i>Larval taste-choice assays</i>	<i>63</i>
Chapter 5: Identification of IR neurons that detect sugars and amino acids	64
Introduction	64
<i>Two methods for measuring neuronal activity in vivo</i>	<i>64</i>
Establishment of calcium imaging methods for adults and larvae	66
<i>IR76b neurons respond to sucrose in adults</i>	<i>66</i>
<i>Calcium imaging on larval SOG</i>	<i>68</i>
Sugar-sensing IR neurons	69
<i>A subset of IR76b neurons selectively respond to sucrose and fructose</i>	<i>69</i>
<i>Identification of IR7e neurons that are sucrose-sensing IR neurons</i>	<i>73</i>
<i>IR76b and IR7e neurons define two different sugar-sensing pathways</i>	<i>75</i>
Amino acids activate IR neurons	76
<i>IR76b neurons respond to amino acids</i>	<i>76</i>
<i>IR60c neurons are activated by the same amino acids as IR76b neurons</i>	<i>78</i>
Discussion and perspectives	80
<i>Optical imaging in the Drosophila larval gustatory system</i>	<i>80</i>
<i>Are IRs involved in sugar perception?</i>	<i>81</i>
<i>IRs as candidate amino acid receptors</i>	<i>83</i>
Material and Methods	84
<i>Fly strains</i>	<i>84</i>
<i>Calcium imaging in adults</i>	<i>85</i>
<i>Calcium imaging in larvae</i>	<i>86</i>
<i>Calcium imaging data analysis</i>	<i>86</i>
Acknowledgements	88
References	90

Chapter 1: Introduction

All organisms need to detect an enormous number of chemicals in their environment in order to recognise nutrients, toxins, preys, predators and conspecifics. A challenge for chemosensory systems is to be able to distinguish between these various chemicals and transmit suitable information to the organism, in order to induce proper behavioural, metabolic or developmental responses. To avoid sensing irrelevant cues and thus to minimise the cost of chemosensory perception, the range of chemicals that can be detected is highly dependent on the ecology of each species and of the developmental and physiological state of the individuals, and is thus likely to fit to their needs. Because of their ecological relevance and impact on a species' behaviour and physiology, chemosensory systems represent one of the major forces of evolution. In addition, chemosensory systems are a good model to study how neuronal networks are assembled and how environmental stimuli are detected, processed, and transduced.

My thesis aims at a better understanding of the mechanisms of chemosensation using the fruit fly *Drosophila melanogaster* as a model system and focussing on a family of receptors called the Ionotropic Receptors (IRs). *Drosophila* has been used extensively in genetic studies for almost a century. In addition to a short generation time, its genome was sequenced and is well annotated, and it has the advantage of having a broad panel of genetic tools that allow thorough molecular and cellular experiments *in vivo*. Moreover, although the global organisation of the

chemosensory systems of fruit flies has many similarities with those of mammals, they is numerically much simpler, thus easier to construe.

In this introduction, I will first discuss the differences and similarities between smell and taste, which are the two ways that animals use to detect chemicals. I will then briefly explain their main features in other model organisms before focussing on their function in *Drosophila*. Finally, I will introduce the Ionotropic Receptors (IRs) as a novel family of chemosensory receptors. In the following chapters, I will present my studies on their evolution, as well as their function as gustatory receptors in *Drosophila* larvae.

Smell and taste as two modalities for sensing chemicals in the environment

Because of the high diversity across animal clades, there is no universal definition of what smell and taste are and of what features differentiate them. Both happen when a receptor protein located in the membrane of a sensory cell detects a chemical. This binding triggers excitation of this cell, which then activates further neuronal processes. Although in humans smell occurs in the nose and taste on the tongue, this is impossible to generalise for all animals, and in many of them these two senses are not even differentiated. However, chemical, molecular and anatomical properties can often distinguish smell from taste. First, the chemical state of olfactory and gustatory ligands is different. Whereas taste usually requires a direct contact with a liquid or solid tastant, olfaction is

dedicated to the detection of volatile gaseous compounds. However, although this is true for terrestrial animals, it does not apply for aquatic ones. Second, as olfactory and gustatory ligands are usually different, the receptors for smell and taste belong to distinct (though sometimes related) families. Finally, in many organisms, the olfactory and gustatory nerves do not innervate the same region of the brain, allowing olfactory and gustatory stimuli to be processed in different ways. In order to illustrate these differences, I will describe the chemosensory systems of three major groups of animals: mammals, nematodes and insects.

Chemosensation in mammals

Vertebrates have five types of olfactory receptors and three types of taste receptors (Figure 1). Although most of them belong to the G protein-coupled receptor (GPCR) family, they all evolved independently from each other (Nei et al., 2008). Amongst receptors detecting odours, the most numerous and the first ones to be discovered are the Odorant Receptors (ORs), which are expressed in the main olfactory epithelium (MOE) (Buck and Axel, 1991). Trace amine-associated receptor (TAARs) are also expressed in the MOE and are thought to be involved in the detection of social pheromones (Figure 1) (Liberles and Buck, 2006). Most mammals also have a second organ dedicated to pheromone detection, called the vomeronasal organ, and expressing two types of vomeronasal receptors: V1Rs (Dulac and Axel, 1995) and V2Rs (Herrada and Dulac, 1997; Matsunami and Buck, 1997). From the MOE, olfactory nerves drive sensory information to the main olfactory bulb (MOB) located slightly more dorsally,

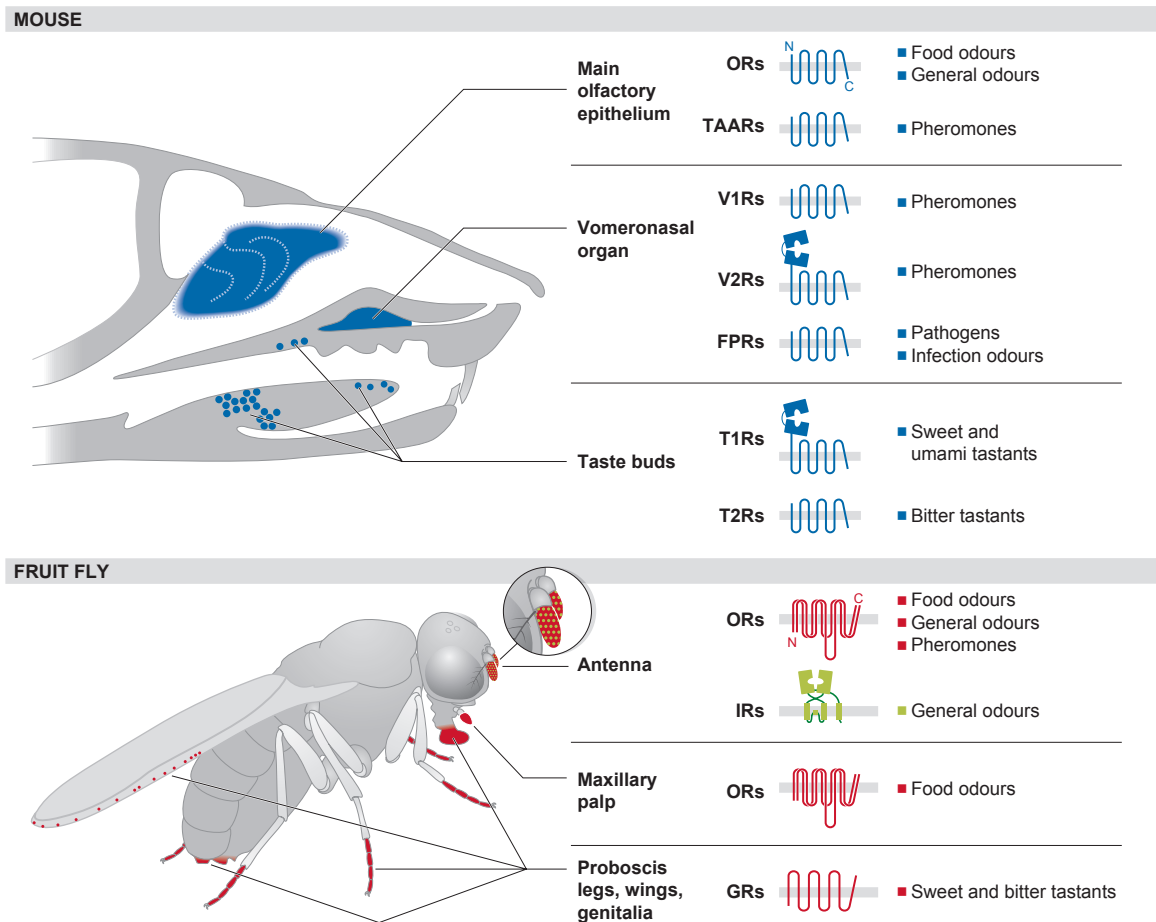


Figure 1 | The main chemosensory organs, receptors and putative ligands in the mouse and the fruit fly. The figure was adapted from Silbering & Benton (2010), the image of the mouse head was adapted from Matsunami & Amrein (2003). FPRs, formyl peptide receptors; GRs, gustatory receptors; IRs, ionotropic receptors; ORs, odorant receptors; T1Rs, taste receptors type 1; T2Rs, taste receptors type 2; TAARs, trace amine-associated receptors; V1Rs, vomeronasal receptors type 1; V2Rs, vomeronasal receptors type 2.

whereas vomeronasal neurons project to the accessory olfactory bulb (AOB) in the same region. There, information is processed and further reaches higher brain centres including the piriform cortex and the amygdala (Lledo et al., 2005).

One particularity of mammals is that tastants are not directly detected by sensory neurons, but by taste buds made of hair cells of epithelial origin. These express three types of receptors, which are responsible for taste perception. T1Rs detect appetitive cues, such as sweet and umami tastants (Li et al., 2002) and T2Rs bitter compounds (Figure 1) (Adler et al., 2000; Matsunami et al., 2000), whereas salts and acids are detected by ion channels (Miyamoto et al., 1998). The cranial nerves VII, IX and X innervate these taste buds and transmit gustatory information further to the gustatory cortex.

Chemosensation in nematodes

In invertebrates like the nematode worm *Caenorhabditis elegans*, smell and taste are not separated. The non-motile cilia of amphid neurons ensure both olfactory and gustatory functions. Each of them is dedicated to particular sensory cues that can be volatile or water-soluble environmental chemicals, pheromones or even temperature (Bargmann, 2006; Inglis et al., 2007). Chemosensory neurons express a broad panel of more than 1200 receptors (Robertson and Thomas, 2006), which implies many receptors to be expressed in the same cell. All axons from these amphid neurons project to the nerve ring, where sensory information is integrated (Ware et al., 1975).

Olfaction in *Drosophila*

Olfactory organs and receptors

In *Drosophila*, odours are recognised by receptors expressed by olfactory sensory neurons (OSNs) that project their dendrites to porous hairs called sensilla. These are located at the surface of the antenna or the maxillary palp, inside the sacculus or on the arista (Vosshall and Stocker, 2007) (Figure 3A). There are three types of sensilla. The basiconic and trichoid sensilla host neurons expressing olfactory receptors (ORs) (Clyne et al., 1999; Gao and Chess, 1999; Vosshall et al., 1999), whereas neurons from the smaller coeloconic sensilla express Ionotropic Receptors (IRs; Figure 3A, see below) (Benton et al., 2009). ORs are a family of ~60 proteins. Like GPCRs, ORs have seven transmembrane domains. However, their topology is inverted compared to GPCRs, with a cytoplasmic N-terminus (Benton et al., 2006). In addition, they are independent of G-proteins and unlike vertebrate ORs they function as ion channels (Sato et al., 2008; Wicher et al., 2008). Each OSN normally expresses a single OR, in addition to the broadly expressed co-receptor ORCO that is present in all OR-expressing neurons (Couto et al., 2005; Larsson et al., 2004).

*Organisation of the *Drosophila* olfactory system*

OSNs project their axons to the primary olfactory centre located in the frontal brain, called the antennal lobe (AL), which is formed of an assembly of glomeruli (Figure 2B, Figure 3A). OSNs expressing the same receptors

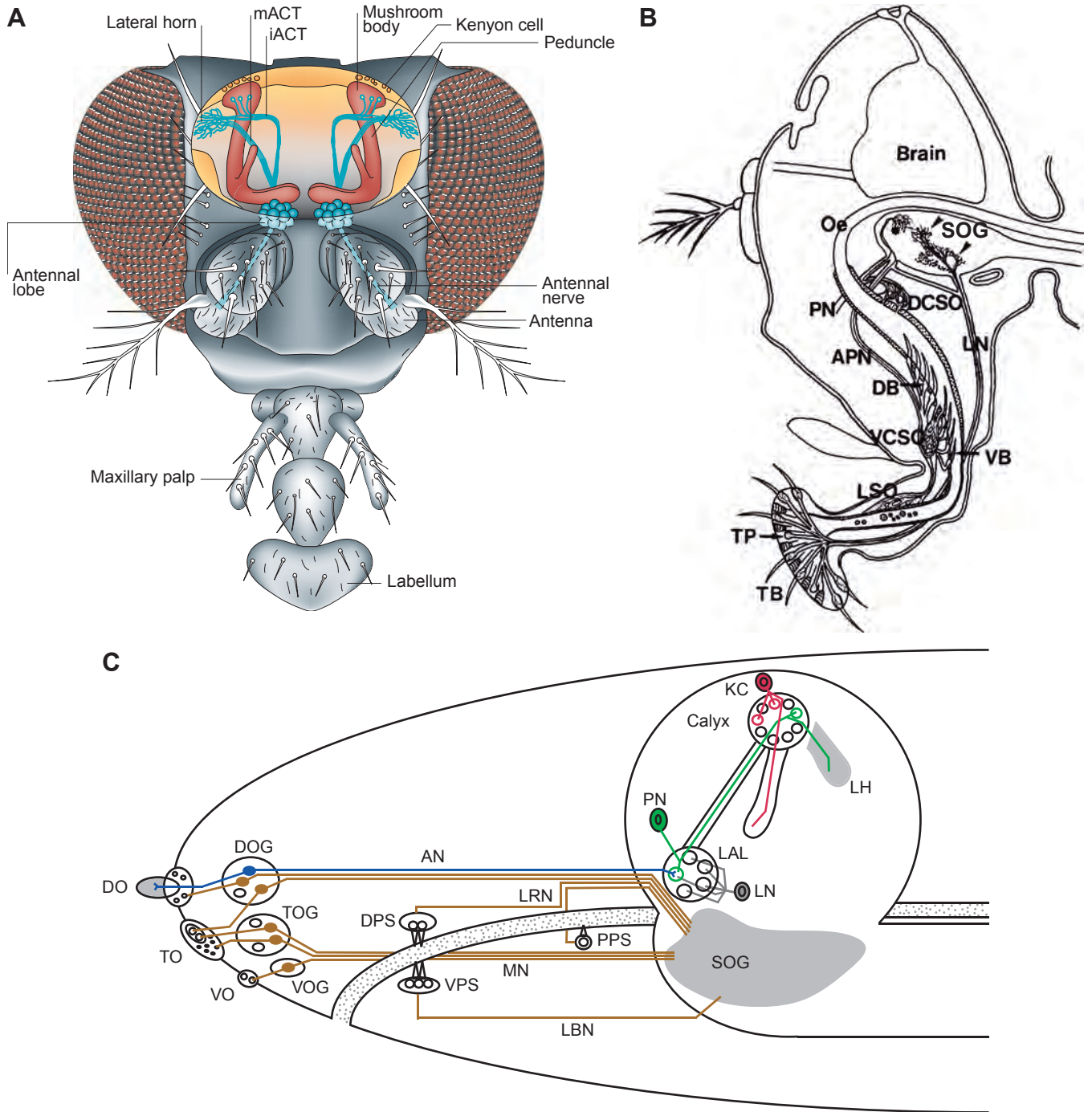


Figure 2 | Organisation of the *Drosophila* olfactory and gustatory systems. (A) Dorsal view of the main olfactory pathways in adult flies. Odours are sensed by OSNs in the antenna that project to the antennal lobe. From there, projection neurons drive information to the higher brain centres. mACT: medial antennocerebral tract, iACT: inner antennocerebral tract. Adapted from Keene & Waddell (2007). (B) Gustatory pathways in adult flies. Taste bristles (TB) and taste pegs (TP) sense chemicals peripherally. The labral nerve (LN) drives the information to the subesophageal ganglion (SOG). The labral (LSO), ventral cibarial (VCSO, with dorsal (DB) and ventral (VB) bristles) and dorsal cibarial (DCSO) sensory organs sense chemicals inside the pharynx and connect the SOG with the pharyngeal (PN) and accessory pharyngeal (APN) nerves. Adapted from Stocker & Schorderet (1981). (C) Olfactory and gustatory pathways in larvae. Odours are sensed by neurons from the dorsal organ ganglion (DOG) in the dorsal organ (DO). Neurons project to the larval antennal lobe (LAL) through the antennal nerve (AN), where they synapse with local (LN) and projection (PN) neurons. PN then drive information to the Kenyon cells (KC) of the mushroom body and to the lateral horn (LH). Tastants are sensed peripherally in the terminal (TO) and ventral (VO) organs by neurons that have their somata in the terminal (TOG) and ventral (VOG) organ ganglia, and sometimes also in the DOG. The dorsal (DPS), ventral (VPS) and posterior (PPS) pharyngeal sensilla are responsible for internal gustatory sensing. Gustatory neurons connect to the SOG through the labral (LRN), maxillary (MN) and labial (LBN) nerves. Adapted from Vosshall & Stocker (2007)

innervate single glomerulus (Fishilevich and Vosshall, 2005; Gao et al., 2000; Hallem et al., 2004), which means that odor processing is anatomically segregated at this stage. Glomeruli are connected to each other by local interneurons that play a generally inhibitory role in modulating odour responses (Chou et al., 2010; Ng et al., 2002). Projection neurons (PNs) synapse with OSNs within particular glomeruli and transmit olfactory information to the protocerebrum, where higher brain centres are located. PNs in the inner antennocerebral tract (iACT) project to the mushroom body (MB) calyx and the lateral horn (LH), whereas the medial antennocerebral tract goes directly to the LH. The MB is responsible for olfactory memory and acquired behaviours (Figure 3A) (Heisenberg et al., 1985; Keene and Waddell, 2007), whereas the LH is thought to be where innate responses to odours are processed (Figure 3A) (Jefferis et al., 2007; Tanaka et al., 2004).

Taste perception in *Drosophila*

Taste organs in Drosophila

In adults, the main gustatory sensory organ is the labellum (Figure 2B), which is located at the tip of the proboscis (Stocker, 1994; Stocker and Schorderet, 1981). As in the olfactory system, there are taste sensilla on the labellum that host the dendrites of gustatory sensory neurons (GSNs) (Shanbhag et al., 2001), but also of mechanosensory and thermosensory neurons (Ishimoto and Tanimura, 2004). There are three types of taste sensilla, which host GSNs that detect different classes of compounds.

Large (L-type) sensilla host four neurons, one to three of them expressing sweet-sensing receptors and responding to sugars, the other neurons detecting water and osmolarity (Amrein and Thorne, 2005; Cameron et al., 2010; Hiroi et al., 2002). Small (S-type) sensilla also have four neurons, and in most of them one neuron senses bitter compounds, another sweet tastants, and the two others water and osmolarity (Amrein and Thorne, 2005; Weiss et al., 2011). Finally, intermediate (I-type) sensilla have only two gustatory neurons: one for sweet and one for bitter compounds (Amrein and Thorne, 2005; Hiroi et al., 2002; Weiss et al., 2011). In addition to sensilla, taste pegs are also present on the labellum, and are thought to be involved in the detection of carbonation (Fischler et al., 2007).

Unlike mammals, insects have more than one gustatory organ. In addition to the labellum, internal organs are also known to detect post-ingested tastants (Stocker and Schorderet, 1981). The labral, ventral cibarial and dorsal cibarial organs are distributed along the pharynx of adult flies (Figure 2B). Their specific function however remains unknown. Furthermore, flies are able to detect tastants through specialised neurons distributed on most of their appendages, such as the legs, wings or female ovipositor (Stocker, 1994). GSNs on the legs sense sugars and are able to induce the extension reflex of the proboscis (Falk and Atidia, 1975; Gordon and Scott, 2009). They also detect female pheromones necessary for courtship behaviour (Bray and Amrein, 2003). GSNs on the ovipositor play a role in the choice of the oviposition site (Yang et al., 2008)

Drosophila gustatory receptors

In *Drosophila*, Gustatory Receptors (GRs) are a family of 68 proteins encoded by a total of 60 genes. Past studies demonstrated many of them to be necessary for taste perception (Montell, 2009), however two of them have an olfactory function in the antenna (Dunipace et al., 2001; Jones et al., 2007; Kwon et al., 2007; Turner and Ray, 2009). In the gustatory system of adult flies, GRs detecting bitter and sweet taste, pheromones (Bray and Amrein, 2003; Miyamoto and Amrein, 2008) have been identified. In addition to GRs, members from the *pickpocket* DEG/ENaC channel family mediate detection of water, salt, osmolarity and pheromones (Cameron et al., 2010; Liu et al., 2003a).

Gr66a is responsible for caffeine detection and broadly marks bitter-sensing neurons (Thorne et al., 2004; Wang et al., 2004). Two other receptors that are co-expressed with Gr66a are activated by caffeine (Gr93a) (Lee et al., 2009) or in a more broader way by most bitter tastants (Gr33a) (Moon et al., 2009). Gr5a is a receptor for trehalose (Chyb et al., 2003; Dahanukar et al., 2001; Ueno et al., 2001) and marks most sweet-sensing neurons, without overlapping with Gr66a. Seven other GRs are involved in sugar perception, six of which are members of the Gr64a-f cluster, which is expressed as a poly-cistronic mRNA (Slone et al., 2007). In this cluster, Gr64a detects sucrose, maltose and glucose (Dahanukar et al., 2007), Gr64e glycerol (Wisotsky et al., 2011), whereas Gr64f is likely to act as a co-receptor along with other sweet-sensing GRs (Jiao et al., 2007).

The suboesophageal ganglion: the primary gustatory centre

Peripheral and internal GSNs from the proboscis all project to the suboesophageal ganglion (SOG), which is the primary gustatory centre in the brain (Stocker and Schorderet, 1981). GSNs from thoracic and abdominal segments project primarily to the ventral nerve cord (VNC), from where second-order neurons can drive the information further to the SOG or directly induce reflex behaviours (Bader et al., 2007; Park and Kwon, 2011).

Unlike the antennal lobe, the SOG is not divided into glomeruli, although some internal sub-structures have been described (Miyazaki and Ito, 2010). However, different GSNs can have very different projection patterns depending on which nerve drives their axons to the SOG and, more importantly, on the quality of the tastant that they detect. Hence, GR66a neurons, that detect caffeine and other aversive compounds, project to the medial part of the SOG, where they form a ring-like structure, whereas sweet-sensing GR5a neurons send their axons more anteriorly and more laterally (Marella et al., 2006; Thorne et al., 2004; Wang et al., 2004). This demonstrates that discrimination between appetitive and aversive cues already occurs at the level of the SOG. Its different areas are involved in the processing of opposite signals and thus likely to induce opposite behaviours.

The further components of the gustatory pathways are mostly unknown. However, GSNs eventually connect (directly or not) to modulatory monoaminergic neurons that are involved in various behaviours such as courtship, learning and memory or feeding (Keene and Waddell, 2007; Koganezawa et al., 2010; Marella et al., 2012). In other insects such as

honeybees, neurons that connect the SOG to MBs have been identified (Schroter et al., 2007), suggesting that they are likely to exist in *Drosophila* as well.

The IRs: a novel family of chemosensory receptors

Expression and function of olfactory IRs

The observation that several chemosensory neurons do not express ORs or GRs, such as the coeloconic sensilla (ac) of the antenna, suggested that other chemosensory proteins might exist. These proteins were discovered in 2009 and are called Ionotropic Receptors (IRs) (Benton et al., 2009), due to their high similarity with ionotropic glutamate receptors (see below and Chapter 2).

IRs are a family of 66 genes, and 16 of them have been shown to be expressed in the coeloconic sensilla of adult antennae, whereas nothing is known about the other ones. Coeloconic sensilla are thus divided in four classes with different electrophysiological response profiles, and that host neurons expressing specific combinations of IRs (Figure 3C) (Benton et al., 2009; Yao et al., 2005). In addition to coeloconic sensilla, IRs also have been identified in the arista and the sacculus, which are two specialised structures on the antenna of unclear function (Benton et al., 2009).

Two IRs (IR25a and IR8a) are more broadly expressed, together with other IRs, and consistently, they have been shown to function as co-receptors (see below). Single-sensillum electrophysiological recordings showed that IRs respond to ligands that are usually not detected by ORs,

mostly amines, carboxylic acids and aldehydes (de Bruyne et al., 2001; Hallem and Carlson, 2006; Silbering et al., 2011). Similarly to OR-expressing ORNs, neurons that express IRs project their axons to the antennal lobe. However, they innervate glomeruli that are not innervated by ORs and that are located on the posterior side of the AL (Figure 3B) (Couto et al., 2005; Silbering et al., 2011). Local interneurons connect IR and OR glomeruli, suggesting that interaction between their respective signals is likely to happen at the level of the antennal lobe (Chou et al., 2010). PNs connecting IR glomeruli project their axons to the MB and LH indistinctly from the OR ones (Silbering et al., 2011)

Functional architecture of the IRs

The overall molecular and structural organization of IRs is very similar to the one of ionotropic glutamate receptors (iGluRs), which have been extensively described (Mayer, 2006; Mayer and Armstrong, 2004) for their function in synaptic transmission in animals, as receptors for variant forms of the excitatory neurotransmitter glutamate. All iGluRs and IRs share two distinct segments (S1 and S2) forming a ligand-binding domain (Armstrong et al., 1998; Stern-Bach et al., 1994), and an ion-channel domain formed of three transmembrane alpha-helices (M1, M2 and M3) and a pore-loop (P) (Kuner et al., 2003; Panchenko et al., 2001) (Figure 1). iGluRs also have an extracellular amino-terminal domain (ATD) involved in channel assembly (Ayalon et al., 2005) and binding of co-factors (Masuko et al., 1999; Paoletti et al., 1997), that is absent from most IRs but the two co-receptors IR8a and

IR25a. These also show a higher sequence similarity to iGluRs than other IRs (Benton et al., 2009; Croset et al., 2010).

Despite the fact that they operate with different effectors, the similarity on domain organization between iGluRs and IRs suggests similarities in channel assembly and function. Using two IRs, IR8a and IR84a, it has been shown that they assemble as heterotetramers. Indeed, when these two IRs are expressed together, they elicit a response to the fruit odours phenylacetaldehyde and phenylacetic acid. The loss of either of these receptors strongly impairs that response, as well as the dendritic localisation of the associated receptor, demonstrating that both the co-receptor and the odor-specific partner are required for proper cilia targeting and channel function (Abuin et al., 2011). Ectopic reconstitution of the IR8a-IR84a complex was achieved in both *Drosophila* OR neurons and *Xenopus* oocytes, showing these receptors to be sufficient to drive odor response (Abuin et al., 2011). Furthermore, it was shown with a photobleaching method coupled with total internal reflection fluorescence microscopy (TIRF) (Ulbrich and Isacoff, 2007, 2008) that the IR8a-IR84a complex contains up to two subunits of each member and thus assembles as a dimer-of-dimers (Abuin et al., 2011).

However, more complex structures have been observed, in which the presence of three IRs (IR25a, IR76a and IR76b) in the same OSN is necessary to drive a proper response to phenylethyl amine (Abuin et al., 2011). This suggests that some IRs can also require two co-receptors, although the functional differences between IRs having one or two co-receptors remain unknown.

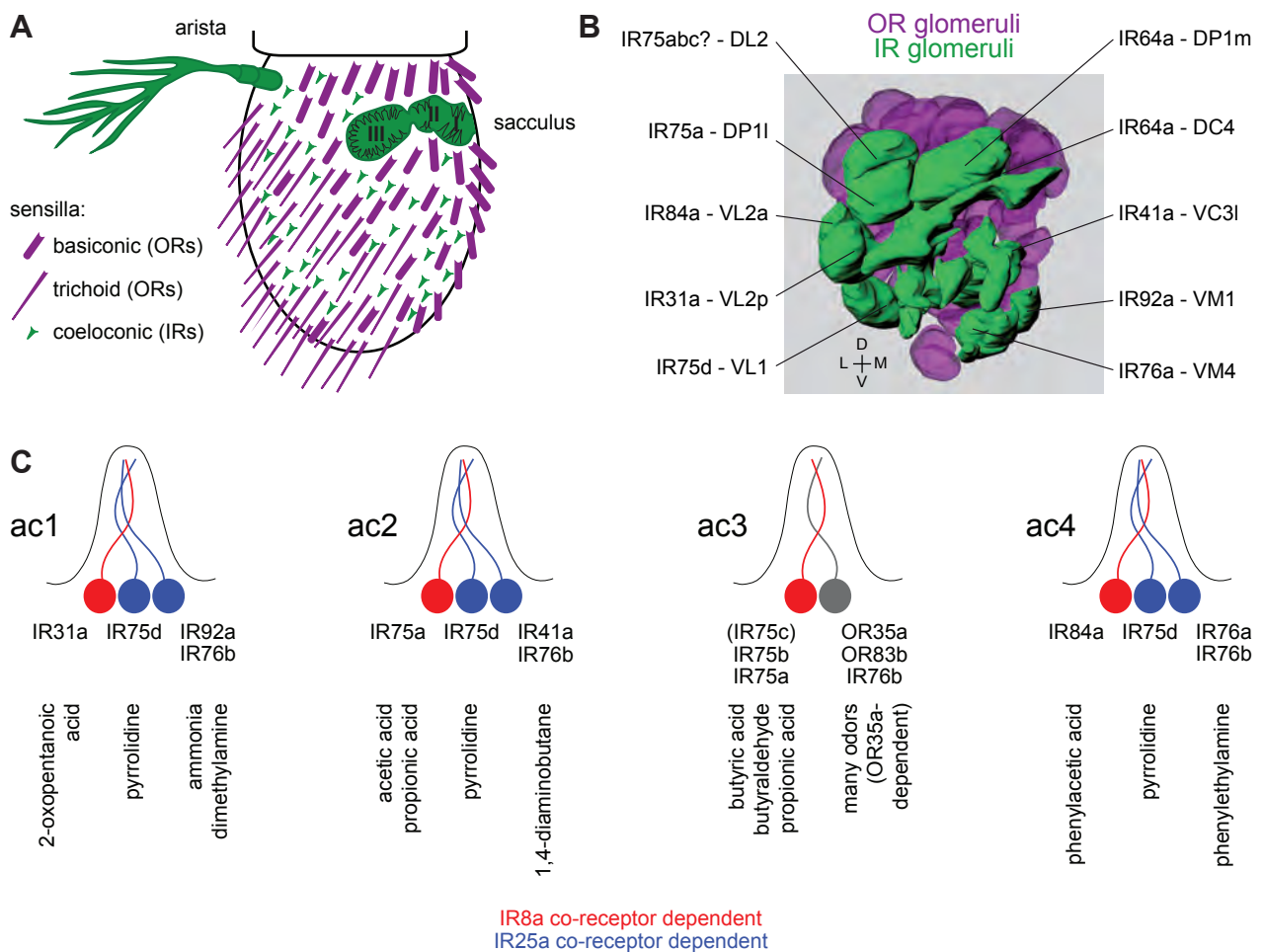


Figure 3 | The function of IRs in the *Drosophila* olfactory system. (A) Scheme of a *Drosophila* antenna showing basiconic and trichoid sensilla expressing ORs in purple, and the IR-expressing coeloconic sensilla in green. The arista and sacculus also express IRs. (B) Innervation of antennal lobe glomeruli by OR (purple) and IR (green) neurons. The names of IR glomeruli are indicated. (C) Expression of olfactory IRs in the four subtypes of coeloconic sensilla. The strongest identified ligands for each neuron are indicated. Red and blue neurons also express the co-receptors IR8a and IR25a, respectively. The figure was adapted from Rytz et al., submitted.

Behavioural function of IRs

IRs have been associated with several kinds of behavioural responses towards their main ligands. IR64a is involved in the detection of acids, and the DC4 glomerulus, which is innervated by IR64a neurons, responds to many acids, including carbonic acid, acetic acid and HCl (Ai et al., 2010). Whereas wild-type flies strongly avoid acids, IR64a mutant flies lack this aversive response (Ai et al., 2010).

Another IR, IR84a has shown to have a function in courtship behaviour. Expressed together with IR8a in *fruitless^M* neurons (Manoli et al., 2005; Ryner et al., 1996; Stockinger et al., 2005), this IR responds to phenylacetic acid and phenylacetaldehyde, which are aromatic compounds typical of fruit odors (Abuin et al., 2011; Grosjean et al., 2011; Silbering et al., 2011). However, deletion of IR84a impairs male courtship behavior. The IR84a pathway is thus turned on by food odours instead of female-specific volatiles, suggesting that females “perfumed” with food odours are more attractive to males (Grosjean et al., 2011).

During my PhD research, I first performed a comparative genomics analysis of IR to understand their evolution as chemosensory receptors and demonstrated that these genes represent a very ancient chemosensory mechanism present in most invertebrates (Chapter 2). Next, I generated Gal4 lines for all non-olfactory IRs in order to map their expression, and thus showed that they are present in diverse gustatory neurons and are

thus likely to function as taste receptors (Chapter 3). Finally, I studied the behavioural and physiological function of IRs and IR neurons, and showed that they are involved in sensing sugars and amino acids (Chapter 4-5).

Chapter 2: The Evolution of IRs

IRs were discovered in *Drosophila*, but in order to better understand the function of this family of receptors, I first asked, whether these receptors are also present in other species. In particular, a study of the conservation of the members of this family across close or distant species is likely to provide relevant information that can be linked with their chemical ecologies. In this chapter, through the description of three research papers that I co-authored, I discuss the main evolutionary properties of the IRs: their conservation, their phylogeny, and the major mechanisms that led to their expansion.

Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction

Summary of results

This article is an extended study of IR evolution that I conducted together with Raphael Rytz, another PhD student in the Benton lab. This work is detailed in the attached publication (Croset et al., 2010). At the start of this work, I spent three months in the group of Toby Gibson at the EMBL in Heidelberg (Germany), where I developed and performed a bioinformatics screen for putative IR sequences in other species than *Drosophila*. A Hidden-Markov-Model (HMM) was built from the iGluR/IR-specific transmembrane domain, and used to screen more than 30 eukaryotic and prokaryotic genomes. IRs were identified in all protostome

species, but not beyond, suggesting that they emerged in their common ancestor, about 540 million years ago.

One IR (IR25a) is conserved across all protostomes and is the IR most similar to the ancestral one. This receptor is expressed in chemosensory organs of diverse protostome species including arthropods, nematodes or molluscs, suggesting a broadly conserved chemosensory function for that receptor, which functions as a co-receptor in *Drosophila* (Abuin et al., 2011; Benton et al., 2009). Several olfactory IRs are conserved in all insect orders with a 1:1 orthology, which contrasts with the high level of species-specificity of ORs (Jones et al., 2005; Robertson et al., 2003). By contrast, a more thorough analysis of IR evolution including twelve sequenced drosophilid species showed that most non-olfactory IRs are only conserved in flies, but not beyond. We also showed that repeated pseudogenisation events occurred in some species, including *D. sechellia*, which feeds exclusively on *Morinda citrifolia* fruit and may have a highly specialised chemosensory system. Finally, we showed that although retroposition probably allowed the first steps of their expansion, gene duplications by non-allelic homologous recombination were the main mechanism for the diversification of IR repertoires.

These results have provided insights into the evolutionary origin, expansion and diversification of IRs. The fact that IRs are conserved across all insect orders represents a major difference compared to ORs, where only the co-receptor ORCO is present in all insects (Jones et al., 2005). IRs thus represent an ancestral mechanism for sensing chemicals, and their tuning to water-soluble acids and amines (Silbering et al., 2011) is

consistent because these were likely to be major signalling molecules for the aquatic ancestor of protostomes (Derby and Sorensen, 2008).

My contribution to this work

I annotated IRs in all non-drosophilid species (Figure 1), assessed the orthology of these genes with the *D. melanogaster* repertoire, demonstrated that IRs emerged from a non-NMDA iGluR (Figure 2), studied the species-specificity of the non-antennal IRs (Figure 4), and the mechanisms of IR evolution (Figure 8B-C). In addition, I performed the expression analysis of the antennal IRs in bees (Figure 2E) and of selected non-antennal IRs in *Drosophila* (Figure 5). I also wrote the first version of the Results, Figure Legends and Methods in the paper corresponding to these experiments and provided input and feedback on other parts of the manuscript.

Ancient Protostome Origin of Chemosensory Ionotropic Glutamate Receptors and the Evolution of Insect Taste and Olfaction

Vincent Croset¹✉, Raphael Rytz¹✉, Scott F. Cummins²✉, Aidan Budd³, David Brawand¹, Henrik Kaessmann¹, Toby J. Gibson³, Richard Benton¹*

1 Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, **2** School of Biological Sciences, The University of Queensland, St. Lucia, Queensland, Australia, **3** Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

Abstract

Ionotropic glutamate receptors (iGluRs) are a highly conserved family of ligand-gated ion channels present in animals, plants, and bacteria, which are best characterized for their roles in synaptic communication in vertebrate nervous systems. A variant subfamily of iGluRs, the Ionotropic Receptors (IRs), was recently identified as a new class of olfactory receptors in the fruit fly, *Drosophila melanogaster*, hinting at a broader function of this ion channel family in detection of environmental, as well as intercellular, chemical signals. Here, we investigate the origin and evolution of IRs by comprehensive evolutionary genomics and *in situ* expression analysis. In marked contrast to the insect-specific Odorant Receptor family, we show that IRs are expressed in olfactory organs across Protostomia—a major branch of the animal kingdom that encompasses arthropods, nematodes, and molluscs—indicating that they represent an ancestral protostome chemosensory receptor family. Two subfamilies of IRs are distinguished: conserved “antennal IRs,” which likely define the first olfactory receptor family of insects, and species-specific “divergent IRs,” which are expressed in peripheral and internal gustatory neurons, implicating this family in taste and food assessment. Comparative analysis of drosophilid IRs reveals the selective forces that have shaped the repertoires in flies with distinct chemosensory preferences. Examination of IR gene structure and genomic distribution suggests both non-allelic homologous recombination and retroposition contributed to the expansion of this multigene family. Together, these findings lay a foundation for functional analysis of these receptors in both neurobiological and evolutionary studies. Furthermore, this work identifies novel targets for manipulating chemosensory-driven behaviours of agricultural pests and disease vectors.

Citation: Croset V, Rytz R, Cummins SF, Budd A, Brawand D, et al. (2010) Ancient Protostome Origin of Chemosensory Ionotropic Glutamate Receptors and the Evolution of Insect Taste and Olfaction. *PLoS Genet* 6(8): e1001064. doi:10.1371/journal.pgen.1001064

Editor: David L. Stern, Princeton University, Howard Hughes Medical Institute, United States of America

Received: February 19, 2010; **Accepted:** July 12, 2010; **Published:** August 19, 2010

Copyright: © 2010 Croset et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: SFC was supported by a University of Queensland Fellowship and the Australian Research Council, RR and DB by Roche Research Foundation Fellowships, and VC by an EMBO Short-term Fellowship and a Boehringer Ingelheim Foundation Fellowship. Research in RB's laboratory is funded by the University of Lausanne, a European Research Council Starting Independent Researcher Grant, and the Swiss National Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Richard.Benton@unil.ch

✉ These authors contributed equally to this work.

✉ Current address: Faculty of Science, Health and Education, University of the Sunshine Coast, Maroochydore, Queensland, Australia

Introduction

Ionotropic glutamate receptors (iGluRs) are a conserved family of ligand-gated ion channels present in both eukaryotes and prokaryotes. By regulating cation flow across the plasma membrane in response to binding of extracellular glutamate and related ligands, iGluRs represent an important signalling mechanism by which cells modify their internal physiology in response to external chemical signals.

iGluRs have originated by combination of protein domains originally encoded by distinct genes (Figure 1A) [1–2]. An extracellular amino-terminal domain (ATD) is involved in assembly of iGluR subunits into heteromeric complexes [3]. This precedes the ligand-binding domain (LBD), whose two half-domains (S1 and S2) form a “Venus flytrap” structure that closes around glutamate and related agonists [4]. Separating S1 and S2

in the primary structure is the ion channel pore, formed by two transmembrane segments and a re-entrant pore loop [5]. S2 is followed by a third transmembrane domain of unknown function and a cytosolic carboxy-terminal tail.

Animal iGluRs have been best characterised for their essential roles in synaptic transmission as receptors for the excitatory neurotransmitter glutamate [1,6]. Three pharmacologically and molecularly distinct subfamilies exist, named after their main agonist: α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), kainate and N-methyl-D-aspartate (NMDA). AMPA receptors mediate the vast majority of fast excitatory synaptic transmission in the vertebrate brain, while Kainate receptors have a subtler modulatory role in this process. NMDA receptors require two agonists for activation, glutamate and glycine, and function in synaptic and neuronal plasticity. Representatives of these iGluR subfamilies have been identified across vertebrates [7], as well as

Author Summary

Ionotropic glutamate receptors (iGluRs) are a family of cell surface proteins best known for their role in allowing neurons to communicate with each other in the brain. We recently discovered a variant class of iGluRs in the fruit fly (*Drosophila melanogaster*), named Ionotropic Receptors (IRs), which function as olfactory receptors in its “nose,” prompting us to ask whether iGluR/IRs might have a more general function in detection of environmental chemicals. Here, we have identified families of IRs in olfactory and taste sensory organs throughout protostomes, one of the principal branches of animal life that includes snails, worms, crustaceans, and insects. Our findings suggest that this receptor family has an evolutionary ancient function in detecting odors and tastants in the external world. By comparing the repertoires of these chemosensory IRs among both closely- and distantly-related species, we have observed dynamic patterns of expansion and divergence of these receptor families in organisms occupying very different ecological niches. Notably, many of the receptors we have identified are in insects that are of significant harm to human health, such as the malaria mosquito. These proteins represent attractive targets for novel types of insect repellents to control the host-seeking behaviors of such pest species.

invertebrates, such as the fruit fly *Drosophila melanogaster*, the nematode worm *Caenorhabditis elegans* and the sea slug *Aplysia californica* [8–10].

While most iGluRs have exquisitely tuned synaptic functions, identification of iGluR-related genes in prokaryotic and plant genomes provided initial indication of more diverse roles for this class of ion channel. A bacterial glutamate receptor, GluR0, was first characterised in the cyanobacterium, *Synechocystis PCC6803* [11]. GluR0 conducts ions in response to binding of glutamate and other amino acids *in vitro*, suggesting a potential function in extracellular amino acid sensing *in vivo*. The flowering plant *Arabidopsis thaliana* has 20 iGluR-related genes, named GLRs [12–13]. Genetic analysis of one receptor, GLR3.3, has implicated it in mediating external amino acid-stimulated calcium increases in roots [14].

We recently described a family of iGluR-related proteins in *D. melanogaster*, named the Ionotropic Receptors (IRs) [15]. Several lines of evidence demonstrated that the IRs define a new family of olfactory receptors. First, the IR LBDs are highly divergent and lack one or more residues that directly contact the glutamate ligand in iGluRs. Second, several IRs are expressed in sensory neurons in the principal *D. melanogaster* olfactory organ, the antenna, that do not express members of the other *D. melanogaster* chemosensory receptor families, the Odorant Receptors (ORs) and Gustatory Receptors (GRs) [16]. Third, IR proteins localise to the ciliated endings of these sensory neurons and not to synapses [15]. Finally, mis-expression of an IR in an ectopic neuron is sufficient to confer novel odour-evoked neuronal responses, providing direct genetic evidence for a role in odour sensing [15].

The identification of the IRs as a novel family of olfactory receptors in *D. melanogaster* provides a potential link between the well-characterised signalling activity of iGluRs in glutamate neurotransmitter-evoked neuronal depolarisation and a potentially more ancient function of this family in environmental chemosensation. In this work, we have combined comparative genomics, molecular evolutionary analysis and expression studies to examine the evolution of the IRs. Four principal issues are addressed: first,

when did olfactory IRs first appear? Are they a recent acquisition as environmental chemosensors in *D. melanogaster*, or do they have earlier origins in insect or deeper animal lineages? Second, what is the most recent common ancestor of IR genes? Do they derive from AMPA, Kainate or NMDA receptors, or do they represent a distinct subfamily that evolved from the ancestral animal iGluR? Third, what mechanisms underlie the expansion and diversification of this multigene family? Finally, do IRs function only as olfactory receptors or are they also involved in other sensory modalities? Through answers to these questions, we sought insights into IR evolution in the context of the origins of iGluRs, the appearance and evolution of other chemosensory receptor repertoires and the changing selective pressures during animal diversification and exploitation of new ecological niches.

Results

A broad phylogenetic survey of iGluR and IR genes

iGluRs and IRs are characterised by the presence of a conserved ligand-gated ion channel domain (the combined Pfam domains PF10613 and PF00060 [17]) (Figure 1A). All iGluRs additionally contain an ATD (Pfam domain PF01094), which is discernible, but more divergent, in only two *D. melanogaster* IRs, IR8a and IR25a. Most IRs have only relatively short N-terminal regions preceding the LBD S1 domain (Figure 1A). To identify novel iGluR/IR-related genes, we therefore constructed a Hidden Markov Model (HMM) from an alignment of the conserved iGluR/IR C-terminal region, which is specific to this protein family. In combination with exhaustive BLAST searches, we used this HMM to screen raw genomic sequences and available annotated protein databases of 32 diverse eukaryotic species and 971 prokaryotic genomes (see Materials and Methods and Table S2 in Supporting Information). These screens identified all previously described eukaryotic iGluRs and all *D. melanogaster* IRs, as well as 23 prokaryotic iGluRs. Novel sequences were manually reannotated and classified by sequence similarity, phylogenetic analysis and domain structure as either non-NMDA (i.e. AMPA and Kainate) or NMDA subfamily iGluRs, or IRs (Figure 1B, Table S3, and Datasets S1 and S2). Like *D. melanogaster* IRs, newly annotated IRs have divergent LBDs that lack some or all known glutamate-interacting residues, supporting their distinct classification from iGluRs.

iGluRs are widespread in eukaryotes, present in all analysed Metazoa (except the sponge, *Amphimedon queenslandica* [18]) and Plantae, but absent in unicellular eukaryotes (Figure 1B, Table S3, and Datasets S1 and S2). Analysis of iGluR subfamilies on the eukaryotic phylogeny suggests that NMDA receptors may have appeared after non-NMDA receptors, as we identified them in Eumetazoa but not in the placozoan *Trichoplax adhaerens*. Further support for this conclusion will require additional genome sequences. One member of the Eumetazoa, the sea urchin *Strongylocentrotus purpuratus*, may have secondarily lost NMDA receptors. Different species contain distinct numbers of each iGluR subfamily: vertebrates, for example, have more NMDA receptor subunits than invertebrates.

Notably, IRs were identified throughout Protostomia, encompassing both Ecdysozoa (e.g. nematodes and arthropods) and Lophotrochozoa (e.g. molluscs and annelids) (Figure 1B, Table S3, and Datasets S1 and S2). There is substantial variation in the size of the IR repertoire, from three in *C. elegans* to eighty-five in the crustacean *Daphnia pulex*. Amongst insects, Diptera (i.e. flies and mosquitoes) generally had a larger number of IRs than other species. We did not identify IRs in Deuterostomia, Cnidaria or Placozoa.

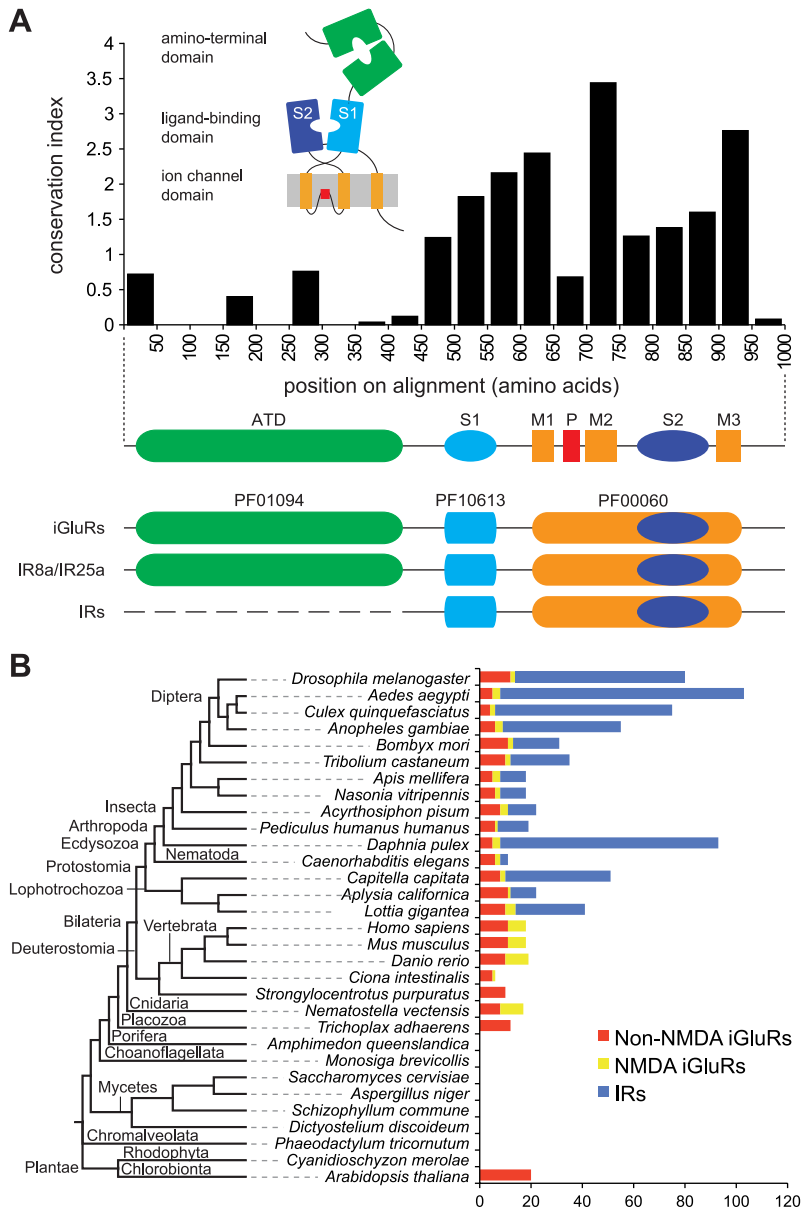


Figure 1. A broad phylogenetic survey of iGluR and IR genes. (A) Top: Histogram showing the mean conservation index (number of conserved physico-chemical properties) [74,91] for 50 amino acid column-blocks of aligned *D. melanogaster* iGluRs and IRs, illustrating the higher conservation of the C-terminal region. The protein domain organisation of iGluRs/IRs is shown in cartoon form above the histogram and in linear form below it. Bottom: illustration of the three Pfam domains present in iGluRs and IRs. IR8a and IR25a contain the Pfam domain corresponding to the iGluR ATD. IR21a, IR40a, IR64a and IR93a also contain long N-termini (~400 amino acids) but these have extremely low primary structural similarity to the ATD. All other IRs have much shorter N-terminal regions (~200 amino acids) that lack homology to the ATD or other protein domains. (B) Histogram of the number of non-NMDA (red), NMDA (yellow) and IR (blue) sequences identified in the indicated eukaryotic species. An unscaled tree showing the phylogenetic relationships between these species is illustrated on the left. doi:10.1371/journal.pgen.1001064.g001

Evolutionary conservation and expression of antennal IRs

To explore the evolutionary origin of the IRs, we examined phylogenetic relationships of the identified protostome IRs. Reciprocal best-hit analysis using *D. melanogaster* sequences as queries revealed that a subset of this species' IRs was conserved in several distant lineages, allowing us to define putative orthologous groups. These include one group containing representatives of all protostome species (IR25a), one represented by all arthropods (IR93a), nine by most or all insects, and three by dipteran insects (Figure 2A and 2B). For most orthologous groups, a single gene for

each species was identified. In a few cases, for example the IR75 group, certain species were represented by several closely related in-paralogues, some of which appeared to be pseudogenes (Figure 2A and 2B, Table S3, and Datasets S1 and S2).

Consistent with its conservation in Protostomia, IR25a is the IR with the most similar primary sequence to iGluRs, suggesting that it is the IR gene most similar to the ancestral IR. Analysis of the phylogenetic relationship of IR25a and eukaryotic iGluRs locates it clearly together with the animal iGluR family, in the non-NMDA receptor clade (Figure 2C). To substantiate this conclu-

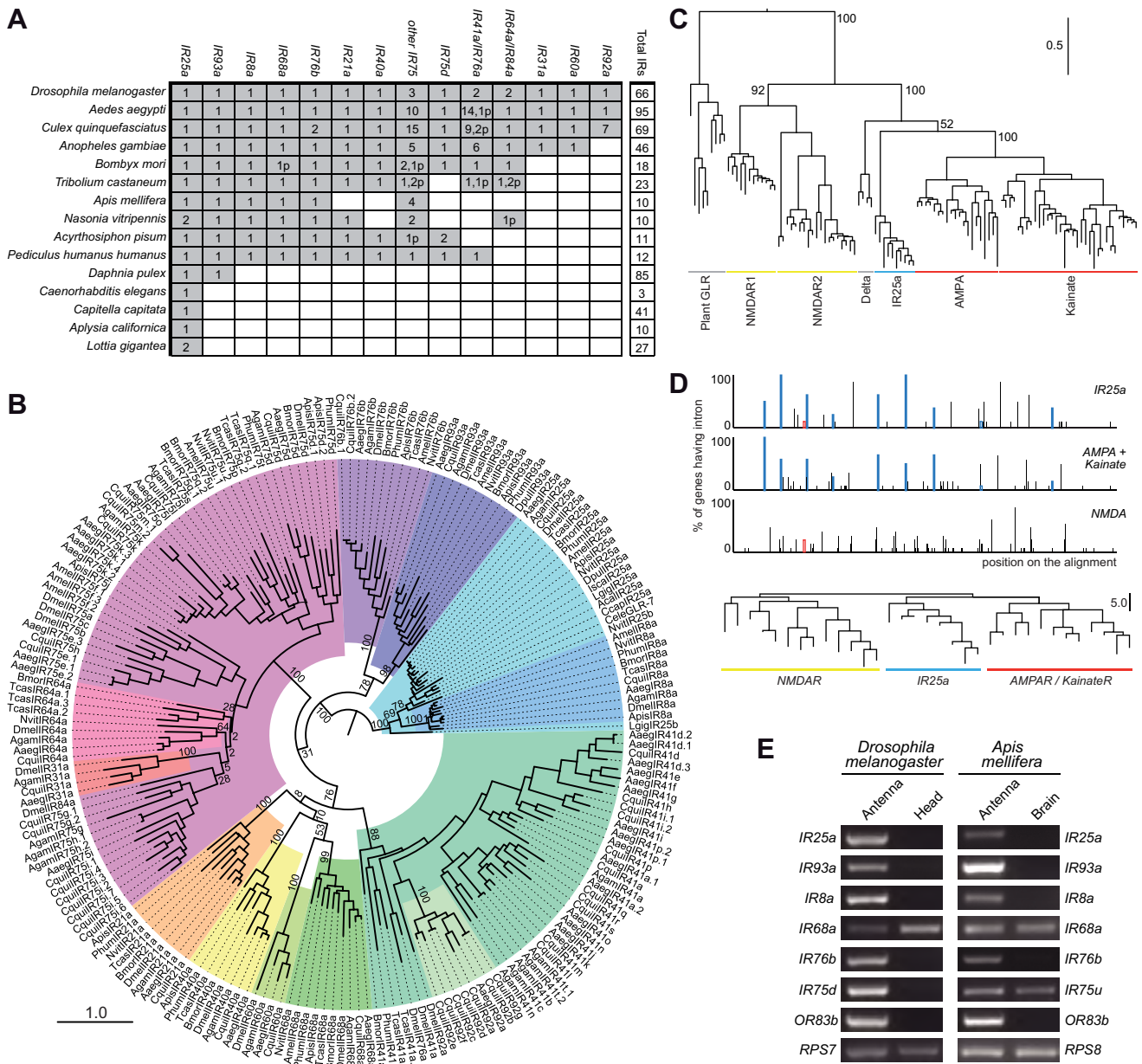


Figure 2. Evolutionary origins and conservation of antennal IRs. (A) Table of antennal IR orthologous groups in the indicated protostome species. A shaded square signifies the presence of at least one representative gene in a species. Figures within a square indicate the existence of multiple functional copies; the "p" suffix indicates the number of pseudogene copies. We considered as pseudogenes only those with frameshift or internal stop codons inside conserved domains due to the difficulty in accurately annotating the termini of these sequences. (B) Phylogenetic relationships of the genes shown in (A). Each colour represents a group of orthologous sequences. The sequences were aligned with PROBCONS and the tree was built with RAXML under the WAG model of substitution with 1000 bootstrap replicates. The scale bar represents the expected number of substitutions per site. (C) Phylogenetic relationships between iGluRs and all IR25a orthologues, excluding low quality or short gene annotations. The tree was built with RAXML under the WAG model of substitution, with 1000 bootstrap replicates. Bootstrap values for selected branches are shown as percentages. The scale bar represents the expected number of substitutions per site. (D) *Top*: Map of intron positions in an alignment of eight IR25a orthologues, 12 AMPA/Kainate receptors and 13 NMDA receptors (see Dataset S3 for alignment file). Coloured boxes illustrate introns whose positions and phases are conserved in at least one member of two different subgroups. Empty coloured boxes indicate introns conserved in position, but not in phase. *Bottom*: Phylogram based on the position of introns in the same subset of sequences as above. The scale bar represents the number of non-conserved intron positions. (E) RT-PCR analysis of antennal IR gene expression of orthologous genes (except *DmellIR75a* and *AmellIR75u*, which are paralogous genes) in *D. melanogaster* and *A. mellifera* tissues. Control RT-PCR products for comparative analysis of gene expression correspond to the ribosomal genes *RPS7* (*D. melanogaster*) and *RPS8* (*A. mellifera*). All RT-PCR products were sequenced to confirm their identity. doi:10.1371/journal.pgen.1001064.g002

sion, we asked whether the IR25a gene structure resembles more closely that of NMDA or non-NMDA receptors. Intron positions and numbers are extremely variable across IR25a orthologues, with multiple cases of intron loss, gain and putative intron sliding

events by a few nucleotides (Figure 2D). Nevertheless, we identified eight intron positions that are conserved between at least subsets of IR25a orthologues and D. melanogaster non-NMDA receptor genes, some of which may represent intron positions

present in a common ancestral gene. By contrast, only a single intron that was conserved in position (but not in phase) was identified between *DmelIR25a* (but not other *IR25a* orthologues) and *DmelNMDAR1* (Figure 2D). A phylogram of intron positions in *IR25a*, non-NMDA and NMDA sequences reveals greater similarity of *IR25a* intron positions to those of non-NMDA receptors than NMDA receptors (Figure 2D). Together, these observations support a model in which *IR25a* evolved from a bilaterian non-NMDA receptor gene.

The conserved *D. melanogaster* IRs encompass the entire subset of its IR repertoire that is expressed in the antenna [15]. Moreover, evidence for antennal expression of the three additional genes, *DmelIR41a*, *DmelIR60a* and *DmelIR68a*, has been obtained by reverse transcription (RT)-PCR analysis, although we have not yet been able to corroborate this by RNA *in situ* hybridisation (data not shown). These combined phylogenetic and expression properties led us to designate this subfamily of receptors the “antennal IRs”.

We examined whether antennal expression of this subfamily of IRs is conserved outside *D. melanogaster* by performing a series of RT-PCR experiments on the honey bee, *Apis mellifera*, for all six putative antennal IR orthologues: *IR8a*, *IR25a*, *IR68a*, *IR75u*, *IR76b* and *IR93a* (see Materials and Methods for the nomenclature of newly-identified IRs). As in *D. melanogaster*, we could reproducibly amplify all of these bee genes from antennal RNA preparations but not in control brain RNA, except for *AmelIR68a* and *AmelIR75u*, which are also detected in the brain (Figure 2E). Thus, antennal expression of this subgroup of IRs is conserved across the 350 million years separating dipteran and hymenopteran insect orders [19], and therefore potentially in all insects.

Conserved IR chemosensory expression in Protostomia

To investigate whether IRs are likely to have an olfactory function beyond insects, we examined expression of the IR repertoire from a representative of a distantly related protostome lineage, *Aplysia* molluscs, whose last common ancestor with *D. melanogaster* probably existed 550–850 million years ago [20]. We first used RT-PCR to analyse the expression of the ten *Aplysia* IR genes in a variety of sensory, nervous and reproductive tissues (Figure 3A). Notably, the *Aplysia IR25a* orthologue is predominantly expressed in the olfactory organs, the rhinophore and oral tentacle [21]. Two other *Aplysia*-specific IR genes, *IR214* and *IR217*, are expressed in the rhinophore and oral tentacle, respectively, and not detected in other tissues, except for the large hermaphroditic duct (*IR214*) and skin (*IR217*). Five additional IRs are also expressed in the oral tentacle, but displayed broader tissue expression in skin and the central nervous system; both of these tissues are likely to contain other types of chemosensory cells [22–23]. Expression of two IR genes, *IR209* and *IR213*, was not detected in this analysis (data not shown).

To further characterise *Aplysia IR25a*, we analysed its spatial expression in the mature *A. dactylorella* rhinophore by RNA *in situ* hybridisation. An antisense probe for *AdacIR25a* labels a small number of cells in rhinophore cryosections. Their size and morphology is typical of neurons, although we lack an unambiguous neuronal marker to confirm this identification (Figure 3B–3D). These cells are found either singly or in small clusters adjacent or close to the sensory epithelial surface in the rhinophore groove, in a similar position to cells expressing other types of chemosensory receptors [21]. A control sense riboprobe showed no specific staining (Figure 3E). Together, these results are consistent with at least some of these molluscan IRs having a chemosensory function.

The expression of putative *IR25a* orthologues has previously been reported in two other Protostomia. An *IR25a*-related gene from the American lobster, *Homarus americanus*, named OET-07, is specifically expressed in mature olfactory sensory neurons [24–25]. In *C. elegans*, a promoter reporter of the *IR25a* orthologue, *GLR-7*, revealed expression in a number of pharyngeal neurons [9], which might have a role in food sensing [26]. While both crustacean and nematode genes were classified in these studies as iGluRs, there is no evidence that they act as canonical glutamate receptors, and we suggest that they fulfil instead a chemosensory function.

Species-specificity of divergent IRs

The antennal IR subfamily accounts for only a small fraction of the IR repertoire in most analysed insects and only 1–2 genes in other Protostomia. The remaining majority of IR sequences are - amongst the genomes currently available - largely species-specific, with low amino acid sequence identity (as little as 8.5%) with other IR genes in either the same or different species. We refer to this group of genes here as the “divergent IRs”. Dipteran insects have particularly large expansions of divergent IRs (Figure 1B). Phylogenetic analysis revealed no obvious orthologous relationships of these genes either between *D. melanogaster* and mosquitoes or amongst the three mosquito species (*Aedes aegypti*, *Culex quinquefasciatus* and *Anopheles gambiae*) (Figure 4). Instead, this subfamily of IRs displays a number of species-specific clades, perhaps reflective of the distinct ecological niches of these insects.

Divergent IRs as candidate gustatory receptors in adult and larval *D. melanogaster*

By contrast to antennal IRs, divergent IR expression has not been detected in *D. melanogaster* olfactory organs [15], leading us to test whether these genes are expressed in other types of chemosensory tissue. As endogenous transcripts of non-olfactory chemosensory genes, such as GRs, are difficult to detect [27–28], we employed a sensitive transgenic approach to investigate divergent IR expression. We transformed flies with constructs containing putative promoter regions for these genes upstream of the yeast transcription factor *GAL4* and used these “driver” transgenes to induce expression of a *GAL4*-responsive *UAS-mCD8:GFP* fluorescent reporter [29]. We sampled divergent IRs from several distinct clades, including *IR7a*, *IR11a*, *IR52b*, *IR56a* and *IR100a* (Figure 4). All IR promoter-*GAL4* constructs were inserted in the same genomic location using the ϕ C31 integrase system [30], eliminating transgene-specific position effects on expression resulting from their site of integration.

Expression of three of these divergent IR reporters was observed in highly selective populations of neurons in distinct gustatory organs (Figure 5A). In the adult, *IR7a* is expressed in at least eleven neurons in the labellum, a sense organ involved in peripheral taste detection (Figure 5B) [31]. Two reporters labelled neurons in internal sense organs in the pharynx: *IR11a* is expressed in one neuron in the ventral cibarial sense organ and *IR100a* is expressed in two neurons in the dorsal cibarial sense organ (Figure 5C and 5D). These internal pharyngeal neurons are thought to play a role in assessment of ingested food prior to entry into the main digestive system [16]. Expression was not detected in any other neurons or other cell types in the adult head (data not shown), although we cannot exclude expression in other regions of the body. *IR52b* and *IR56a* reporters were not detected in these experiments.

We also examined expression of these reporters at an earlier stage in the *D. melanogaster* life cycle, third instar larvae, which display robust gustatory responses [16]. The same three IR reporters were exclusively detected in unique bilaterally-symmetric

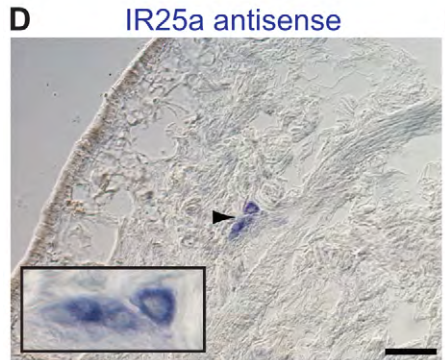
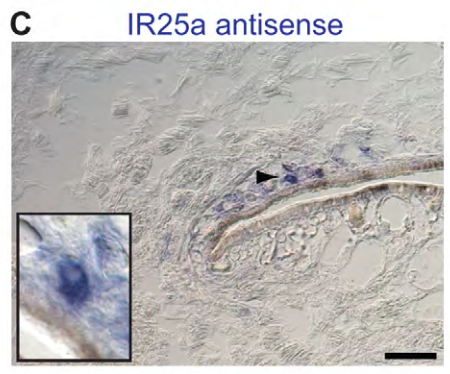
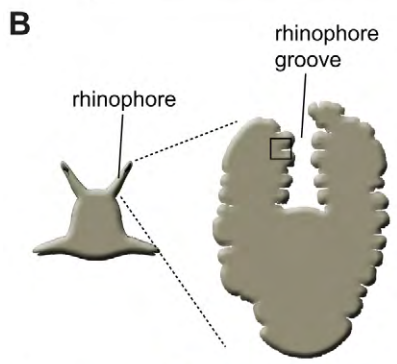
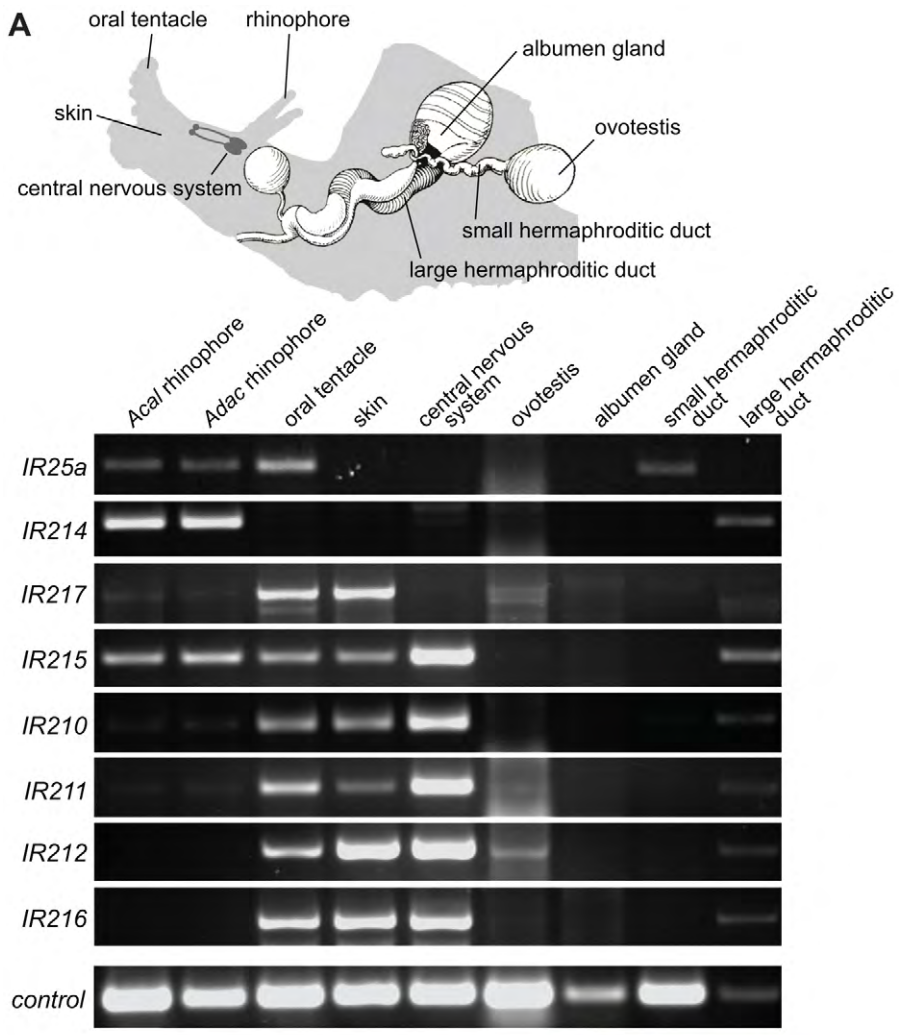


Figure 3. Olfactory expression of IRs in *Aplysia molluscs*. (A) *Top*: Schematic representation of *Aplysia*, illustrating the location of selected sensory, neuronal and reproductive tissues used for RNA isolation and RT-PCR (adapted from [21]). The central nervous system samples comprised pooled cerebral, pleural, buccal, pedal and abdominal ganglia. The skin samples were taken from the side of the head. *Bottom*: RT-PCR analysis of *Aplysia* IR gene expression from the indicated species and tissues. Only rhinophores from *A. californica* (*Acal*) were tested due to limited availability of animals, while rhinophore and other tissues were examined for the closely related species *A. dactylorella* (*Adac*) [92]. Nucleotide sequence identity of IR orthologues between these species is >85%. Control RT-PCR corresponds to β -actin. (B) Schematic of *Aplysia* rhinophore showing the approximate location of the field of views of the rhinophore groove olfactory tissue in (C–E). (C,D) RNA *in situ* hybridisation on *A. dactylorella* rhinophore sections using a digoxigenin-labelled antisense RNA probe for *AdaclR25a*. Micrographs reveal *IR25a* expression (blue) in small clusters of cells of a characteristic neuronal morphology close to the sensory epithelial surface. Higher magnifications of specific cellular staining (arrowhead) are shown in the insets. The scale bars represent 100 μ m. (E) Control RNA *in situ* hybridisation on an *A. dactylorella* rhinophore section with a digoxigenin-labelled sense riboprobe for *AdaclR25a*. No signal is apparent. The scale bar represents 100 μ m.
doi:10.1371/journal.pgen.1001064.g003

larval gustatory organs: *IR7a* was expressed in two neurons in the terminal organ at the periphery, *IR11a* in a single neuron in the ventral pharyngeal sense organ and *IR100a* in two neurons in the posterior pharyngeal sense organ (Figure 5E–5H). Notably, all of these neurons in both adult and larval tissues (except for a single *IR7a*-expressing cell in the terminal organ) co-express *IR25a*, as revealed by a specific antibody against this receptor (Figure 5) [15]. *IR25a* is also expressed in several other cells in each of the gustatory organs, which may express other divergent IRs not examined here. Together these results support a role for divergent IRs as taste receptors in distinct taste organs and stages of the *D. melanogaster* life cycle.

IR evolution on the *Drosophila* phylogeny

To obtain more detailed insights into the processes underlying the expansion and diversification of IR repertoires, we investigated their evolution over a shorter timescale by comparative analysis of *D. melanogaster* with 11 additional sequenced drosophilid species [32–33]. The last common ancestor of these drosophilids is estimated to have existed 40 million years ago [34], by contrast to the ~250 million years since the last common ancestor of *D. melanogaster* and the mosquito *A. gambiae* [35]. Certain species may have diverged much more recently, such as *D. simulans* and *D. sechellia*, whose last common ancestor may have existed only 250,000 years ago [36].

We used *D. melanogaster* sequences as queries in exhaustive BLAST searches of the drosophilid genomes. Retrieved sequences were manually reannotated to unify gene structure predictions across species and, in some cases, genes were partially resequenced to close sequence gaps or verify them as pseudogenes (see Materials and Methods, Table S3, and Datasets S1 and S2). Although predicted full-length gene sequences could be annotated for most genes, 28 sequences remain incomplete - but assumed in further analysis to be functional - because of a lack of sequence data or difficulty in precise annotation of exons in divergent regions of these genes. Of the 926 drosophilid sequences identified (including those of *D. melanogaster*), 49 genes were classified as pseudogenes because they consisted of only short gene fragments or contained frameshift mutations and/or premature stop codons. We clustered all genes into orthologous groups by examining their sequence similarity, phylogenetic relationships and, in the case of *IR47a*, *IR47b*, *IR47c*, *IR56e* and *IR60f*, their micro-syntentic relationships (Table S1 and Figure 6). For drosophilid species that are most distant from *D. melanogaster*, definition of precise orthologous relationships was not always possible, particularly for groups of closely related IR genes (e.g. *IR52a–f*, *IR60b–f*) (Table S1). Orthologous groups were named after their *D. melanogaster* representatives or a logical variant in groups where no *D. melanogaster* gene was identified (see Materials and Methods).

This analysis identified 14 iGluR and 58–69 IR genes in each of the twelve drosophilid species (Figure 6A and Table S1). iGluRs are highly conserved, with a mean amino acid sequence identity of

$89 \pm 1\%$ s.e.m., and a single representative for each species in every orthologous group. Antennal IRs are also well conserved (mean sequence identity = $76 \pm 2\%$) and amongst these genes we identified only a single pseudogenisation event, in *D. sechellia* *IR75a*, and a single gene duplication event, of *D. mojavensis* *IR75d*. By contrast, divergent IRs, though also largely classifiable into monophyletic groups, display a more dynamic pattern of evolution (mean sequence identity = $61 \pm 2\%$), with multiple cases of gene loss, pseudogenisation or duplication (Figure 6 and Table S1).

Species-specific rates of IR gene loss and gain

We reconciled the gene phylogeny with the drosophilid species phylogeny to estimate the number of IR gene gain and loss events. While this analysis is necessarily constrained by our ability to accurately define gene orthology, we estimated across the entire phylogeny there to be sixteen gene gain events (gene birth rate, $B = 0.0006/\text{gene}/\text{million years}$) and 76 gene loss events (gene death rate, $D = 0.0030/\text{gene}/\text{million years}$) (Figure 7A, see Materials and Methods). Most (46/76) gene losses are pseudogenisation events, which indicates that many of these events must have occurred relatively recently, as drosophilid species appear to eliminate pseudogenes rapidly from their genomes [37–38]. Notably, 13 gene loss events – 12 of which reflect the presence of just one or a small number of premature stop codons or frameshift mutations – occur on the branch leading to the specialist *D. sechellia*. Consequently, the gene loss rate on this branch is remarkably high compared with its generalist sister species *D. simulans* (Figure 7A and 7B).

Selective forces acting on drosophilid IR genes

We studied the selective forces acting on drosophilid iGluRs and IRs by calculating the ratio of nonsynonymous to synonymous nucleotide substitution rates (d_N/d_S , ω_1) in these genes from all 12 species. All tested iGluR, antennal IR and divergent IR genes are evolving under strong purifying selection ($\omega_1 < 1$) (Figure 7C, left and Table S4), suggesting that they all encode functional receptors. iGluRs have the lowest estimated d_N/d_S ratio (median $\omega_1 = 0.060$), consistent with a conserved role in synaptic communication. Antennal IRs have an intermediate d_N/d_S ratio (median $\omega_1 = 0.107$) and divergent IRs the highest (median $\omega_1 = 0.149$), suggesting that divergent IRs have evolved under weaker purifying selection and/or contain more sites that have been shaped by positive selection. Amongst the IRs, *IR25a* has the lowest d_N/d_S ratio ($\omega_1 = 0.028$), consistent with its high sequence conservation in and beyond drosophilids (Figure 2).

To compare these properties with those of other insect chemosensory receptor families (ORs and GRs) [39], we also calculated d_N/d_S ratios for IR genes from only the five sequenced species of the *melanogaster* subgroup (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. erecta* and *D. yakuba*). For this subset of sequences, the relative differences between median d_N/d_S ratios (ω_2) for the iGluR and IR gene subfamilies observed with all twelve species was

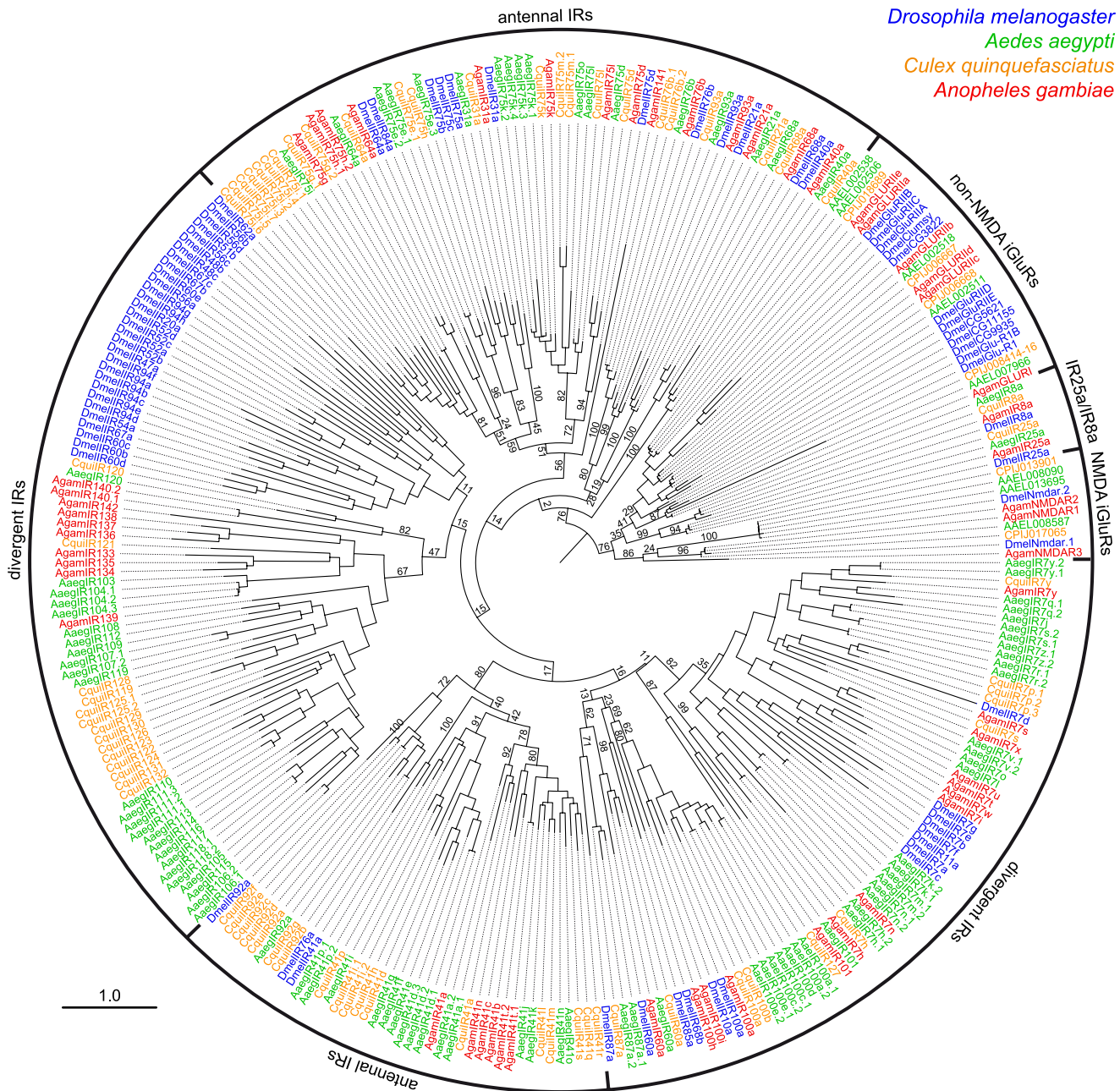


Figure 4. Species-specificity of divergent IR repertoires. Phylogenetic tree of all iGLuRs and IRs from *D. melanogaster* (blue), *A. aegypti* (green), *C. quinquefasciatus* (orange) and *A. gambiae* (red). Sequences were aligned with PROBCONS and the tree was built with RAxML under the WAG model of substitution, with 500 bootstrap replicates. The scale bar represents the expected number of substitutions per site. Note that due to the high divergence and number of sequences analysed, bootstrap values in several of the most internal nodes are extremely low and the position of certain large clades of IR genes on the tree are distinct from trees in other figures.
doi:10.1371/journal.pgen.1001064.g004

reproduced (Figure 7C, right). The GR gene family has previously been noted to evolve under weaker purifying selection than ORs [39]. Notably, we found that the median d_N/d_S ratios for antennal IRs ($\omega_2 = 0.120$) is statistically indistinguishable from that of ORs ($\omega_2 = 0.137$) ($p > 0.4$, Wilcoxon rank-sum test), and that the median d_N/d_S ratio of divergent IRs ($\omega_2 = 0.176$) is statistically indistinguishable from that of GRs ($\omega_2 = 0.217$) ($p > 0.5$, Wilcoxon rank-sum test). Thus, the selective forces acting on the IR receptor gene subfamilies parallel those on the ORs and GRs and appear to correlate with their putative distinct chemosensory functions in

olfaction and gustation (Figure 7C, right). The reason for this difference is unknown, but might reflect reduced evolutionary constraints on co-expressed and partially redundant taste receptor genes or selection for higher diversity in taste receptor sequences to recognise more variable non-volatile chemosensory ligands in the environment.

Most residues of IR proteins can be expected to have evolved under purifying selection to maintain conserved structural and signalling properties, which may mask detection of positive selection ($\omega > 1$) at a small number of sites that contribute to their

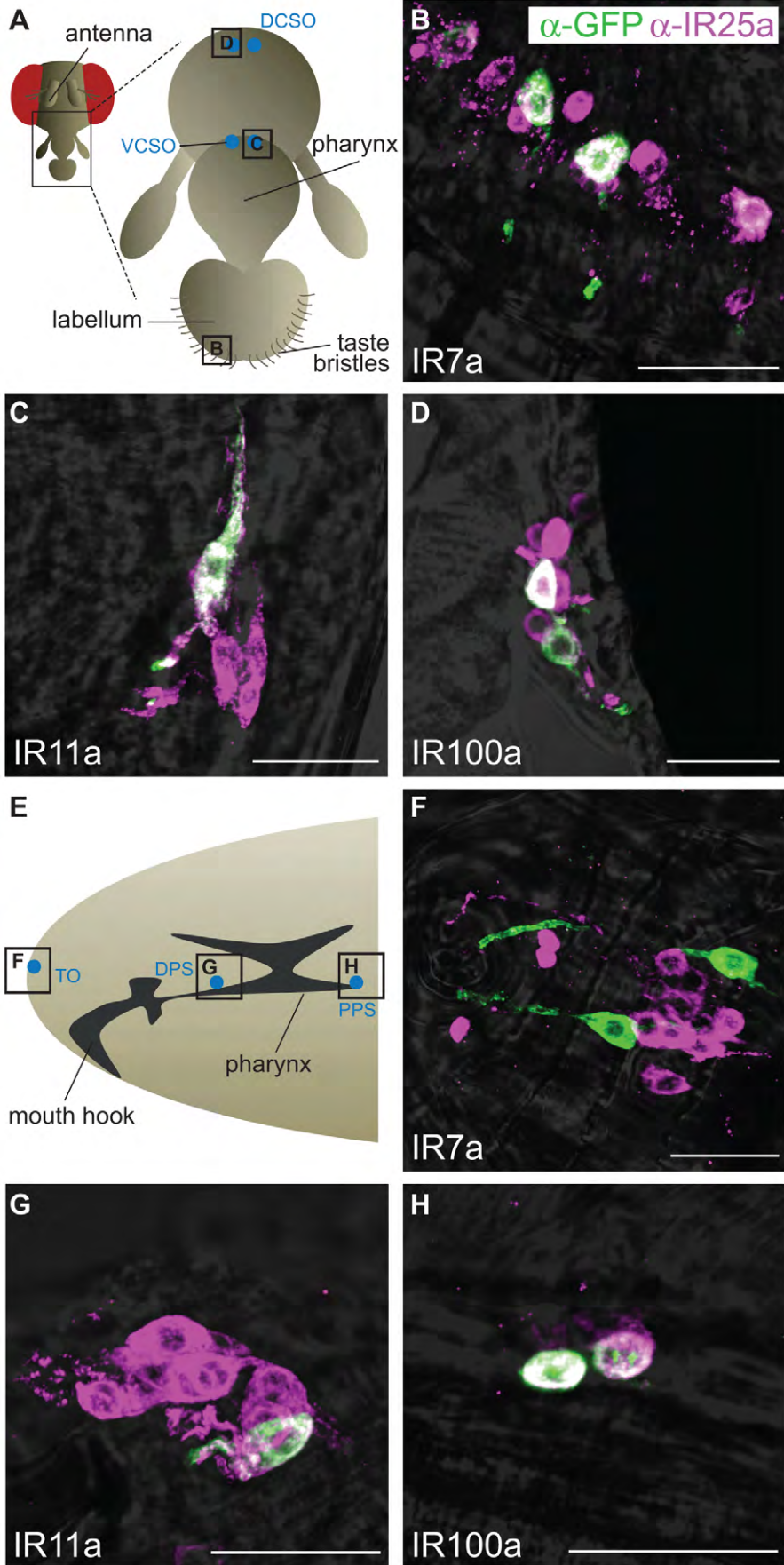


Figure 5. Expression of divergent IRs in *D. melanogaster* adult and larval gustatory organs. Immunofluorescence with anti-GFP (green) and anti-IR25a (magenta) antibodies (overlaid on bright-field images) on whole-mount tissues from animals expressing a membrane targeted GFP reporter transgene (*UAS-mCD8:GFP*) under the control of the indicated *IR promoter-GAL4* driver transgenes. The scale bars represent 20 μm . (A) Schematic of the adult *D. melanogaster* proboscis showing the location of the field of views in (B–D). DCSO: dorsal cibarial sense organ, VCSO: ventral cibarial sense organ. (B) *IR7a-GAL4* drives expression of mCD8:GFP in the labellum. (C) *IR11a-GAL4* drives expression of mCD8:GFP in the VCSO. (D) *IR100a-GAL4* drives expression of mCD8:GFP in the DCSO. (E) Schematic of the *D. melanogaster* larval head showing the location of the field of views in (F–H). TO: terminal organ, DPS: dorsal pharyngeal sense organ, PPS: posterior pharyngeal sense organ. (F) *IR7a-GAL4* drives expression of mCD8:GFP in the TO. (G) *IR11a-GAL4* drives expression of mCD8:GFP in the DPS. (H) *IR100a-GAL4* drives expression of mCD8:GFP in the PPS.
doi:10.1371/journal.pgen.1001064.g005

functional diversity. To obtain evidence for site-specific selection we applied site class models M7 and M8 in PAML to analyse 49 sets of orthologous IR genes of the six species of the *melanogaster* group. This test did not identify any sites significantly under positive selection after Bonferroni correction (Table S4), a result consistent with orthologous IR genes having the same function across drosophilids.

Site-specific positive selection may be more easily detectable in relatively recent IR gene duplicates potentially undergoing functional divergence. We therefore analysed the sole duplication of an antennal IR, *IR75d.1* and *IR75d.2* in *D. mojavensis*. Assuming an estimated divergence time of 35 My between *D. virilis* and *D. mojavensis* [40], and based on analysis of d_s of *IR75d* genes in these species (see Materials and Methods), we estimated this duplication to have occurred relatively recently, approximately 2.6–5.1 My ago. Using a branch-site test we identified two sites ($p < 0.05$) that have evolved under positive selective pressure, where DmojIR75d.1 and DmojIR75d.2 appear to contain the ancestral and derived residues, respectively: DmojIR75d.2-S670 maps to the third transmembrane domain and DmojIR75d.2-Q365 maps to the putative ligand binding domain. Functional characterisation of these variant receptors will be required to determine their significance.

Expansion of the IR repertoire by gene duplication and retroposition

From potentially one ancestral IR, what genetic processes underlay the generation of large repertoires of IR genes? We initially sought evidence for these mechanisms through analysis of the *D. melanogaster* IR family. Several monophyletic groups of IR genes exist in clusters in the genome suggesting an important role of gene duplication by non-allelic homologous recombination. For example, eight divergent IRs of the IR94 orthologous groups are located in three close, but separate, tandem arrays on chromosome arm 3R (Figure 8A). Other genes in the same clade are also found scattered on other chromosome arms (X, 2R, 3L) (Figure 6 and Figure 8A), indicating that interchromosomal translocation has also occurred frequently, most likely both during and after formation of the tandem arrays. Similar patterns are observed in the orthologous/paralogous sequences of these IRs in other drosophilid species (Figure 8A), as well as for other IR clades (data not shown). These features are also observed in IR repertoires in other insects, although incomplete genome assembly prevented a more precise analysis. For example, in *Aedes aegypti* the 23 IR7 clade members are found in arrays of 1, 1, 2, 5, 7 and 7 genes on 6 different supercontigs (data not shown).

We also noticed an unusual pattern in *D. melanogaster* IR gene structures, in which antennal IRs (as well as iGluRs) contain many (4–15) introns, while the vast majority of divergent IRs are single exon genes (Figure 8B). Drastic intron loss in multigene families is a hallmark of retroposition, where reverse-transcription of spliced mRNAs from parental, intron-containing genes and reinsertion of the resulting cDNA at a new genomic location may give rise to a functional, intronless retrogene [41]. The few introns that are

present in these IRs in *D. melanogaster* have a highly biased distribution towards the 5' end of the gene (19/25 introns in the first 50% of IR gene sequences) (Figure 8C), which is characteristic of recombination of partially reverse-transcribed cDNAs (a process which initiates at the 3' end) with parental genes [42]. Sequence divergence of IRs prevented us from identifying parental gene-retro gene relationships. Nevertheless, these observations together suggest that divergent IRs arose by at least one, and possibly several, retroposition events of ancestral antennal IRs. Once “born”, single exon IRs could presumably readily further duplicate by non-allelic homologous recombination.

Discussion

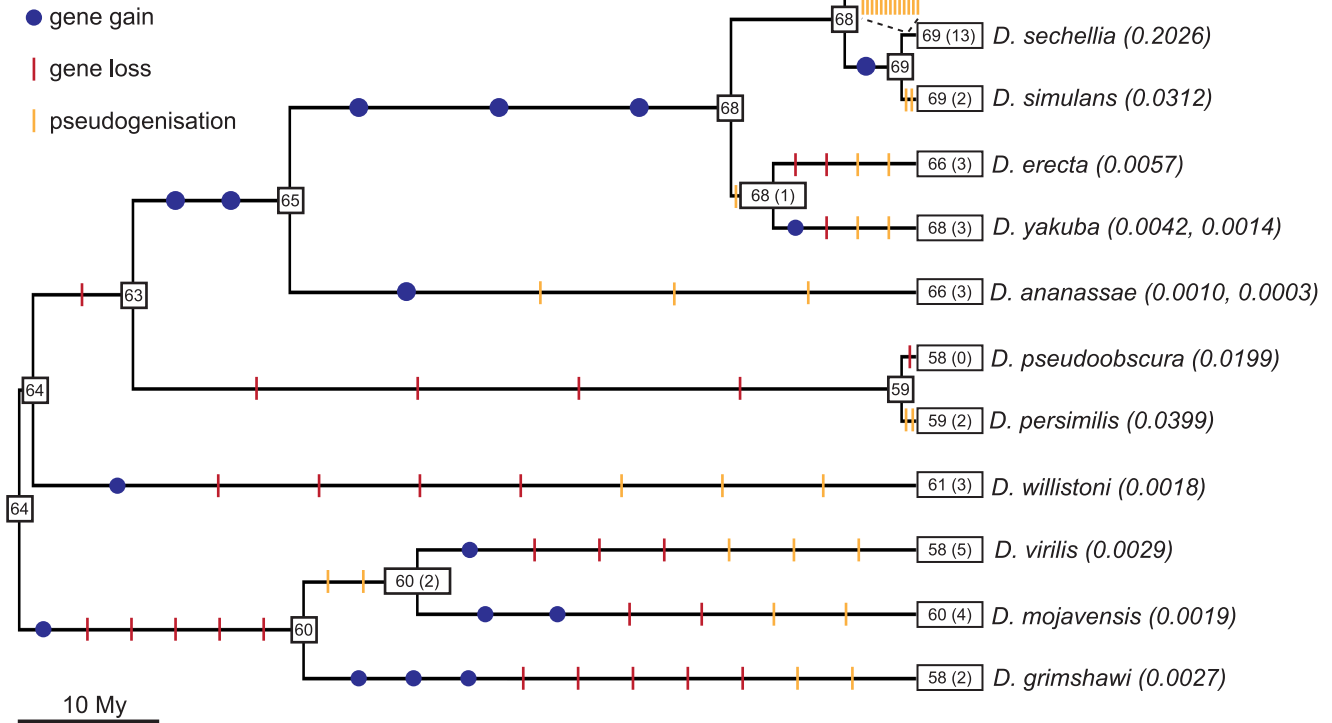
A model for iGluR and IR evolution

Our comprehensive survey and phylogenetic analysis of iGluR/IR-like genes permits development of a model for their evolution (Figure 9). The shared, unusual “S1-ion channel-S2” domain organisation of prokaryotic GluR0 and eukaryotic iGluRs is suggestive of a common ancestor of this family by fusion of genes encoding the separate domains that were present in very early life forms (Figure 9) [11]. However, we have found prokaryotic glutamate receptors in only a very small number of bacterial species. Thus, if an iGluR evolved in the common ancestor of prokaryotes and eukaryotes, it must have subsequently been lost in a large number of prokaryotic lineages. It is possible, therefore, that iGluRs only originated in eukaryotes and were acquired by certain prokaryotic species by horizontal gene transfer [43]. If the latter hypothesis is true, the presence of closely related iGluRs in both plants and animals implies their early evolution within eukaryotes, potentially in the last common eukaryotic ancestor [44]. However, the absence of iGluRs in sponges and all examined unicellular eukaryotes raises the alternative possibility that animal and plant receptors evolved independently, or were acquired by horizontal transmission, perhaps from prokaryotic sources. Whatever the precise origin of iGluRs in animals, their subsequent divergence into AMPA, Kainate and NMDA subfamilies also occurred early, although variation in the size of these subfamilies suggests continuous adaptation of the synaptic communication mechanisms they serve to nervous systems of vastly different complexities.

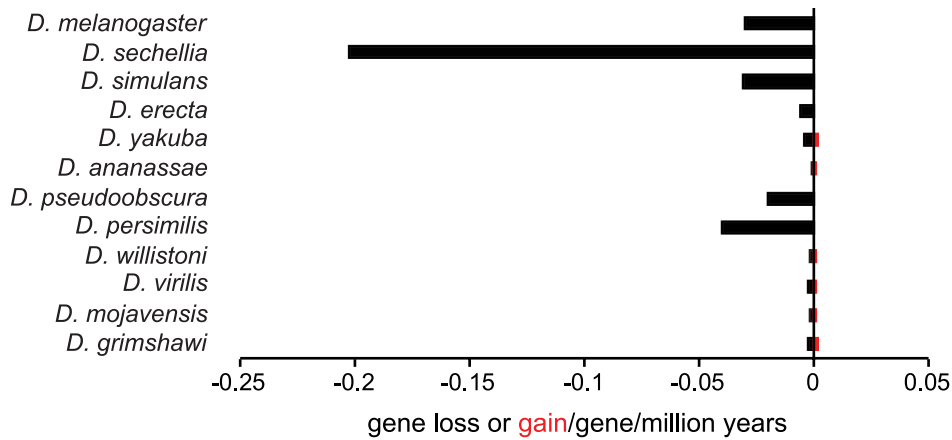
Several outstanding issues regarding IR evolution can now be addressed. First, we have shown that the IRs were very likely to have been present in the last common ancestor of Protostomia, an estimated 550–850 million years ago [20]. *IR25a* represents the probable oldest member of this repertoire and conservation of chemosensory organ expression of *IR25a* orthologues in molluscs, nematodes, crustaceans and insects strongly suggests that this receptor may have fulfilled a chemosensing function in the protostome ancestor.

Second, the apparent absence of IRs in Deuterostomia suggests the parsimonious model that IRs evolved from an animal iGluR ancestor rather than representing a family of chemosensing receptors that was present in a common ancestor of Animalia and lost in non-protostomes. Our phylogenetic and gene structure

A



B



C

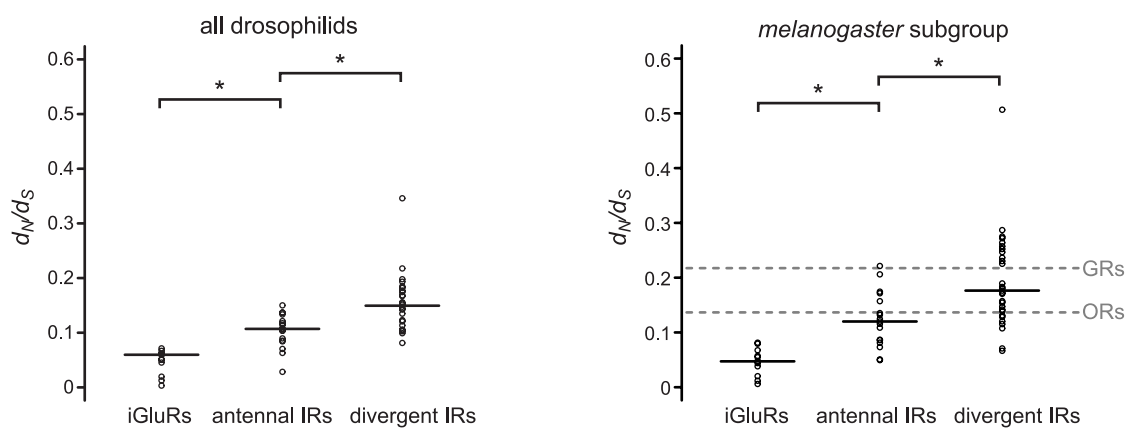


Figure 7. Gene loss and gain and selective pressures in drosophilid IR repertoires. (A) Estimates of the number of IR loci (number of pseudogenes is indicated in parentheses) on internal nodes of the drosophilid phylogeny and gene gain (blue dots), gene loss (red slashes) and pseudogenisation (orange slashes) events on each branch. The gene loss and gene gain rates on the terminal branches are indicated in parentheses after the species names. (B) Histogram of the gene gain (red) and loss (black) rates estimated for the terminal branches of the phylogeny. (C) Distribution and median (horizontal line) of d_N/d_S rates of iGluR and IR genes estimated for all twelve drosophilid species (left) or five *melanogaster* subgroup species (right). d_N/d_S values were significantly different between iGluRs, antennal IRs and divergent IRs ($p < 0.01$, Wilcoxon rank-sum tests). In the right-hand plot, the dashed grey lines represent the median values calculated from the d_N/d_S values for the *melanogaster* subgroup OR and GR genes, as reported in [39]. d_N/d_S values were significantly different both between antennal IRs and GRs and between divergent IRs and ORs ($p < 0.01$, Wilcoxon rank-sum tests).
doi:10.1371/journal.pgen.1001064.g007

arose simply by a change in expression of an iGluR from an interneuron (where it detected amino acid signals from a pre-synaptic partner) to a sensory neuron (where it could now detect chemical signals from the external environment).

Third, our analyses of IR repertoires across both divergent and relatively closely related species provide insights into the mechanistic basis for the expansion and functional diversification of the IR repertoire. Gene duplication by non-allelic homologous recombination is a widespread mechanism for growth of most multigene families in chemosensory systems [46], and this is also true for the IRs. Our implication of retroposition as a second mechanism in the evolution of IR repertoires offers two advantages for functional diversification. First, by arising from random re-insertion of reverse transcribed copies of parental genes, retrogenes normally lack endogenous promoter sequences, and can therefore potentially acquire novel expression patterns from genomic sequences flanking their insertion site that are distinct from their parental ancestor [41]. Indeed, in *D. melanogaster*, retrogene or retrogene-derived IRs - the divergent IRs - are apparently no longer expressed in antennal neurons like their ancestors, but instead in gustatory (and perhaps other) tissues. Second, release from the evolutionary constraints of the preservation of splicing signals near exon boundaries may have contributed to the more rapid divergence of the protein sequences of these intronless IRs [47].

Analysis of IR repertoires across the well-defined drosophilid phylogeny provides clear evidence for a birth-and-death model of evolution, in which, following gene duplication, individual family members progressively diverge in sequence and, in some cases, are lost by pseudogenisation and/or deletion [48–49]. Differential rates of these processes will ultimately shape the precise IR repertoire of an individual species (discussed below).

Evolutionary and functionally distinct IR subfamilies: olfactory and gustatory receptors, and ligand-binding receptors and co-receptors

Our molecular evolutionary analysis has distinguished two subfamilies in the IR repertoire: conserved, antennal IRs and the species-specific, divergent IRs. Their distinct evolutionary properties may correspond to fundamental functional differences, as we provide here the first evidence, to our knowledge, for expression of divergent IR subfamily members in subsets of neurons in both peripheral and internal gustatory organs at both adult and larval stages of *D. melanogaster*. The selective and non-overlapping expression patterns observed in the small sample of IR genes examined indicate that a large fraction of the divergent IR repertoire may be expressed in gustatory neurons. It is also possible that some of these IRs may be expressed in non-chemosensory tissues. Although subsets of GR genes have been implicated in the detection of sweet or bitter compounds in peripheral taste bristles in *D. melanogaster* [31], a comprehensive understanding of the physiological breadth and molecular logic of taste detection is lacking. Our results introduce further complexity

into the molecular mechanisms of taste detection and demand comprehensive and comparative expression and functional analysis of divergent IRs and GRs in this sensory system.

Although many gustatory-expressed divergent IRs in *D. melanogaster* are recently derived in drosophilids, the ancestral chemosensory function of IRs is likely to be not in the detection of airborne volatiles but rather water-soluble, non-volatile compounds, as the last common ancestor of Protostomia was probably aquatic. Indeed, the strikingly similar expression of IR genes in internal pharyngeal neurons in *D. melanogaster* and *C. elegans* suggests a conserved role for these receptors in sensing chemical signals from ingested food. In this light, the derivation of IRs from receptors detecting amino acid-related neurotransmitters invites the attractive hypothesis that ligands for these gustatory IRs (as well as species-specific IRs in other protostomes) are also amino acids. Almost nothing is known about sensory responses to this class of chemical signals in *D. melanogaster*, despite their vital importance for normal insect physiology and metabolism [50], but amino acids are chemosensory stimulants in other insects, lobsters and molluscs [51–53].

Our evolutionary and expression studies have highlighted IR25a as an atypical member of the repertoire, displaying deep conservation and broad expression in many olfactory and gustatory neurons. While we cannot exclude the possibility that IR25a recognises a specific chemical ligand, co-expression of this receptor with other cell-type specific IRs favours a model in which this acts as a co-receptor, analogous both to the heteromeric assembly of iGluR subunits into functional complexes [1], as well as to the pairing of ligand-specific ORs with the common OR83b co-receptor [54–55]. An insect- and antennal-specific homologue of IR25a, IR8a, may play a similar role specifically for olfactory IRs.

A common insect nose and species-specific IR repertoires

In addition to IR25a and IR8a, many other *D. melanogaster* antennal IRs are highly conserved in insects, both in sequence and expression pattern. These properties contrast starkly with the insect OR repertoires, which probably evolved only in terrestrial insects [56], and which contain only one member displaying orthology across multiple orders, the atypical OR83b co-receptor [57]. ORs are an expanded lineage of the ancestral GR repertoire whose evolutionary origins are unknown [56]. Homologues of GR genes exist in *D. pulex* and *C. elegans* [56,58], but in the latter species these receptors may not be involved in chemosensation [59–60]. These observations suggest that, in insects, the IRs represent the first olfactory receptor family, whose members were fixed functionally early in their evolution to detect olfactory stimuli that are important for all species of this animal class. Consistent with this, the antenna of the mayfly *Rhithrogena semicolorata* – an insect belonging to the Paleoptera and not the Neoptera that encompasses all species described here – bears coeloconic sensilla (potentially housing IR-expressing neurons) but not trichoid or basiconic sensilla (which house OR-expressing neurons in all other insects examined) [61]. Available data on ligands for IR sensory

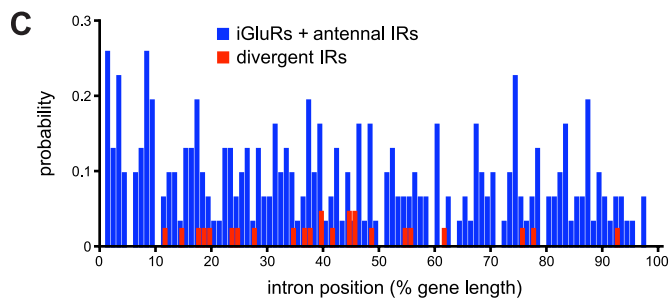
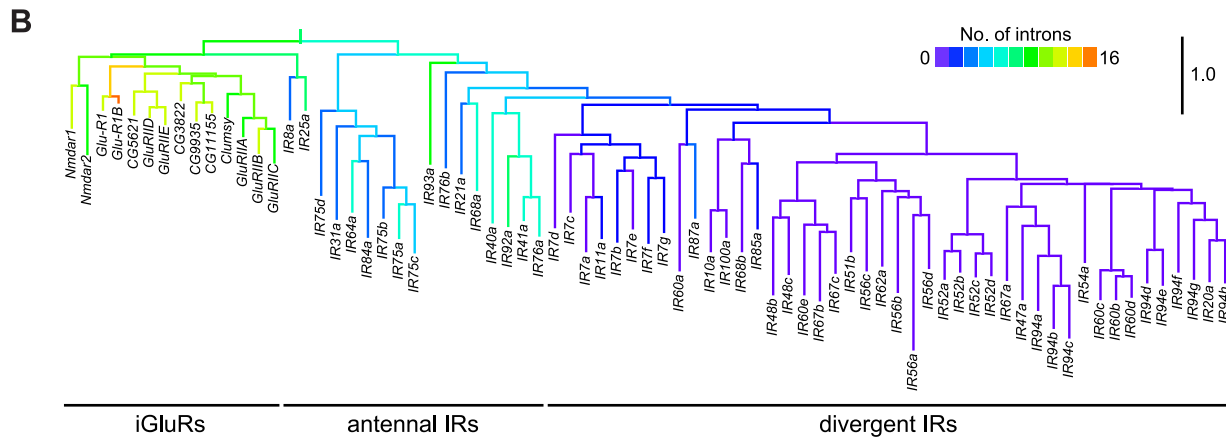
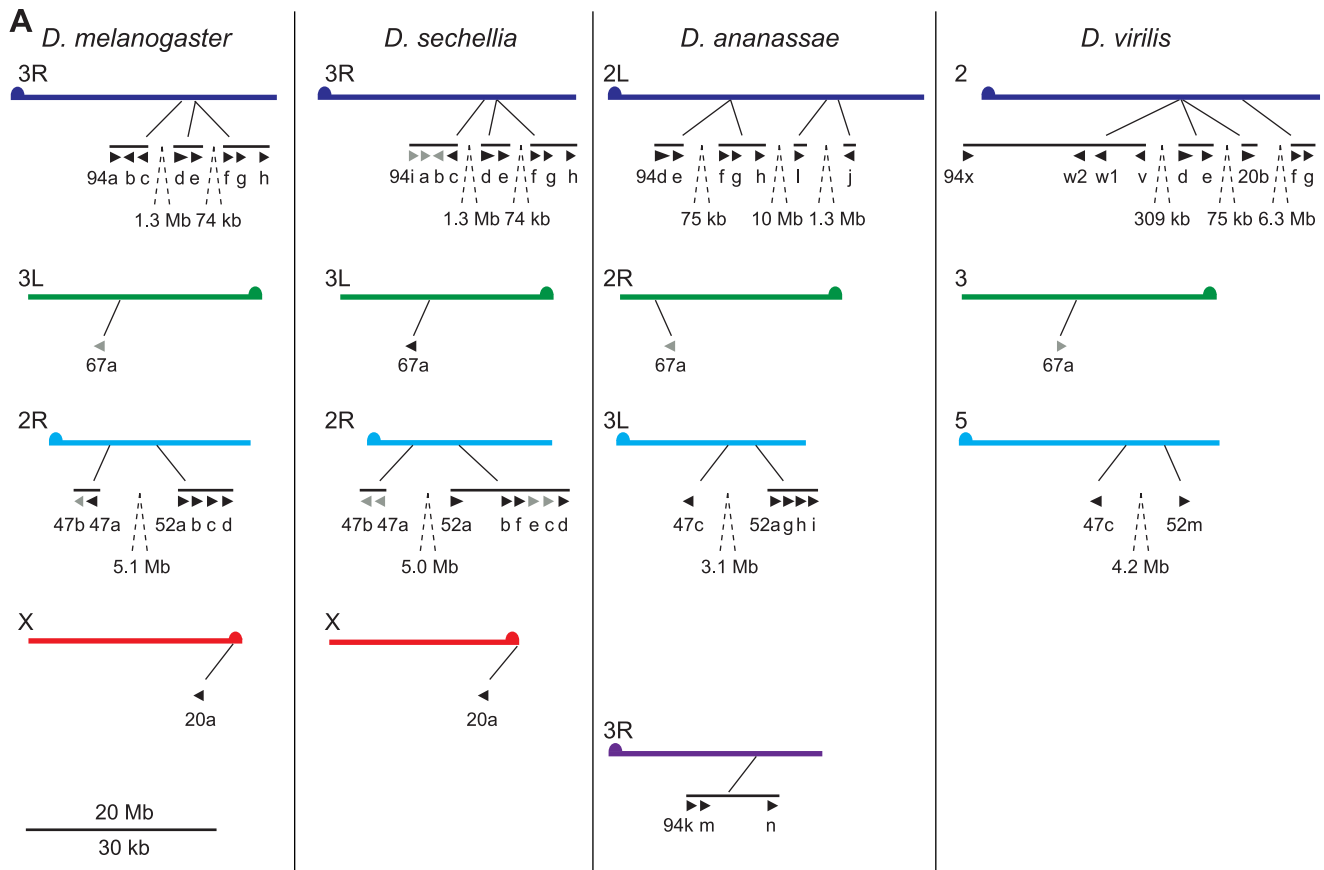


Figure 8. Mechanisms of IR repertoire expansion. (A) Genomic location of the IR genes (black arrowheads; pseudogenes in grey) belonging to the IR94 and IR52 clades in *D. melanogaster*, *D. sechellia*, *D. ananassae* and *D. virilis*. Equivalent chromosome arms (Muller elements) (labelled on the left of each chromosome arm) between the species are indicated by colour and horizontal alignment [93]. Tandem arrays of genes are indicated by horizontal black lines, and the distances between close arrays are shown. The “IR” and some number prefixes for gene names are omitted in clusters

where space is limiting. The scale bar represents 20 Mb for the chromosomes and 30 kb for gene lengths and distances between genes within the same tandem array. (B) Phylogenetic tree of *D. melanogaster* iGluRs and IRs, in which branches are colour-coded by the number of introns in each extant gene sequence or predicted ancestor. The tree was built with RAXML under the WAG model of substitution, with 1000 bootstrap replicates, and the colours representing intron numbers were inferred and displayed with Mesquite. Pseudogenes were excluded from this analysis. The scale bar represents the expected number of substitutions per site. (C) Histogram illustrating the distribution of intron positions as a percentage of protein length for iGluRs and antennal IRs (blue) and divergent IRs (red). Each bar represents the probability of occurrence of an intron at a given percentile of the protein.

doi:10.1371/journal.pgen.1001064.g008

neurons - and the role of specific IRs within these neurons - are limited, but include stimuli such as carboxylic acids, water and ammonia, which are known to be physiologically and behaviourally important in many insect species [62]. ORs, by contrast, may be primarily dedicated to detection of species-specific odour cues. In this light, the IRs are attractive molecular targets for novel, broad-spectrum chemical regulators of insect odour-driven behaviours, with applications in the control of disease vectors, such as mosquitoes, and agricultural pests.

Given the general conservation of the antennal IRs, what is the significance of the more recently evolved, species-specific variation in this family of chemosensory receptors? It is particularly informative to consider this question in the evolutionarily closely related drosophilid species. These display prominent differences in their global geographical distribution and chemosensory-driven behaviours [63–64], and include both generalists, which feed and

breed on a wide range of substrates, and specialists, which have highly restricted ecological niches. The chemical ecology is best-understood for *D. sechellia*, a species endemic to the Seychelles that utilises the acid-rich fruit of *Morinda citrifolia* as its sole food source and oviposition site, a remarkable specialisation as this fruit is repulsive and toxic for other drosophilids [64–65]. Genetic hybrids between *D. sechellia* and *D. simulans* indicate that host specialisation is due to loss-of-function mutations, rather than gain of new chemosensory perception abilities [65]. The accelerated rate of IR gene loss in *D. sechellia* compared to its sibling *D. simulans* (and other drosophilids) bears the hallmark of genetic adaptation of this chemosensory repertoire to the restricted host fruit. Notably, one of the *D. sechellia* pseudogenes is *IR75a*, an antennal IR expressed in a neuron responsive to several acids [62]. Thus, *DsecIR75a* represents an interesting gene whose mutation may be directly linked to host specialisation of this species. Future study of this

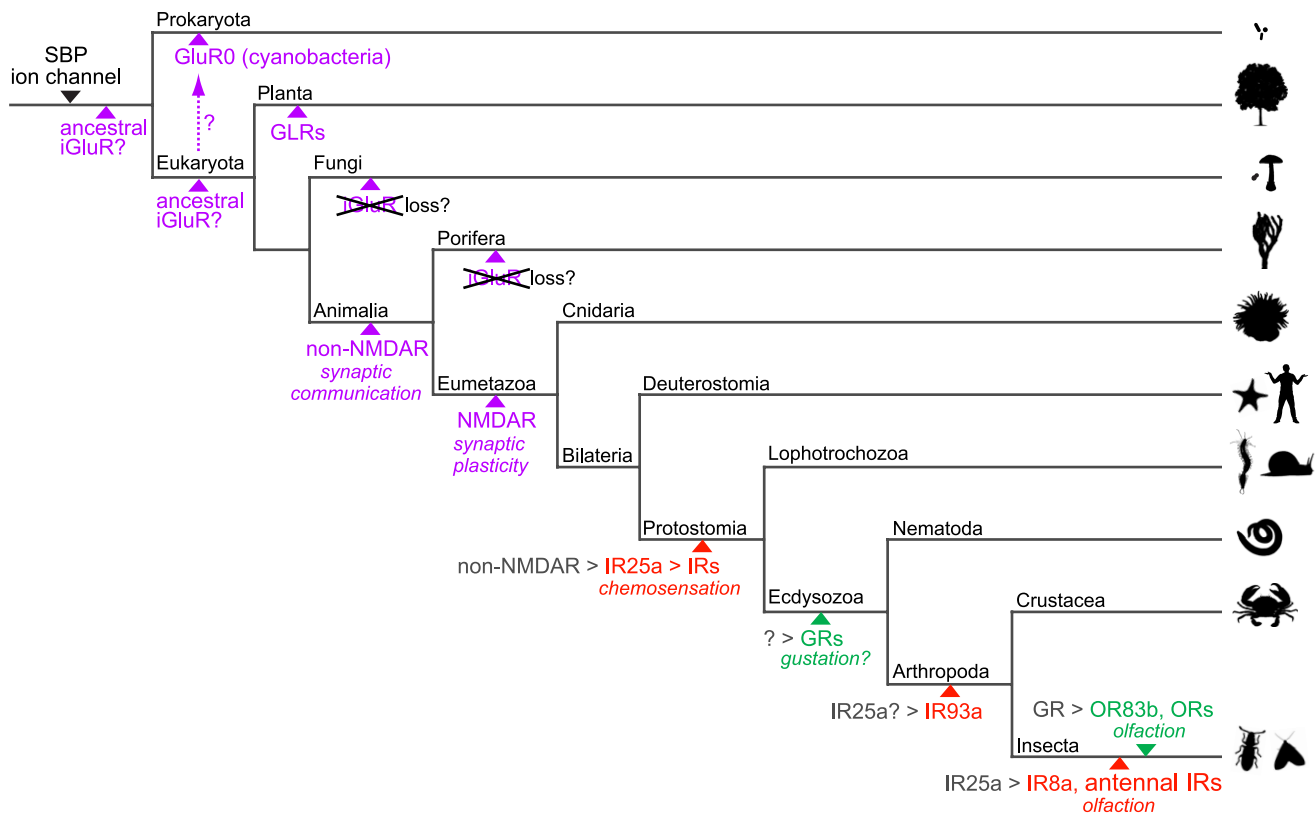


Figure 9. A model for the evolution of iGluRs and IRs. Schematic phylogenetic tree highlighting the branches along which specific gene families or genes appeared with their putative functions, inferred from their presence or absence in sequenced genomes of extant species (see Figure 1). Solute binding proteins (SBPs, which exhibit the same protein fold as the iGluR/IR amino terminal domain and ligand-binding domain) and ion channels were likely present in primitive life forms as related protein domains exist in Eukaryota, Bacteria and Archaea [94]. iGluRs are shown in purple, IRs in red and insect GRs and ORs in green. Various speculative models for the origins of iGluRs are shown. Putative genetic ancestors from which IRs, GRs and ORs derived are shown in grey followed by a ">" symbol. The resolution of the phylogeny is necessarily biased towards invertebrate lineages and branch lengths contain no temporal information.

doi:10.1371/journal.pgen.1001064.g009

receptor, and other species-specific IRs, may offer novel models to link genetic changes with phenotypic adaptation during animal evolution.

Genetic insights into the origins of animal olfactory systems

Finally, our results may shed light into the outstanding question of the evolutionary origin of animal olfactory systems. Common neuroanatomical features have long been appreciated in animal olfactory circuitry, notably glomeruli, which represent sites of synaptic connection of OSNs of identical molecular and physiological specificity with second order neurons [66]. Whether these represent homologous or analogous structures across phyla is unclear. Revelations of fundamental distinctions in the structure, function and regulation of mammalian and insect ORs support a theory of convergent evolution of the neuronal circuits in which these receptors act [67–68].

Our demonstration that most, if not all, insect olfactory systems comprise two molecularly distinct receptor families, the ORs and IRs, indicates that the evolution of receptor repertoires can be uncoupled from a presumed common origin of the OR and IR neuronal circuits within the insect ancestor. Thus, during a significantly greater timescale across animal phyla, profound molecular differences between olfactory receptor genes do not necessarily imply distinct evolutionary origins of the neuronal circuitry in which they are expressed. Our discovery of IRs in mollusc olfactory organs reveals this to be an interesting potential “hybrid” organism in olfactory system evolution. The *A. californica* rhinophore and oral tentacle also express a large family of GPCR-family candidate chemosensory receptors, belonging to the same Rhodopsin superfamily as vertebrate ORs [21]. The co-existence of both insect-like and vertebrate-like olfactory receptors in this species provides evidence for the occurrence of an evolutionary transition between these distinct olfactory receptor families. Thus, while extant animal olfactory systems display an enormous diversity in their receptor repertoires, there may remain - perhaps unexpectedly - a sufficient genetic trace within receptor gene families themselves to open the possibility of a common evolutionary origin of this sensory system.

Materials and Methods

Gene identification and annotation

Eukaryota (non-drosophilids). Genomic and available annotated protein databases for each eukaryotic species were downloaded from the sources described in Table S2 (spring 2009 versions). Prokaryotic genome and protein sequences were downloaded from NCBI. We built and calibrated an HMM with HMMER [69] for iGluR/IR gene identification by adding sequences of the *D. melanogaster* PF00060 domain (iGluR ligand-gated ion channel) to those of the PF00060 domains from the Pfam database [17]. This HMM (LC05) was used to screen protein databases using HMMER. For each species, all significant hits (HMMER E value $< e^{-5}$) were subsequently used, in addition to *D. melanogaster* iGluR and IR sequences [15], as queries in exhaustive PSI-BLAST searches with standard parameters until convergence. All identified sequences (below an arbitrary threshold E value $< e^{-5}$) were then used as queries in TBLASTN searches of genomic DNA databases. For each DNA hit (E value $< e^{-3}$), we analysed a genomic region of approximately 20 kb spanning this sequence for the presence of a *bona fide* iGluR or IR gene, by using the LC05 HMM and homology analysis with *D. melanogaster* iGluRs and IRs to annotate exons in these regions using GeneWise [70]. Predicted proteins were verified by analysing the number and placement of

transmembrane segments using the TMHMM Server v2.0 [71], and domain organisation using the Pfam database. Most annotated sequences (Datasets S1 and S2) appear to be incomplete at their 5' ends as they do not encode N-terminal signal sequences, as determined by analysis with SignalP 3.0 [72], and we were normally not able to annotate this part of the protein with confidence. However, as this region is highly divergent in amino acid sequence, its absence is likely to have little influence on our phylogenetic analyses.

Drosophilids. *D. melanogaster* iGluR and IR sequences were used as queries in exhaustive PSI-BLAST and TBLASTN searches of the genome assemblies described in Table S2. PSI-BLAST was carried out for 20 iterations or until no new sequences with an E value $< e^{-3}$ were recovered. For genes that were apparently missing in some species, we used manual syntenic analysis to determine whether this represented a real absence from the genome. Genes were manually reannotated to ensure the presence of appropriate structural features as described above, as well as reasonable splice site signals and start/stop codons. Missing or mis-annotated exons in one species were usually easily corrected by comparison with homologous sequences in other species. Genes containing nonsense mutations were manually resequenced (see below) to confirm or refute their annotation as pseudogenes (Table S3). We also resequenced parts of genes where there were gaps in the genome assembly (Table S3).

Phylogenetic analyses

Protein tree building. The amino acid sequences of the selected iGluRs/IRs were aligned with PROBCONS [73] and examined in Jalview [74]. The alignments were cleaned manually to obtain final high-quality alignments of 150–300 residues, depending on the sequences analysed (see Dataset S3 for all alignments pre- and post-cleaning). We used ProtTest [75] to assess the best model of substitution to infer the phylogeny. The trees were then calculated with PhyML [76] or RAxML [77] and viewed and graphically edited with FigTree (tree.bio.ed.ac.uk), Mesquite [78] or iTOL [79]. For trees of drosophilid iGluRs/IRs, pseudogenes and incomplete genes were excluded from alignments, and we applied the JTT model of amino acid substitution in PhyML. Bootstrap values were estimated using an approximate likelihood ratio test.

Character matrix tree building. Selected protein sequences were aligned using MUSCLE [80] and the positions of introns were reported on the alignment. A character matrix was built according to the presence of introns at each potential intronic site. The tree was built using the PARS software from the PHYLIP package (evolution.genetics.washington.edu/phyip.html).

Orthology determination. Genes were defined as orthologous when they were best reciprocal BLAST hits and when they grouped in the same clade in phylogenetic trees. Because we could not unambiguously assign orthologues to some IRs, we classified those genes as members of larger orthologous groups encompassing several members in some species.

Gene and protein nomenclature

IR genes were named according to a unified nomenclature system based upon a foundation of the cytologically derived *D. melanogaster* IR gene names [15]. Receptor names are preceded by a four-letter species abbreviation consisting of an uppercase initial letter of the genus name and three lower case initial letters of the species name (e.g. *Anopheles gambiae* = *Agam*; *Daphnia pulex* = *Dpul*). Orthologues of *D. melanogaster* sequences are given the same name (e.g. *CquiIR25a*, *AcallIR25a*). If multiple copies of an orthologue of a *D. melanogaster* gene exist for a species (based on sequence, not

function), they are given the same name followed by a point and a number (e.g. *ApisIR75d.1*, *ApisIR75d.2*). If several in-paralogues exist both in *D. melanogaster* and other species, these are all given the same number (indicating their grouping within a common clade), but different final letterings. For novel, species-specific IRs, we defined new names numbering from 101 upwards to avoid confusion with *D. melanogaster* gene names, which number up to IR100a. For species-specific IRs that form monophyletic clades and had high (>60%) amino acid identity, we gave these the same name with an additional number suffix after a point (e.g. *AaegIR75e.1*, *AaegIR75e.2*). We did not rename genes with previously published names (e.g. *C. elegans* GLR-7 and GLR-8 [9]).

For vertebrate iGluRs, we used the NC-IUPHAR nomenclature [81]: each species name is followed by “Glu”, a letter representing the subtype of the receptor (K for Kainate, A for AMPA and N for NMDA), and a number, reflecting predicted orthology with mammalian iGluRs. We did not name (or rename) invertebrate iGluRs in this study, except for newly predicted gene sequences (Table S3), where logical variants of NC-IUPHAR nomenclature were assigned.

Evolutionary analysis

Gene birth and death rate estimation. To estimate the gene birth and death rates of IRs on the drosophilid phylogeny we used the gene numbers listed in Table S1. Incomplete genes (i.e. genes for which we could not annotate full-length sequences because of lack of sequence data) were classified as present. To estimate the number of gene gain and loss events for each orthologous group we estimated gene numbers on internal branches using a maximum likelihood method [82] implemented in the software CAFÉ [83]. These numbers were then summed to estimate the number of IR gene gains and losses on each branch of the phylogeny. The divergence times for the species tree were taken from the published estimates [40,84]. The gene birth and death rates per million years on the terminal branches were calculated as number of gene losses or gene gains divided by the number of genes on the respective internal node divided by the length in million years of the respective terminal branch. The gene death rates, D , averaged over the whole species tree were calculated as in [85]: $D = \sum_{i=1}^n \left(\frac{L_i}{C_i} \right) / t$, where n is the number of branches in the tree, L_i is the number of gene losses on branch i , C_i is the number of gene copies at the internal node of branch i and t is the total time of the phylogeny. For the estimation of the gene gain rate, B , L_i was replaced by the numbers of gene gains, G_i , on branch i .

Analysis of selective forces. We inferred the d_N/d_S ratio (ω) by maximum likelihood as implemented in PAML [86]. All PAML analyses were run three times using different input parameters to avoid local optima. To create multiple sequence alignments of orthologous genes, we first aligned the amino acid sequences using MUSCLE. Pseudogenes and incomplete genes were avoided in these analyses, and if genes had multiple annotated isoforms we used only those conserved with the other species. The resulting alignments were then used to guide the nucleotide coding region alignments using custom-written software [87]. Columns with gaps were omitted for the d_N/d_S calculations. For all analyses, we assumed the topology illustrated in Figure 7A. We applied model M0 to estimate the global selective pressure acting on the IR and iGluR genes. To compare our data with a previous analysis of drosophilid ORs [39], we applied a branch model to estimate the global selective pressure acting on the IR and iGluR genes. In this model, one d_N/d_S ratio was assigned to the five *melanogaster* subgroup species and one

ratio was assigned to *D. ananassae* (model = 2, NSsites = 0). The *D. ananassae* ratio was then discarded to leave one d_N/d_S ratio depicting the selective pressure acting on the respective gene in the *melanogaster* subgroup.

To identify positively selected sites we applied models M7 (beta) and M8 (beta & ω) in PAML and compared them using a maximum likelihood ratio test (LRT). If M8 fitted the data significantly better than M7, we applied a Bayes Empirical Bayes (BEB) estimation method as implemented in PAML to identify the sites that are estimated to be under positive selection.

We applied another test to analyse the duplication of *IR75d* in *D. mojavensis*. To test if residues of these genes evolved under positive selective pressure, we first compared a model that assigns one single d_N/d_S ratio to all branches with a model that assigns one additional ratio to the branches following the duplication. If this second model fitted the data significantly better than the first model we used branch-site model A (model = 2, NSsites = 2) with $\omega = 1$ fixed on the branches after the duplication as null model and compared it to this same model A but allowing $\omega > 1$ on the branches following the duplication. To estimate the age of the *IR75d* duplication in *D. mojavensis*, we applied model M0 to estimate d_S on the branch before the duplication and on the two branches after the duplication. By relating these d_S values to each other and assuming a divergence time of 35 My between *D. mojavensis* and *D. virilis*, we obtained two estimates of the timing of the duplication event.

Re-sequencing of drosophilid IR genes

Genomic DNA was extracted from the sequenced drosophilid genome strains (obtained from the *Drosophila* Species Stock Center, University of California-San Diego) using a standard DNA extraction protocol. PCR primers were designed to amplify ~500 bp regions covering putative nonsense or missense mutations or spanning gaps in the genome sequence (oligonucleotide sequences are listed in Table S5). PCR amplifications were performed using Taq DNA Polymerase (PEQLAB Biotechnologie GmbH) in a MasterCycler Gradient Thermocycler (Eppendorf) with the following programme: 95°C for 3 min, 35 cycles of (95°C for 30 sec, 55°C for 1 min, 72°C for 1 min) and 72°C for 10 min, with minor modifications of annealing temperature and elongation times for different primer pairs and amplicon sizes. Products were gel purified (Machery-Nagel) and sequenced with BigDye Terminator v3.1 according to the manufacturers' protocols.

Reverse-transcription PCR

Insects: total RNA was extracted from hand-dissected tissues of wildtype *A. mellifera* and *D. melanogaster* (w^{1118} strain) using the RNeasy Mini Kit (Qiagen), and reverse-transcribed using oligo-dT primers and the SuperScript III First-Strand Synthesis System (Invitrogen). Genomic DNA was extracted using standard procedures. Primers were designed to amplify short regions overlapping an intron, if possible at the 3' end of the coding sequence (Table S5). PCR product amplification and purification were performed as described above and sequenced to verify their identity. Multiple independent cDNA preparations were analysed for each primer pair.

Aplysia. Mature *Aplysia dactylomela* (100–300 g) were collected from Kings Beach, Caloundra, Queensland, Australia. Animals were anaesthetised in 337 mM MgCl₂ equivalent to 50% of their weight. Tissues were removed and snap frozen in liquid nitrogen for RNA isolation. Adult *Aplysia californica* (100–500 g) were obtained from Marine Research and Educational Products (Escondido, CA, USA), and the rhinophore was removed and

stored in RNAlater (Qiagen). Total RNA was extracted from samples using TRI Reagent (Sigma). One μg of total RNA was treated with DNase I (Invitrogen) and cDNA was synthesised from 0.5 μg DNase-treated RNA using 200 ng random pentadecamers and the Superscript III Reverse Transcriptase System (Invitrogen). No-RT controls were also carried out for each RNA sample using 0.5 μg DNase-treated RNA to confirm the absence of genomic DNA contamination. PCR amplification using primer pairs for individual *Aplysia* IRs or for a β -actin control (Table S5) were performed using REDTaq DNA polymerase (Sigma) according to the manufacturer's protocol.

Construction of *IR-GAL4* transgenes

Primers were designed to amplify putative promoter regions from *Oregon-R D. melanogaster* genomic DNA with flanking restriction sites, extending from immediately upstream of the predicted start codon to the following 5' extents: *IR7a* (2318 bp), *IR11a* (2099 bp), *IR52b* (446 bp), *IR56a* (2400 bp) and *IR100a* (512 bp) (Table S5). Gel purified PCR products were T:A cloned into *pGEM-T Easy* (Promega), end-sequenced, and subcloned into a *pGAL4-attB* vector, comprising the *GAL4 ORF-hsp70-3' UTR* in the *patB* vector [30]. These constructs were integrated into the attP2 landing site [88], by standard transformation procedures (Genetic Services, Inc.). *IR-GAL4* transgenic flies were double-balanced and crossed with flies bearing a *UAS-mCD8:GFP* transgene [89] to visualise driver expression.

Histology

RNA *in situ* hybridisation on *Aplysia*. A 743 bp region of *A. dactylopera IR25a* cDNA was amplified and cloned into *pGEM-T* (Promega) as a template for synthesis of sense and antisense digoxigenin-labelled RNA probes (Roche). *In situ* hybridisation on 12 μm rhinophore cryosections was performed essentially as described [90]. Sections were photographed using an Olympus BX60 with Nomarski optics and a Nikon Digital Sight DS-U1 camera.

Immunofluorescence on larval and adult *Drosophila*. Third instar larvae were placed in a Petri dish containing 1 \times PBS/0.1% Triton (P/T) and their head regions containing chemosensory organs were removed with forceps. For adults, probosci were pulled off the head with forceps and the labellum and the more proximal parts separated. Dissected tissues were placed in a 1.5 ml microcentrifuge tube and fixed in 4% PFA in 1 \times PBS for 1 hour at 4°C, washed 3 \times 10 minutes in P/T, blocked for 30 minutes in 5% heat-inactivated goat serum in P/T (P/T/S) and incubated overnight at 4°C with mouse anti-GFP (Invitrogen) and rabbit anti-IR25a [15], both diluted to 1:500 in P/T/S. Tissues were washed and blocked as above and incubated with Alexa488-anti mouse and Cy3-anti rabbit secondary antibodies (Milan Analytica AG), both diluted to 1:500 in P/T/S for 2 hours at room temperature. Samples were mounted on glass slides with 100 μl Vectashield. Images were collected with a Zeiss LSM 510 Meta upright confocal microscope (Zeiss, Oberkochen, Germany), using a Plan-APOCHROMAT 63 \times /1,40 Oil DIC objective.

Supporting Information

Dataset S1 iGluR and IR predicted protein sequences. Sequences are in FASTA format. The header line of each sequence displays i) the new sequence name (except for previously annotated non-vertebrate iGluRs), ii) the old sequence name (for

previously annotated sequences) and, in some cases, iii) comments, separated by spaces. Internal stop codons and frameshifts are indicated by an 'X'. Unknown residues (due to gaps in genomic sequence data) are indicated by an 'x'.

Found at: doi:10.1371/journal.pgen.1001064.s001 (1.16 MB TXT)

Dataset S2 iGluR and IR predicted transcripts. Sequences are in FASTA format. The header line of each sequence displays the new sequence name, except for previously annotated non-vertebrate iGluRs.

Found at: doi:10.1371/journal.pgen.1001064.s002 (3.40 MB TXT)

Dataset S3 Alignments used for phylogeny. This folder contains the multiple sequence alignments used for phylogenetic analyses, before and after alignment cleaning in FASTA and PHYLIP format, respectively, as well as the intron alignment file used in Figure 2D.

Found at: doi:10.1371/journal.pgen.1001064.s003 (1.21 MB ZIP)

Table S1 Drosophilid iGluR and IR repertoires.

Found at: doi:10.1371/journal.pgen.1001064.s004 (0.04 MB XLS)

Table S2 Sources of eukaryotic genomic and protein sequence data.

Found at: doi:10.1371/journal.pgen.1001064.s005 (0.70 MB DOC)

Table S3 Nomenclature of newly annotated and previously identified iGluR and IR genes.

Found at: doi:10.1371/journal.pgen.1001064.s006 (0.19 MB XLS)

Table S4 Nonsynonymous to synonymous substitution rates of IR genes.

Found at: doi:10.1371/journal.pgen.1001064.s007 (0.03 MB XLS)

Table S5 Oligonucleotides.

Found at: doi:10.1371/journal.pgen.1001064.s008 (0.05 MB XLS)

Acknowledgments

We thank Jason Pitts and Larry Zwiebel for communication and discussion on their independent identification and annotation of *A. gambiae* IRs. For provision of sequence data resources (Table S2), we gratefully acknowledge the US Department of Energy Joint Genome Institute (jgi.doe.gov), the Wellcome Trust Sanger Institute (sanger.ac.uk), Vectorbase (vectorbase.org), Flybase (flybase.org), the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program (Federica Di Palma, and Kerstin Lindblad-Toh), and B. Degnan. We thank Liliane Abuin for assistance in generating *IR-GAL4* transgenes, Giovanni Galizia and Kevin Wanner for *A. mellifera*, and the Vital-IT (vital-it.ch) Center of the Swiss Institute of Bioinformatics for computational support. We are grateful to Hugh Robertson for advice on gene annotation and nomenclature, Marco Pagni for advice on bioinformatic methods, and Bernard Degnan for advice and support. We are also grateful to Carolina Gomez-Diaz, Sophie Martin, Gregg Nagle, Pavan Ramdya, Michael Saina, Ana Silbering, and other members of the Benton lab for discussions and comments on the manuscript.

Author Contributions

Conceived and designed the experiments: VC RR SFC AB DB HK TJG RB. Performed the experiments: VC RR SFC. Analyzed the data: VC RR SFC AB RB. Contributed reagents/materials/analysis tools: VC RR SFC AB DB RB. Wrote the paper: VC RR RB.

References

- Gereau RW, Swanson GT (2008) The Glutamate Receptors. Totowa N.J.: Humana Press. pp xi, 576.
- Sobolevsky AI, Rosconi MP, Gouaux E (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* 462: 745–756.
- Jin R, Singh SK, Gu S, Furukawa H, Sobolevsky AI, et al. (2009) Crystal structure and association behaviour of the GluR2 amino-terminal domain. *Embo J* 28: 1812–1823.
- Armstrong N, Sun Y, Chen GQ, Gouaux E (1998) Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* 395: 913–917.
- Kuner T, Seeburg PH, Guy HR (2003) A common architecture for K⁺ channels and ionotropic glutamate receptors? *Trends Neurosci* 26: 27–32.
- Mayer ML, Armstrong N (2004) Structure and function of glutamate receptor ion channels. *Annu Rev Physiol* 66: 161–181.
- Tikhonov DB, Magazanik LG (2009) Origin and molecular evolution of ionotropic glutamate receptors. *Neurosci Behav Physiol* 39: 763–773.
- Moroz LL, Edwards JR, Puthanveetil SV, Kohn AB, Ha T, et al. (2006) Neuronal transcriptome of aplysia: neuronal compartments and circuitry. *Cell* 127: 1453–1467.
- Brockie PJ, Madsen DM, Zheng Y, Mellem J, Maricq AV (2001) Differential expression of glutamate receptor subunits in the nervous system of *Caenorhabditis elegans* and their regulation by the homeodomain protein UNC-42. *J Neurosci* 21: 1510–1522.
- Littleton JT, Ganetzky B (2000) Ion channels and synaptic organization: analysis of the *Drosophila* genome. *Neuron* 26: 35–43.
- Chen GQ, Cui C, Mayer ML, Gouaux E (1999) Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* 402: 817–821.
- Lam HM, Chiu J, Hsieh MH, Meisel L, Oliveira IC, et al. (1998) Glutamate-receptor genes in plants. *Nature* 396: 125–126.
- Chiu JC, Brenner ED, DeSalle R, Nitabach MN, Holmes TC, et al. (2002) Phylogenetic and expression analysis of the glutamate-receptor-like gene family in *Arabidopsis thaliana*. *Mol Biol Evol* 19: 1066–1082.
- Qi Z, Stephens NR, Spalding EP (2006) Calcium entry mediated by GLR3.3, an *Arabidopsis* glutamate receptor with a broad agonist profile. *Plant Physiol* 142: 963–971.
- Benton R, Vannice KS, Gomez-Diaz C, Voshall LB (2009) Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136: 149–162.
- Voshall LB, Stocker RF (2007) Molecular Architecture of Smell and Taste in *Drosophila*. *Annu Rev Neurosci* 30: 505–533.
- Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–288.
- Sakarya O, Armstrong KA, Adamska M, Adamski M, Wang IF, et al. (2007) A post-synaptic scaffold at the origin of the animal kingdom. *PLoS One* 2: e506. 10.1371/journal.pone.0000506.
- Wiegmann BM, Trautwein MD, Kim JW, Cassel BK, Bertone MA, et al. (2009) Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol* 7: 34.
- Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22: 2971–2972.
- Cummins SF, Erpenbeck D, Zou Z, Claudianos C, Moroz LL, et al. (2009) Candidate chemoreceptor subfamilies differentially expressed in the chemosensory organs of the mollusc *Aplysia*. *BMC Biol* 7: 28.
- Elliott CJ, Susswein AJ (2002) Comparative neuroethology of feeding control in molluscs. *J Exp Biol* 205: 877–896.
- Preston RJ, Lee RM (1973) Feeding behavior in *Aplysia californica*: role of chemical and tactile stimuli. *J Comp Physiol Psychol* 82: 368–381.
- Hollins B, Hardin D, Gimelbrant AA, McClintock TS (2003) Olfactory-enriched transcripts are cell-specific markers in the lobster olfactory organ. *J Comp Neurol* 455: 125–138.
- Stepanyan R, Hollins B, Brock SE, McClintock TS (2004) Primary culture of lobster (*Homarus americanus*) olfactory sensory neurons. *Chem Senses* 29: 179–187.
- Albertson DG, Thomson JN (1976) The pharynx of *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* 275: 299–325.
- Scott K, Brady R, Jr., Cravchik A, Morozov P, Rzhetsky A, et al. (2001) A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* 104: 661–673.
- Clyne PJ, Warr CG, Carlson JR (2000) Candidate Taste Receptors in *Drosophila*. *Science* 287: 1830–1834.
- Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118: 401–415.
- Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* 104: 3312–3317.
- Montell C (2009) A taste of the *Drosophila* gustatory receptors. *Curr Opin Neurobiol* 19: 345–353.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.
- Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, et al. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* 15: 1–18.
- Russo CA, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* 12: 391–404.
- Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19: 748–761.
- McDermott SR, Kliman RM (2008) Estimation of isolation times of the island species in the *Drosophila simulans* complex from multilocus DNA sequence data. *PLoS One* 3: e2442. 10.1371/journal.pone.0002442.
- Harrison PM, Milburn D, Zhang Z, Bertone P, Gerstein M (2003) Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res* 31: 1033–1037.
- Petrov DA, Chao YC, Stephenson EC, Hartl DL (1998) Pseudogene evolution in *Drosophila* suggests a high rate of DNA loss. *Mol Biol Evol* 15: 1562–1567.
- McBride CS, Arguello JR, O'Meara BC (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177: 1395–1416.
- Powell JR (1997) Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press.
- Kaessmann H, Vinckenbosch N, Long M (2009) RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet* 10: 19–31.
- Coulombe-Huntington J, Majewski J (2007) Intron loss and gain in *Drosophila*. *Mol Biol Evol* 24: 2842–2850.
- Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* 55: 709–742.
- Chiu J, DeSalle R, Lam HM, Meisel L, Coruzzi G (1999) Molecular evolution of glutamate receptors: a primitive signaling mechanism that existed before plants and animals diverged. *Mol Biol Evol* 16: 826–838.
- Shaham S (2009) Chemosensory organs as models of neuronal synapses. *Nat Rev Neurosci*.
- Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9: 951–963.
- Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD (2007) Splicing and the evolution of proteins in mammals. *PLoS Biol* 5: e14. 10.1371/journal.pbio.0050014.
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39: 121–152.
- Sanchez-Gracia A, Vieira FG, Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103: 208–216.
- Grandison RC, Piper MD, Partridge L (2009) Amino-acid imbalance explains extension of lifespan by dietary restriction in *Drosophila*. *Nature* 462: 1061–1064.
- Shiraishi A, Kuwabara M (1970) The effects of amino acids on the labellar hair chemosensory cells of the fly. *J Gen Physiol* 56: 768–782.
- Chiken S, Kuwasawa K, Kurokawa M, Ohsuga K (2001) Amino acid-induced reflexes and their neural pathways in an opisthobranch mollusc *Pleurobranchaea japonica*. *Zoological Science* 18: 456–473.
- Burgess MF, Derby CD (1997) Two novel types of L-glutamate receptors with affinities for NMDA and L-cysteine in the olfactory organ of the Caribbean spiny lobster *Panulirus argus*. *Brain Res* 771: 292–304.
- Benton R, Sachse S, Michnick SW, Voshall LB (2006) Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol* 4: e20. 10.1371/journal.pbio.0040020.
- Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, et al. (2004) *Or33b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* 43: 703–714.
- Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 100 Suppl 2: 14537–14542.
- Jones WD, Nguyen TA, Kloss B, Lee KJ, Voshall LB (2005) Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr Biol* 15: R119–121.
- Penalva-Arana DC, Lynch M, Robertson HM (2009) The chemoreceptor genes of the waterflea *Daphnia pulex*: many Grs but no Ors. *BMC Evol Biol* 9: 79.
- Edwards SL, Charlie NK, Milfort MC, Brown BS, Gravlin CN, et al. (2008) A novel molecular solution for ultraviolet light detection in *Caenorhabditis elegans*. *PLoS Biol* 6: e198. 10.1371/journal.pbio.0060198.
- Moresco JJ, Koelle MR (2004) Activation of EGL-47, a Galphao-coupled receptor, inhibits function of hermaphrodite-specific motor neurons to regulate *Caenorhabditis elegans* egg-laying behavior. *J Neurosci* 24: 8522–8530.
- Rebora M, Piersanti S, Gaino E (2009) The antennal sensilla of adult mayflies: *Rhithrogena semicolorata* as a case study. *Micron* 40: 571–576.
- Yao CA, Ignell R, Carlson JR (2005) Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *J Neurosci* 25: 8359–8367.
- Markow TA, O'Grady PM (2007) *Drosophila* biology in the genomic age. *Genetics* 177: 1269–1276.
- Stensmyr MC (2009) *Drosophila sechellia* as a model in chemosensory neuroecology. *Ann N Y Acad Sci* 1170: 468–475.
- Jones CD (2005) The genetics of adaptation in *Drosophila sechellia*. *Genetica* 123: 137–145.
- Zou DJ, Chesler A, Firestein S (2009) How the olfactory bulb got its glomeruli: a just so story? *Nat Rev Neurosci* 10: 611–618.
- Strausfeld NJ, Hildebrand JG (1999) Olfactory systems: common design, uncommon origins? *Curr Opin Neurobiol* 9: 634–639.

68. Benton R (2006) On the ORigin of smell: odorant receptors in insects. *Cell Mol Life Sci* 63: 1579–1585.
69. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–763.
70. Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. *Genome Res* 14: 988–995.
71. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567–580.
72. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783–795.
73. Do CB, Mahabhashyam MS, Brudno M, Batzoglou S (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* 15: 330–340.
74. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.
75. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.
76. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
77. Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456–463.
78. Maddison WP, Maddison DR (2009) Mesquite: A modular system for evolutionary analysis. In: <http://mesquiteproject.org>, editor. 2.6 ed.
79. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23: 127–128.
80. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
81. Collingridge GL, Olsen RW, Peters J, Spedding M (2009) A nomenclature for ligand-gated ion channels. *Neuropharmacology* 56: 2–5.
82. Hahn MW, De Bic T, Stajich JE, Nguyen C, Cristianini N (2005) Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Research* 15: 1153–1160.
83. De Bic T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22: 1269–1271.
84. Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Molecular Biology and Evolution* 21: 36–44.
85. Vieira FG, Sanchez-Gracia A, Rozas J (2007) Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biology* 8.
86. Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 13: 555–556.
87. Brawand D, Wahli W, Kaessmann H (2008) Loss of egg yolk genes in mammals and the origin of lactation and placentation. *PLoS Biol* 6: e63. 10.1371/journal.pbio.0060063.
88. Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N (2008) Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40: 476–483.
89. Lee T, Luo L (1999) Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* 22: 451–461.
90. Cummins SF, Leblanc L, Degnan BM, Nagle GT (2009) Molecular identification of candidate chemoreceptor genes and signal transduction components in the sensory epithelium of *Aplysia*. *J Exp Biol* 212: 2037–2044.
91. Livingstone CD, Barton GJ (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* 9: 745–756.
92. Medina M, Collins TM, Walsh PJ (2001) mtDNA ribosomal gene phylogeny of sea hares in the genus *Aplysia* (Gastropoda, Opisthobranchia, Anaspidea): implications for comparative neurobiology. *Syst Biol* 50: 676–688.
93. Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, et al. (2008) Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* 179: 1601–1655.
94. Kokoeva MV, Storch KF, Klein C, Oesterhelt D (2002) A novel mode of sensory transduction in archaea: binding protein-mediated chemotaxis towards osmoprotectants and amino acids. *Embo J* 21: 2312–2322.

IRs in insect genomes

Summaries of results

As an extension of our research article about the evolution of IRs, our lab was contacted to participate in the annotation of IRs in two other insect species: the Argentine ant *Linepithema humile* (Smith et al., 2011a), the red harvester ant *Pogonomyrmex barbatus* (Smith et al., 2011b). We used our IR annotation pipeline to identify these genes in these species, classified them and examined their orthology with IRs from other species. We found that IR repertoires are well conserved in these two ant species. However, although they share several of the conserved insect olfactory IRs, ants have more IRs than other Hymenoptera such as bees or wasps, and thus seem to have developed their own IR subfamily. A deeper investigation into this expansion would be likely to provide interesting insights into ant evolution and speciation.

My contribution to this work

For these two species, I annotated the IRs, confirmed their genomic location and manually annotated missing elements. I tested their orthology with other insect IRs and created phylogenetic trees comparing IRs across the species in question and other neighbouring or relevant species.

In the *Linepithema humile* paper, I provided Figure 3B and table S12. In the *Pogonomyrmex barbatus* paper, I provided Figure S17 and the text about IRs in the chapter 14 of the supplementary data.

Draft genome of the red harvester ant *Pogonomyrmex barbatus*

Chris R. Smith^{a,1}, Christopher D. Smith^{b,1}, Hugh M. Robertson^c, Martin Helmkamp^d, Aleksey Zimin^e, Mark Yandell^f, Carson Holt^f, Hao Hu^f, Ehab Abouheif^g, Richard Benton^h, Elizabeth Cash^d, Vincent Croset^h, Cameron R. Currie^{ij}, Eran Elhaik^k, Christine G. Elsik^l, Marie-Julie Favé^g, Vilaiwan Fernandes^g, Joshua D. Gibson^d, Dan Graur^m, Wulfila Gronenbergⁿ, Kirk J. Grubbs^j, Darren E. Hagen^l, Ana Sofia Ibarraⁿ, Brian R. Johnson^o, Reed M. Johnson^p, Abderrahman Khila^g, Jay W. Kim^b, Kaitlyn A. Mathis^o, Monica C. Munoz-Torres^l, Marguerite C. Murphy^q, Julie A. Mustard^d, Rin Nakamura^b, Oliver Niehuis^r, Surabhi Nigam^q, Rick P. Overson^d, Jennifer E. Placek², Rajendran Rajakumar^g, Justin T. Reese^l, Garret Suen^{ij}, Shu Tao^l, Candice W. Torres^o, Neil D. Tsutsui^o, Lumi Viljakainen^s, Florian Wolschin^{d,t}, and Jürgen Gadau^{d,2}

^aDepartment of Biology, Earlham College, Richmond, IN 47374; ^bDepartment of Biology, San Francisco State University, San Francisco, CA 94132; ^cDepartment of Entomology, University of Illinois Urbana-Champaign, Urbana, IL 61801; ^dSchool of Life Sciences, Arizona State University, Tempe, AZ 85287; ^eInstitute for Physical Science and Technology, University of Maryland, College Park, MD 20742; ^fDepartment of Human Genetics, University of Utah, Salt Lake City, UT 84112; ^gDepartment of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada; ^hCenter for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; ⁱDOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53706; ^jDepartment of Bacteriology, University of Wisconsin, Madison, WI 53706; ^kThe Johns Hopkins University School of Medicine, Baltimore, MD 21205; ^lDepartment of Biology, Georgetown University, Washington, DC 20057; ^mDepartment of Biology and Biochemistry, University of Houston, Houston, TX 77204; ⁿDepartment of Neuroscience, University of Arizona, Tucson, AZ 85721; ^oDepartment of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720; ^pDepartment of Entomology, University of Nebraska, Lincoln, NE 68583; ^qDepartment of Computer Science, San Francisco State University, San Francisco, CA 94132; ^rCenter for Molecular Biodiversity, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany; ^sDepartment of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853; and ^tDepartment of Biotechnology, Chemistry, and Food Science, Norwegian University of Life Sciences, 1492 Ås, Norway

Edited* by Gene E. Robinson, University of Illinois at Urbana-Champaign, Urbana, IL, and approved November 9, 2010 (received for review June 5, 2010)

We report the draft genome sequence of the red harvester ant, *Pogonomyrmex barbatus*. The genome was sequenced using 454 pyrosequencing, and the current assembly and annotation were completed in less than 1 y. Analyses of conserved gene groups (more than 1,200 manually annotated genes to date) suggest a high-quality assembly and annotation comparable to recently sequenced insect genomes using Sanger sequencing. The red harvester ant is a model for studying reproductive division of labor, phenotypic plasticity, and sociogenomics. Although the genome of *P. barbatus* is similar to other sequenced hymenopterans (*Apis mellifera* and *Nasonia vitripennis*) in GC content and compositional organization, and possesses a complete CpG methylation toolkit, its predicted genomic CpG content differs markedly from the other hymenopterans. Gene networks involved in generating key differences between the queen and worker castes (e.g., wings and ovaries) show signatures of increased methylation and suggest that ants and bees may have independently co-opted the same gene regulatory mechanisms for reproductive division of labor. Gene family expansions (e.g., 344 functional odorant receptors) and pseudogene accumulation in chemoreception and P450 genes compared with *A. mellifera* and *N. vitripennis* are consistent with major life-history changes during the adaptive radiation of *Pogonomyrmex* spp., perhaps in parallel with the development of the North American deserts.

chemoreceptor | de novo genome | eusociality | genomic evolution | social insect

The formation of higher-level organization from independently functioning elements has resulted in some of the most significant transitions in biological evolution (1). These include the transition from prokaryotes to eukaryotes and from uni- to multicellular organisms, as well as the formation of complex animal societies with sophisticated division of labor among individuals. In eusocial insects such as ants, distinct morphological castes specialize in either reproduction or labor (2). Currently, very little is known of the genetic basis of caste and reproductive division of labor in these societies, where individuals follow different developmental trajectories, much like distinct cell lines in an organism (3). The resulting phenotypes, queens and workers, can differ greatly in morphology, physiology, and behavior, as well as in order of magnitude differences in life span and reproductive po-

tential (2). Ants, of all social insects, arguably exhibit the highest diversity in social complexity, such as queen number, mating frequency, and the degree of complexity of division of labor (2), and most social traits have independent origins within the ants, making them well suited to comparative genomic analyses.

The sequencing of the honey bee (*Apis mellifera*) genome marked a milestone in sociogenomics (4, 5), facilitating research on the evolution and maintenance of sociality from its molecular building blocks. Since then, genomes of three closely related species of solitary parasitic hymenopterans, *Nasonia* spp., were published and similarities and differences were extensively discussed in the context of the evolution of eusociality (6). However, *A. mellifera* represents only 1 of at least 10 independent evolutionary origins of eusociality within the order Hymenoptera (7–11), and thus it remains unclear whether differences between the honey bee and *Nasonia* spp. truly reflect differences inherent in sociality. With at least six ant genomes on the horizon (12), among other solitary and social insects, sociogenomic comparisons are likely to yield exciting insights into the common molecular basis for the social lifestyle. Ant genomics will also allow us to gain a better understanding of variation in social organization, of elaborate variations of physical and behavioral divisions of labor, of invasion biology, and of the convergent evolution of life histories and diets. It also remains a major question whether there are many evolutionary routes to eusociality, especially at the molecular level,

Author contributions: C.R.S., C.D.S., and J.G. designed research; C.R.S., C.D.S., H.M.R., M.H., A.Z., M.Y., C.H., H.H., M.-J.F., W.G., F.W., and J.G. performed research; C.R.S., C.D.S., H.M.R., M.H., A.Z., M.Y., C.H., H.H., E.A., R.B., E.C., V.C., C.R.C., E.E., C.G.E., V.F., J.D.G., D.G., W.G., K.J.G., D.E.H., A.S.I.V., B.R.J., R.M.J., A.K., J.W.K., K.A.M., M.C.M.-T., M.C.M., J.A.M., R.N., O.N., S.N., R.P.O., J.E.P., R.R., J.T.R., G.S., S.T., C.W.T., N.D.T., L.V., F.W., and J.G. analyzed data; and C.R.S., M.H., and J.G. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the Hymenoptera Genome Database: <http://HymenopteraGenome.org/pogonomyrmex> (NCBI Genome Project #45803, Assembly Project ID 45797, Transcriptome Project ID 46577).

¹C.R.S. and C.D.S. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: jgadau@asu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1007901108/-DCSupplemental.

or whether we can extract generalities and rules for the molecular evolution of eusociality (3, 4, 13). Although it is likely that much variation in social structure is due to changes in the regulation of conserved pathways, it is undetermined what, if any, role novel genes or pathways have played in the solitary-to-social transition and diversification of social phenotypes (14).

The genus *Pogonomyrmex* contains species that vary greatly in social organization (15), is among the best studied of ant genera (16, 17), is sister to almost all other genera in the diverse subfamily Myrmicinae (8, 11), and contains species of major ecological importance as granivores in both North and South America (18, 19). Colonies can contain over 10,000 workers and a single multiply mated queen that may live for decades. Some *Pogonomyrmex barbatus* populations have a unique system of genetic queen-worker caste determination (Fig. 1) where individuals are essentially hard-wired to develop as either queens or workers, a contrast to environmentally determined diphenism (20–24) (*SI Appendix, Chapter 1*). As a consequence, individuals can be genotyped using genetic markers to determine their caste even before caste differentiation. This unique system of caste determination provides a means of studying the genes and regulatory networks used in caste determination.

Results and Discussion

Genome coverage is 10.5–12 \times on the basis of the estimates of genome size for *Pogonomyrmex* ants as 250–284 Mb (25). The assembly consists of 4,646 scaffolds (mean contig/scaffold: 7.22) spanning 235 Mb (~88%) of the genome that harbor 220 Mb (~83%) of DNA sequence (15 Mb of which are gaps within scaffolds). The N50 scaffold size of the assembly is 793 kb, and the largest scaffold is 3.8 Mb in length; the N50 contig size is 11.6 kb. The transcriptome assembly yielded 7,400 isogroups with a N50 contig size of 1.3 kb.

The MAKER annotation pipeline predicted 16,331 genes and 16,404 transcripts. InterProScan (26) identified additional genes from the in silico prediction programs, which were added to the MAKER predicted genes. The final official gene set, OGS1.1, which was used for computational analyses, consisted of 17,177 genes encoding 17,250 transcripts. Of these, 7,958 (>46%) had complete or partial EST support from the *P. barbatus* transcriptome assembly. The results of the assembly and annotation of the *P. barbatus* genome are well within the range of other insect genomes (Table 1).

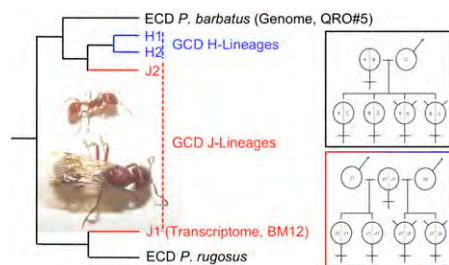


Fig. 1. A pictorial description of the phylogenetic position of the samples used for the genome and transcriptome sequencing, with each put in the context of environmental and genetic caste determination (for a more complete phylogenetic tree, see *SI Appendix, Chapter 1*). The dependent lineages (H1/H2 or J1/J2) obligately co-occur because hybridization between them is necessary to produce workers, although within either J or H, the constituent lineages are reproductively isolated because interlineage hybrids cannot become queens (red/blue box). In the boxes to the right, workers are represented by “horned” female symbols. In all *P. barbatus*, the queen mates multiply; polyandry in genetic caste determining (GCD) colonies is obligate to produce both female castes (queens originate from intralineage matings and workers from interlineage matings). In environmental caste determination (ECD), alleles from any father have an equal chance to be in queens or workers (black box). Photo of gyne and worker *P. barbatus* by C. R. Smith.

More than 1,200 genes have been manually annotated to improve models generated by MAKER (*SI Appendix, Chapter 2*) and were used in gene family-centered analyses (see discussion below and *SI Appendix, Chapters 3, 6–8, 14, and 16–29*). There are two fundamentally different reasons for our choice of gene families: One set comprises highly conserved gene families for quality assessment (e.g., sequencing error, genome completeness), whereas the second set is based on biologically interesting functional groups associated with the evolution and regulation of social behavior or adaptations of *P. barbatus* to a desert seed-harvesting lifestyle.

Quality of Genome Assembly. The core eukaryotic gene-mapping approach (CEGMA) (27) provides a method to rapidly assess genome completeness because it comprises a set of highly conserved, single-copy genes, present in all eukaryotes. In *P. barbatus*, 245 of the 248 (99%) CEGMA genes were found, and 229 of the 248 genes were complete (92%). Cytoplasmic ribosomal protein genes are another highly conserved set of genes that are widely distributed across the physical genome in animals (28, 29). A full complement of 79 proteins was found within the *P. barbatus* genome encoded by 86 genes (*SI Appendix, Chapter 6*). Because ribosomal proteins are highly conserved, their manual annotation also provided an estimate of sequencing errors, such as frameshift-inducing homopolymers (a potential problem inherent to pyrosequencing) (30). Six erroneous frameshifts were found in ribosomal protein genes (only one homopolymer); extrapolating from the number of nucleotides encoding the ribosomal genes suggests that 1 in 7,200 coding nucleotide positions (0.014%) may be affected by frameshifts. Analyses of other highly conserved gene families, including the oxidative phosphorylation (31) pathway and the *Hox* gene cluster (32, 33), also suggest high coverage and good genome assembly (*SI Appendix, Chapters 7 and 8*). Interestingly, the mitochondrial genome did not auto-assemble into scaffolds greater than 2 kb, but 71% of the mitochondrial genome could be manually assembled with the longest contig containing 5,835 bp (*SI Appendix, Chapter 9, Dataset S1*). The largest missing fragment of the mitochondrial genome is typically very high in AT content (96% in *A. mellifera ligustica*) (34) and may not have been sequenced due to PCR biases.

In silico-predicted gene models gain significant support through EST sequences. Another way to confirm predicted gene models is a proteomics approach, which has the additional benefit that it demonstrates that a gene is not only transcribed but also translated. A proteomic analysis of the poison gland and antennae confirmed 165 gene and protein models with at least two peptides (*SI Appendix, Chapter 10*). It also resulted in the identification of proteins likely associated with nest defense (poison gland) and chemoreception (antenna).

Chromosomal coverage in the current draft assembly was assessed by the identification of telomeres. Most insects outside of the Diptera have telomeres consisting of TTAGG repeats. On the basis of the karyotype data ($n = 16$), we expected 32 telomeres in *P. barbatus* (35). We searched the assembled genome and mate pair reads for TTAGG repeats and extended these where possible (6). In total, 27 of the expected 32 telomeres (88%) were found (*SI Appendix, Chapter 11*). These telomeres are even simpler than those of *A. mellifera* (36). Whereas most other insect telomeres commonly include retrotransposon insertions, these seem to be absent from the telomeres of *P. barbatus*.

Genome-Wide Analyses. The mean GC content of the *P. barbatus* genome is 36.5% and the mean ratio of observed-to-expected CpG [CpG(o/e)] is 1.57, both of which are within the ranges reported for other Hymenoptera (5, 6). We define compositional domains as the sequence stretches of variable lengths that differ widely in their GC compositions. A comparison of GC compositional-domain lengths among insects shows that *P. barbatus* and *A. mellifera* have similar compositional domain-length distributions (*SI Appendix, Chapter 4*). Among the compared insect genomes, the hyme-

Table 1. Comparison of metrics for recently sequenced insect genomes

Species	Order/name	Fold coverage	N50 scaffold (kb)	No. of genes	Gene set	Source
<i>Pogonomyrmex barbatus</i>	Hymenoptera (red harvester ant)	12	793	17,177	OGS1.1	This study
<i>Nasonia vitripennis</i>	Hymenoptera (jewel wasp)	6.8	709	18,822	OGS1.2	(6)
<i>Apis mellifera</i>	Hymenoptera (honey bee)	7.5	362	10,156/21,001	OGS1/OGS2	(5)
<i>Acyrtosiphon pisum</i>	Sternorrhyncha (pea aphid)	6.2	88.5	34,604	OGS1	(37)
<i>Tribolium castaneum</i>	Coleoptera (red flower beetle)	7.3	990	16,404	Consensus set	(38)

noptera have the smallest proportion (0.1–0.5%) of long compositional domains (>100 kb) as well as the widest range in GC compositional domains. Similar to the other sequenced hymenopteran genomes, but in contrast to other insect orders, genes in *P. barbatus* occur in the more GC-poor regions of the genome. Although the mean CpG(o/e) values of hymenopteran genomes are among the highest observed, species-specific patterns of CpG (o/e) within each genome are not consistent between the hymenopterans studied (Fig. 2). The distribution of CpG(o/e) in *P. barbatus* exons is similar to that in insects without CpG methylation (although with greater variance) (39) and suggests little germline methylation despite the presence of a complete methylation toolkit (see below and *SI Appendix, Chapter 24*). We used an indirect method [single nucleotide polymorphisms (SNP) frequency: CpG – TpG] and a direct method [methylation-sensitive amplified fragment length polymorphism (AFLP) assay; *SI Appendix, Chapter 4*] to determine the presence and frequency of active CpG methylation in *P. barbatus*. We found that CpG/TpG (and vice versa) SNPs constitute 84% of all CpG-to-NpG polymorphisms. This is an indirect measure of CpG methylation because it has been shown that a methylated cytosine in a CpG has a higher probability to mutate into thymine (*SI Appendix, Chapter 30*). The more direct measure of CpG methylation comes from an AFLP analysis that used methylation-sensitive and -insensitive restriction enzymes. In a comparison of 209 individuals from every female and developmental caste, 33% of all AFLP fragments showed a signature of methylation (*SI Appendix, Chapter 4*). These findings suggest a role of DNA methylation in genome regulation, but additional data are necessary to confirm these predictions and discern the biological role of DNA methylation in *P. barbatus*.

Gene ontology analyses detected significant enrichments in genes associated with sensory perception of smell, cognition, and neurological processes (*SI Appendix, Chapter 5*). These enrichments may reflect the heavy reliance on chemical communication in ants. Consistent with this and detailed analyses of chemosensory and cytochrome P450 gene families (see below), a gene orthology analysis including *Drosophila melanogaster*, *A. mellifera*, and *Nasonia vitripennis* found expansions of genes involved in responses to chemical stimuli and electron transport. The orthology analysis also found a small fraction of genes (3.2% of those in the analysis) common to both social insects studied (*SI Appendix, Chapter 5*); these genes may be important in processes related to the evolution or maintenance of sociality.

Repetitive DNA. Previous results for the *A. mellifera* (5) and *N. vitripennis* (6) genomes illustrate two extreme cases of genomic repeat composition for Hymenoptera: *A. mellifera* is devoid of all except a few *mariner* (40) and rDNA-specific R2 (41) transposable elements whereas *N. vitripennis* has an unusual abundance of repetitive DNA (6). The *P. barbatus* genome assembly contains 18.6 Mb (8% of genome) of interspersed elements (*SI Appendix, Chapter 12*). A total of 9,324 retroviral element fragments and 13,068 DNA transposons were identified; however, the majority of interspersed elements (55,373, 8.8 Mb, 3.75% of genome) could not be classified into a specific transposable element family. Gypsy/DIR1 and L2/CR1/Rex elements were the most abundant transposable elements; however, we discovered most families of

known insect retrotransposable elements. Nearly 1% (269 loci/1 Mb) of the scaffolded genome is microsatellite DNA (*SI Appendix, Chapter 13*), greater than in most insects (42), which are valuable markers for mapping and population genetic studies.

Chemoreceptor Gene Family Expansions. One special focus of the manual annotation was the proteins involved in chemoreception, which plays an important role in colony communication, a cor-

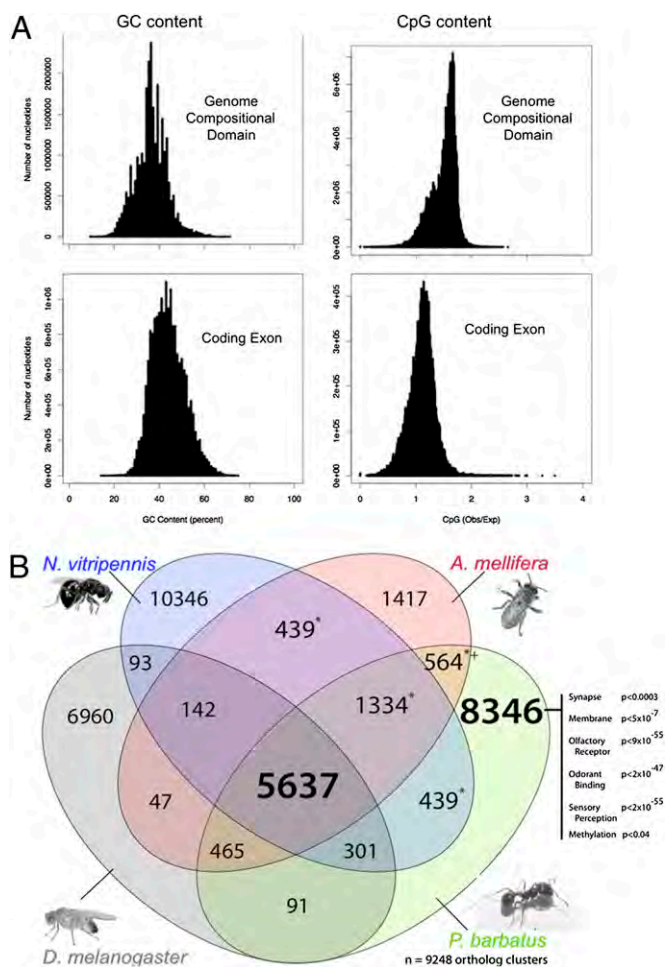


Fig. 2. Genome-wide analyses of nucleotide and relative gene content. (A) Synopsis of GC and CpG(o/e) content of the *P. barbatus* genome. (Upper panels) Comparison of genome regions with the same GC composition. (Lower panels) Comparison of the same features for exons. These distributions are similar to those found in other hymenopterans, except that *P. barbatus* shows no evidence of bimodality in CpG(o/e) for either exons (like *A. mellifera*) or introns (like *N. vitripennis*) (for comparisons, see *SI Appendix, Chapter 4*). (B) A Venn diagram displaying overlap in orthologous genes in three hymenopteran and one dipteran insect (for a detailed description of the method, see *SI Appendix, Chapter 5*). A subset of gene ontology terms significantly enriched in *P. barbatus* are displayed at the right. (*) Hymenoptera-specific genes; (**) social Hymenoptera-specific genes.

nerstone of social living. Below we report insights derived from four gene families involved in chemoreception: the ionotropic receptors (IRs), gustatory receptors (Gr), odorant receptors (Ors), and cytochrome P450s.

The IR family in *P. barbatus* consists of 24 genes, compared with 10 in *A. mellifera* and 10 in *N. vitripennis* (43). Phylogenetic analysis and sequence comparison of IRs identified putative orthologs of conserved IRs that are present in other insect genomes and that are expressed in insect antennae (e.g., IR25a, IR8a, IR93a, IR76b) (44), but a number of ant-specific divergent IRs display no obvious orthology to other hymenopteran or insect receptors (*SI Appendix, Chapter 14*). Some of these IRs may fulfill contact chemosensory functions by analogy to the gustatory neuron expression of species-specific IRs in *D. melanogaster* (43).

The *P. barbatus* Gr family contains 73 genes compared with just 11 in *A. mellifera* and 58 in *N. vitripennis*. Phylogenetic analysis of the Gr proteins (*SI Appendix, Chapter 14*) supports several conclusions about the evolution of this gene family. *A. mellifera* has lost multiple Gr lineages and failed to expand any of them (45, 46), but gene losses are not restricted to *A. mellifera*, with some occurring in *N. vitripennis* and/or *P. barbatus*. The existence of at least 18 Gr lineages is inferred, with *A. mellifera* having lost function in 10 of them, *P. barbatus* in 4, and *N. vitripennis* in 5. *P. barbatus* has expanded two gene lineages independently of the two expansions seen in *N. vitripennis*. Expansion A is considered to be orthologous to the NvGr48-50 gene lineage and a large set of ≈ 50 highly degraded pseudogenes in *A. mellifera* (represented by AmGrX-Z), and expansion B is somewhat younger. We hypothesize that these are bitter taste receptors that lost function in *A. mellifera* at the time at which they transitioned to nectar feeding, ≈ 100 Mya (47). Bitter taste perception may be essential for *P. barbatus* to avoid unpalatable seeds (e.g., plant secondary compounds).

The Or family also appears to be considerably expanded in *P. barbatus*, with 344 apparently functional genes among a total of 399 genes (the largest total known for any insect) compared with a total of 166 in *A. mellifera* and 225 in *N. vitripennis* (Dataset S2). We counted 365 ± 10 and 345 ± 10 glomeruli in five queens and five workers, respectively (*SI Appendix, Chapter 15*), supporting an $\approx 1:1$ relationship of Or genes to glomeruli resulting from convergence of the axons of all neurons expressing a particular Or on one glomerulus (48, 49). A particularly large expansion of a nine-exon gene subfamily to 169 genes suggests that these genes might comprise the cuticular hydrocarbon receptors (*SI Appendix, Chapter 14*). Cuticular hydrocarbons have gained many novel functions important in the context of social behavior, such as colony recognition and queen signaling (50, 51).

P. barbatus has 72 genes in the cytochrome P450 superfamily, compared with 46 in *A. mellifera* and 92 in *N. vitripennis* (5, 6). P450 subfamilies involved in detoxification of xenobiotics show some expansion, whereas those implicated in pheromone metabolism are enigmatically less expanded (*SI Appendix, Chapter 16*).

Evolutionary Rate and Pseudogene Accumulation. An evolutionary rate analysis based on amino acid substitutions of the three hymenopteran species with a genome sequence, with *D. melanogaster* as an outgroup, showed that a significant part of the *P. barbatus* genome (4,774 orthologous genes conserved over approximately 350 million y) evolves at a similar rate as the *A. mellifera* genome, and the *A. mellifera* and *P. barbatus* genomes show slightly higher substitution rates than the *N. vitripennis* genome (Fig. 3 and *SI Appendix, Chapter 31*). This analysis suggests that the slow evolutionary rate reported for *A. mellifera* may not be associated with sociality, but rather is specific to the Hymenoptera.

A notable feature of *P. barbatus* chemosensory and P450 genes is that the pseudogenes commonly have multiple major mutations suggesting that they are mostly “middle-aged” pseudogenes. Normally a range of pseudogene ages can be inferred in the chemoreceptor gene families from young pseudogenes with

single mutations to gene fragments. We estimated the relative ages of the pseudogenes in Ors, Grs, and cytochrome P450s in *P. barbatus*, *A. mellifera*, and *N. vitripennis* by counting the number of obvious pseudogene-causing (“pseudogenizing”) mutations per gene (stop codons, intron boundary mutations, small frame-shift insertions or deletions, or large insertions or deletions). As shown in Fig. 3, there is a contingent of considerably older pseudogenes in these gene families in *P. barbatus*. The pattern in *P. barbatus* is in contrast to *A. mellifera* and *N. vitripennis*, which have a greater number of young pseudogenes. We hypothesize that the ant lineages that gave rise to *P. barbatus* experienced a major change in chemical ecology ≈ 10 – 30 Mya, possibly as a consequence of the increase in elevation of the Sierras and Andes to their present height (52, 53). These western mountain ranges created rain shadows on their eastern sides and spawned the great American deserts. The North American members of the genus *Pogonomyrmex* underwent a significant radiation adapting to these new habitats (16), so the gene expansions in the chemoreceptors and P450s might be adaptations to novel seeds and plant families and their associated toxic components and chemical signatures. Accumulated pseudogenes may therefore reflect a shift toward a more specialized diet concurrent with the adaptive radiation of *Pogonomyrmex* spp. (54).

Innate Immunity Genes. Social insects live in dense groups with high connectivity, putting them at increased risk for disease outbreaks, but they also have social immunity to minimize the introduction and spread of pathogens (55, 56). Very efficient social defenses (e.g., hygienic behaviors) or novel immune pathways were hypothesized to explain the presence of few (roughly half) innate immunity genes in *A. mellifera* compared with *D. melanogaster* (and more recently in the red flour beetle, *Tribolium castaneum*) (5, 38). However, the more recently sequenced genomes of *N. vitripennis* (6) and *Acyrtosiphon pisum* (pea aphid) (37) also have “depauperate” complements of immune genes relative to flies and beetles, which suggests that the gene complement of flies and beetles might be a derived condition within insects. Indeed, the number of innate immune genes in *P. barbatus* is more similar to the other hymenopterans (*SI Appendix, Chapter 17*). Although all of the major signaling pathways are present in *P. barbatus* (IMD, Toll, Jak/STAT, and JKN), only a few recognition proteins were identified, which suggests either a highly focused immune system or an alternative unknown pathogen recognition system. Interestingly, we found expansions of antimicrobial peptides relative to *A. mellifera*. These expansions may correspond to a transition to living within the soil and an increased exposure to bacterial and fungal pathogens.

Developmental Networks and Polyphenism. The production of alternative phenotypes during development may occur through the regulation of several key nodes in specific networks during development (57–59). In ant colonies, queens and workers fill divergent adaptive roles—dispersal and reproduction vs. colony maintenance—and their functional differences are reflected in differences in morphology, physiology, and behavior, such as in wings and ovaries. *P. barbatus* workers are completely devoid of wings at the adult stage and have ovaries a fraction of the size of the queen’s. In analogy to honey bees (60), we hypothesized that CpG DNA methylation may play a role in the differential regulation of genes in the wing and reproductive development networks of workers and queens. This hypothesis was computationally evaluated by examining the CpG dinucleotide content (39) of wing and reproductive developmental pathway genes relative to the genome (*SI Appendix, Chapter 18*). These developmental networks contain significantly fewer CpGs than random genes, suggesting that they are more methylated than most genes because methylated cytosines are more prone to deamination (6, 39, 61). These results are in contrast to data on *A. mellifera*, where housekeeping

ACKNOWLEDGMENTS. A very special thanks to S. Pratt for comments on the manuscript. We are thankful to the Earlham College Evolutionary Genomics class, which annotated genes and did preliminary analyses. R. Jones, B. Mott, and T. Holbrook collected specimens. We are grate-

ful for allocated computer time from the Center for High Performance Computing at the University of Utah. We also thank Mike Wong from the Center for Computing for Life Science at San Francisco State University for assistance with custom scripts and hardware configuration. National Science Foundation Grant IOS-0920732 (to J.G. and C.R.S.) funded the sequencing of the genome, and National Institutes of Health Grant 5R01HG004694 (to M.Y.) funded the MAKER annotation.

1. Maynard Smith J, Szathmáry E (1995) *The Major Transitions in Evolution* (W. H. Freeman/Spektrum, New York).
2. Hölldobler B, Wilson EO (1990) *The Ants* (Belknap Press of Harvard University Press, Cambridge, MA).
3. Smith CR, Toth AL, Suarez AV, Robinson GE (2008) Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* 9:735–748.
4. Robinson GE, Grozinger CM, Whitfield CW (2005) Sociogenomics: Social life in molecular terms. *Nat Rev Genet* 6:257–270.
5. Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931–949.
6. Werren JH, et al.; Nasonia Genome Working Group (2010) Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* 327:343–348.
7. Hines HM, Hunt JH, O'Connor TK, Gillespie JJ, Cameron SA (2007) Multigene phylogeny reveals eusociality evolved twice in vespid wasps. *Proc Natl Acad Sci USA* 104:3295–3299.
8. Brady SG, Schultz TR, Fisher BL, Ward PS (2006) Evaluating alternative hypotheses for the early evolution and diversification of ants. *Proc Natl Acad Sci USA* 103:18172–18177.
9. Brady SG, Sipes S, Pearson A, Danforth BN (2006) Recent and simultaneous origins of eusociality in halictid bees. *Proc Biol Sci* 273:1643–1649.
10. Schwarz MP, Richards MH, Danforth BN (2007) Changing paradigms in insect social evolution: Insights from halictine and allodapine bees. *Annu Rev Entomol* 52:127–150.
11. Moreau CS, Bell CD, Vila R, Archibald SB, Pierce NE (2006) Phylogeny of the ants: Diversification in the age of angiosperms. *Science* 312:101–104.
12. Smith CD, Smith CR, Mueller U, Gadau J (2010) Ant genomics: Strength and diversity in numbers. *Mol Ecol* 19:31–35.
13. Toth AL, Robinson GE (2007) Evo-devo and the evolution of social behavior. *Trends Genet* 23:334–341.
14. Page RE, Jr., Amdam GV (2007) The making of a social insect: Developmental architectures of social design. *Bioessays* 29:334–343.
15. Johnson RA (2000) Seed-harvester ante (Hymenoptera: Formicidae) of North America: An overview of ecology and biogeography. *Sociobiology* 36:89–122.
16. Taber SW (1998) *The World of the Harvester Ants* (Texas A&M University Press, College Station, TX).
17. Gordon DM (1999) *Ants at Work* (The Free Press, New York).
18. Pirk GI, Lopez de Casenave J (2006) Diet and seed removal rates by the harvester ants *Pogonomyrmex rastratus* and *Pogonomyrmex pronotalis* in the central Monte desert, Argentina. *Insectes Soc* 53:119–125.
19. MacMahon JA, Mull JF, Crist TO (2000) Harvester ants (*Pogonomyrmex* spp.): Their community and ecosystem influences. *Annu Rev Ecol Syst* 31:265–291.
20. Anderson KE, Linksvayer TA, Smith CR (2008) The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecol News* 11:119–132.
21. Helms Cahan S, et al. (2002) Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proc Biol Sci* 269:1871–1877.
22. Julian GE, Fewell JH, Gadau J, Johnson RA, Larrabee D (2002) Genetic determination of the queen caste in an ant hybrid zone. *Proc Natl Acad Sci USA* 99:8157–8160.
23. Volny VP, Gordon DM (2002) Genetic basis for queen-worker dimorphism in a social insect. *Proc Natl Acad Sci USA* 99:6108–6111.
24. Cahan SH, et al. (2004) Loss of phenotypic plasticity generates genotype-caste association in harvester ants. *Curr Biol* 14:2277–2282.
25. Tsutsui ND, Suarez AV, Spagna JC, Johnston JS (2008) The evolution of genome size in ants. *BMC Evol Biol* 8:64.
26. Quevillon E, et al. (2005) InterProScan: Protein domains identifier. *Nucleic Acids Res* 33(Web Server issue):W116–W120.
27. Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067.
28. Uechi T, Tanaka T, Kenmochi N (2001) A complete map of the human ribosomal protein genes: Assignment of 80 genes to the cytogenetic map and implications for human disorders. *Genomics* 72:223–230.
29. Marygold SJ, et al. (2007) The ribosomal protein genes and Minute loci of *Drosophila melanogaster*. *Genome Biol* 8:R216.
30. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 8:R143.
31. Saraste M (1999) Oxidative phosphorylation at the fin de siècle. *Science* 283:1488–1493.
32. Hughes CL, Kaufman TC (2002) Hox genes and the evolution of the arthropod body plan. *Evol Dev* 4:459–499.
33. Gellon G, McGinnis W (1998) Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *Bioessays* 20:116–125.
34. Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: Complete sequence and genome organization. *Genetics* 133:97–117.
35. Taber SW, Cokendolpher JC, Francke OF (1988) Karyological study of North-American *Pogonomyrmex* (Hymenoptera, Formicidae). *Insectes Soc* 35:47–60.
36. Robertson HM, Gordon KH (2006) Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera*. *Genome Res* 16:1345–1351.
37. International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 8:e1000313.
38. *Tribolium* Genome Sequencing Consortium et al. (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452:949–955.
39. Elango N, Hunt BG, Goodisman MAD, Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci USA* 106:11206–11211.
40. Robertson HM (1993) The mariner transposable element is widespread in insects. *Nature* 362:241–245.
41. Kojima KK, Fujiwara H (2005) Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* 22:2157–2165.
42. Pannebakker BA, Niehuis O, Hedley A, Gadau J, Shuker DM (2010) The distribution of microsatellites in the *Nasonia* parasitoid wasp genome. *Insect Mol Biol* 19(Suppl 1):91–98.
43. Crosset V, et al. (2010) Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet* 6:e1001064.
44. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB (2009) Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136:149–162.
45. Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: Expansion of the odorant, but not gustatory, receptor family. *Genome Res* 16:1395–1403.
46. Robertson HM, Gadau J, Wanner KW (2010) The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol Biol* 19(Suppl 1):121–136.
47. Poinar GO, Jr., Danforth BN (2006) A fossil bee from Early Cretaceous Burmese amber. *Science* 314:614.
48. Mombaerts P (1999) Molecular biology of odorant receptors in vertebrates. *Annu Rev Neurosci* 22:487–509.
49. Gao Q, Yuan B, Chess A (2000) Convergent projections of *Drosophila* olfactory neurons to specific glomeruli in the antennal lobe. *Nat Neurosci* 3:780–785.
50. Endler A, et al. (2004) Surface hydrocarbons of queen eggs regulate worker reproduction in a social insect. *Proc Natl Acad Sci USA* 101:2945–2950.
51. Hefetz A (2007) The evolution of hydrocarbon pheromone parsimony in ants (Hymenoptera: Formicidae): Interplay of colony odor uniformity and odor idiosyncrasy. *Myrmecol News* 10:59–68.
52. Poulsen CJ, Ehlers TA, Insel N (2010) Onset of convective rainfall during gradual late Miocene rise of the central Andes. *Science* 328:490–493.
53. Cassel EJ, Graham AA, Chamberlain CP (2009) Cenozoic tectonic and topographic evolution of the northern Sierra Nevada, California, through stable isotope paleoaltimetry in volcanic glass. *Geology* 37:547–550.
54. McBride CS (2007) Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci USA* 104:4996–5001.
55. Walker TN, Hughes WO (2009) Adaptive social immunity in leaf-cutting ants. *Biol Lett* 5:446–448.
56. Fefferman NH, Traniello JFA (2008) Social insects as models in epidemiology: Establishing the foundation for an interdisciplinary approach to disease and sociality. *Insect Sociology*, eds Gadau J, Fewell J (Harvard University Press, Cambridge, MA), pp 545–571.
57. Davidsein EH (2006) The sea urchin genome: Where will it lead us? *Science* 314:939–940.
58. Abouheif E, Wray GA (2002) Evolution of the gene network underlying wing polyphenism in ants. *Science* 297:249–252.
59. Khila A, Abouheif E (2008) Reproductive constraint is a developmental mechanism that maintains social harmony in advanced ant societies. *Proc Natl Acad Sci USA* 105:17884–17889.
60. Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827–1830.
61. Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R (2009) Epigenetic regulation of the honey bee transcriptome: Unravelling the nature of methylated genes. *BMC Genomics* 10:472.
62. Lu HL, Vinson SB, Pietrantonio PV (2009) Oocyte membrane localization of vitellogenin receptor coincides with queen flying age, and receptor silencing by RNAi disrupts egg formation in fire ant virgin queens. *FEBS J* 276:3110–3123.
63. Schwander T, Cahan SH, Keller L (2007) Characterization and distribution of *Pogonomyrmex* harvester ant lineages with genetic caste determination. *Mol Ecol* 16:367–387.
64. Anderson RC (1945) A study of the factors affecting fertility of lozenge females of *Drosophila melanogaster*. *Genetics* 30:280–296.
65. Perrimon N, Mohler D, Engstrom L, Mahowald AP (1986) X-linked female-sterile loci in *Drosophila melanogaster*. *Genetics* 113:695–712.
66. Bloch Qazi MC, Heifetz Y, Wolfner MF (2003) The developments between gametogenesis and fertilization: Ovation and female sperm storage in *Drosophila melanogaster*. *Dev Biol* 256:195–211.
67. Khila A, Abouheif E (2010) Evaluating the role of reproductive constraints in ant social evolution. *Philos Trans R Soc Lond B Biol Sci* 365:617–630.
68. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24:2818–2824.
69. Cantarel BL, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18:188–196.

Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*)

Christopher D. Smith^{a,1}, Aleksey Zimin^b, Carson Holt^c, Ehab Abouheif^d, Richard Benton^e, Elizabeth Cash^f, Vincent Croset^e, Cameron R. Currie^{g,h}, Eran Elhaikⁱ, Christine G. Elsik^j, Marie-Julie Fave^d, Vilaiwan Fernandes^d, Jürgen Gadau^f, Joshua D. Gibson^f, Dan Graur^k, Kirk J. Grubbs^f, Darren E. Hagen^l, Martin Helmkampf^f, Jo-Anne Holley^l, Hao Hu^c, Ana Sofia Ibarra-Viniegra^d, Brian R. Johnson^m, Reed M. Johnson^l, Abderrahman Khila^d, Jay W. Kim^a, Joseph Laird^l, Kaitlyn A. Mathisⁿ, Joseph A. Moeller^{g,h}, Monica C. Muñoz-Torres^j, Marguerite C. Murphyⁿ, Rin Nakamura^a, Surabhi Nigamⁿ, Rick P. Overson^f, Jennifer E. Placek^a, Rajendhran Rajakumar^d, Justin T. Reese^l, Hugh M. Robertson^o, Chris R. Smith^p, Andrew V. Suarez^o, Garret Suen^{g,h}, Elissa L. Suhr^l, Shu Tao^l, Candice W. Torres^m, Ellen van Wilgenburg^m, Lumi Viljakainen^q, Kimberly K. O. Walden^l, Alexander L. Wild^l, Mark Yandell^c, James A. Yorke^r, and Neil D. Tsutsui^{m,1}

Departments of ^aBiology and ^bComputer Science, San Francisco State University, San Francisco, CA 94132; ^cInstitute for Physical Science and Technology and ^dDepartment of Mathematics, University of Maryland, College Park, MD 20742; ^eDepartment of Human Genetics, Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112; ^fDepartment of Biology, McGill University, Montreal, QC, Canada H3A 1B1; ^gCenter for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland; ^hSchool of Life Sciences, Arizona State University, Tempe, AZ 85287; ⁱDepartment of Bacteriology and ^jDepartment of Energy Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53706; ^kThe Johns Hopkins University School of Medicine, Baltimore, MD 21205; ^lDepartment of Biology, Georgetown University, Washington, DC 20057; ^mDepartment of Biology and Biochemistry, University of Houston, Houston, TX 77204; ⁿDepartment of Entomology and ^oInstitute of Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ^pDepartment of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720-3114; ^qDepartment of Biology, Earlham College, Richmond, IN 47374; and ^rDepartment of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853

Edited* by Gene E. Robinson, University of Illinois, Urbana, IL, and approved January 11, 2011 (received for review June 17, 2010)

Ants are some of the most abundant and familiar animals on Earth, and they play vital roles in most terrestrial ecosystems. Although all ants are eusocial, and display a variety of complex and fascinating behaviors, few genomic resources exist for them. Here, we report the draft genome sequence of a particularly widespread and well-studied species, the invasive Argentine ant (*Linepithema humile*), which was accomplished using a combination of 454 (Roche) and Illumina sequencing and community-based funding rather than federal grant support. Manual annotation of >1,000 genes from a variety of different gene families and functional classes reveals unique features of the Argentine ant's biology, as well as similarities to *Apis mellifera* and *Nasonia vitripennis*. Distinctive features of the Argentine ant genome include remarkable expansions of gustatory (116 genes) and odorant receptors (367 genes), an abundance of cytochrome P450 genes (>110), lineage-specific expansions of yellow/major royal jelly proteins and desaturases, and complete CpG DNA methylation and RNAi toolkits. The Argentine ant genome contains fewer immune genes than *Drosophila* and *Tribolium*, which may reflect the prominent role played by behavioral and chemical suppression of pathogens. Analysis of the ratio of observed to expected CpG nucleotides for genes in the reproductive development and apoptosis pathways suggests higher levels of methylation than in the genome overall. The resources provided by this genome sequence will offer an abundance of tools for researchers seeking to illuminate the fascinating biology of this emerging model organism.

Hymenoptera | invasive species | transcriptome | chemoreception | sociality

Ants are pivotal players in the Earth's terrestrial ecosystems. They include >14,000 described species, comprise about half of all insect biomass in the tropics, are the most important predators of insects and other arthropods, and turn and aerate more soil than earthworms (1). The diversity of lifestyles displayed by ants is equally impressive, including minute, acorn-dwelling species (and the ants that parasitize them), fungus-growing leafcutters with complex division of labor, trap-jaw ants that exhibit the fastest animal movements (2), slave-making ants that kidnap and enslave the young of other ants, and the countless teeming multitudes of army ants that prowl the leaf litter and subterranean habitats (1).

Some ant species have been introduced to new geographic ranges by human activities, and a few have emerged as damaging

and destructive invasive species (3). The Argentine ant (*Linepithema humile*) is one of the most widely distributed of these invaders and is established in nearly every Mediterranean-type climate in the world (4). Introduced Argentine ants form enormous "supercolonies," often across hundreds or thousand of kilometers (5), and workers from the largest supercolonies on different continents even accept each other as cononymates (6). At a local scale, the absence of aggression among workers within supercolonies allows them to direct resources toward colony growth (7) and attain high population densities. These introduced populations are then able to outcompete and eliminate native ants, which imperils plants and animals that normally interact with the native ants (5). In contrast, Argentine ants in their native South American range display high levels of aggression and genetic structure among colonies across substantially smaller spatial scales (8, 9). Argentine ants are also significant pests and thus are targets of heavy insecticidal control, leading to environmental contamination and harm to nontarget organisms (10).

Despite their importance and prominence, there are essentially no genomic resources for Argentine ants. Here we report the transcriptome (GenBank accession no. 46575) and de novo ge-

Author contributions: C.D.S. and N.D.T. designed research; C.D.S., A.Z., C.H., E.A., R.B., E.C., V.C., E.E., C.G.E., M.-J.F., V.F., J.G., J.D.G., D.G., D.E.H., M.H., H.H., A.S.I.V., B.R.J., R.M.J., A.K., J.W.K., J.L., M.C.M.-T., R.N., R.P.O., J.E.P., R.R., J.T.R., H.M.R., C.R.S., A.V.S., S.T., C.W.T., E.v.W., L.V., K.K.O.W., M.Y., and N.D.T. performed research; C.D.S., A.Z., C.H., C.G.E., M.C.M., S.N., H.M.R., M.Y., and N.D.T. contributed new reagents/analytic tools; C.D.S., A.Z., C.H., E.A., R.B., E.C., V.C., C.R.C., E.E., C.G.E., M.-J.F., V.F., J.G., J.D.G., D.G., K.J.G., D.E.H., M.H., J.-A.H., H.H., A.S.I.V., B.R.J., R.M.J., A.K., J.W.K., J.L., K.A.M., J.A.M., M.C.M.-T., R.N., R.P.O., J.E.P., R.R., J.T.R., H.M.R., C.R.S., A.V.S., G.S., E.L.S., S.T., C.W.T., E.v.W., L.V., K.K.O.W., A.L.W., M.Y., J.A.Y., and N.D.T. analyzed data; and C.D.S., A.Z., C.H., E.A., R.B., E.C., V.C., C.R.C., E.E., C.G.E., M.-J.F., V.F., J.G., J.D.G., D.G., D.E.H., M.H., H.H., A.S.I.V., R.M.J., A.K., K.A.M., M.C.M.-T., R.P.O., J.E.P., R.R., J.T.R., H.M.R., C.R.S., G.S., S.T., C.W.T., E.v.W., L.V., M.Y., and N.D.T. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. 46575 (transcriptome), 45799 (genome), and 45805 (genome project)].

¹To whom correspondence may be addressed. E-mail: ntsutsui@berkeley.edu or smithcd@sfsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1008617108/-DCSupplemental.

nome sequence of the Argentine ant (GenBank accession no. 45799), generated using a combination of Roche 454 and Illumina sequencing. Project data are archived at the Hymenoptera Genome Database (http://hymenopteragenome.org/linepithema/genome_consortium). This project (GenBank accession no. 45805) represents an organizational advance in emerging model organism genomics (11), as it was not supported by a federal grant or genome sequencing center but, rather, was accomplished using small sums of discretionary funds provided by members of the insect genomics and Argentine ant research communities. This genome sequence provides a powerful resource for future analysis of gene families and phenotypes and candidate SNP markers for future population genetic and association mapping studies.

Results and Discussion

Genomic Features. We sequenced the Argentine ant genome ($2n = 16$) to $\sim 23\times$ coverage using source material from a single nest of the large California supercolony (Table 1 and *SI Appendix S1, Table S1*). The combined assembly of Roche 454 ($\sim 9\times$ coverage) and Illumina ($\sim 14\times$) sequencing resulted in 215.6 megabases (Mb) of scaffolded sequence (86% of the 250.8-Mb genome; ref. 12). We also recovered 12,516 bp of the mitochondrial genome in three scaffolds (*SI Appendix S1, Fig. S1*). We assembled the combined 454 and Illumina data using the Roche gsAssembler (454 Life Sciences) and Celera CABOG assemblers (13) (Table 1, *SI Appendix S1, Table S1*). Early 454-only assemblies contained homopolymer errors, but addition of Illumina sequence data produced marked improvements (Table 1) with an overall error rate estimated at 0.012% (*SI Appendix*). Overall, the Celera assembler yielded the best assembly, and addition of the Illumina data dramatically increased contig and scaffold length (Table 1). We also generated a 454 transcriptome sequence from ants of mixed age, caste, and geographic location and used it to train hidden Markov models for a custom MAKER (14) annotation pipeline (*SI Appendix S1, Table S2*).

The assembled genome appears to be relatively complete. First, our genome assembly captured 99% (246 of 248) of the core CEGMA genes (15), and 96% of them (239 of 248) were complete. Second, annotation of the cytoplasmic ribosomal protein genes revealed 83 genes, including the full set of 79 cytoplasmic ribosomal proteins (16, 17), and 4 duplicated genes (*RpS16*, *RpS23*, *RpS28*, and *RpS30*) (*SI Appendix S1, Table S2*). Third, annotation of the 67 nuclear-encoded oxidative phosphorylation genes shows that the *L. humile* genome assembly is only missing *cox7a* (*SI Appendix S1, Table S2*). This gene is also absent from the *Apis mellifera* genome, suggesting that it has been lost from the *L. humile* genome.

We generated a de novo repeat library of 523 elements and performed a whole-genome repeat annotation (Table 2). We found that 41 predictions were classified as retroint transposable elements (TEs) and 42 as DNA transposons, and 424 could not be classified. Although it has been hypothesized that the paucity of TEs in the *A. mellifera* genome is a product of eusociality, their abundance in the *L. humile* genome suggests otherwise. The *L. humile* genome also possesses remnants of viruses and viroids from a variety of families (*SI Appendix S1, Table S4*). The total repeat content was 28.7 Mb (13%), but without de novo repeat libraries, only 3.78% (8.3 Mb) of the *L. humile* assembly was identified as repetitive, with only 1.4% (3 Mb) identified as TEs. The *L. humile* genome was 62% AT, which is intermediate between *A. mellifera* (67%) and *Nasonia vitripennis* (58%). When combined with the 37.7 Mb of missing and ambiguous nucleotides, the total repetitive fraction (~ 66.4 Mb; 30.8% of the genome) was significantly higher than *A. mellifera* (10%), but lower than *N. vitripennis* (40%).

We described 16,123 genes (16,177 transcripts) in the *L. humile* official gene set (OGS1.1); 8,303 (51%) were supported by EST evidence, which confirmed 70% and 76% of the splice sites and exons for these genes, respectively. Of these genes, 1,364 were shared with *A. mellifera* and *N. vitripennis*, but not

Drosophila melanogaster (Fig. 1), and 589 were shared by the two aculeate Hymenoptera, *L. humile* and *A. mellifera*, but not with *Nasonia* (the outgroup genome for the Aculeata) or *Drosophila*. Although *L. humile* and *A. mellifera* are both social, Aculeata also contains solitary species. A total of 7,184 genes (45%) were unique to *L. humile* relative to these three other species.

We used Interproscan (18) and KEGG (19) (*SI Appendix S1, Table S5 and Fig. S2*) to identify putative functional domains and compared Gene Ontology (GO; *SI Appendix S1, Fig. S3*) terms for all *L. humile* genes relative to *D. melanogaster*, *A. mellifera*, and *N. vitripennis* (*SI Appendix S1, Tables S6 and S7*) and for the *L. humile*-specific genes. Of 16,123 genes in the OGS1.1, a total of 7,514 (44%) were annotated with at least one GO term (average = 3). Of the 7,184 (16%) genes unique to *L. humile*, 1,174 (Fig. 1 and *SI Appendix S1, Fig. S3*) had ontology terms. In the subset of genes found only in *L. humile*, 99 terms were enriched ($P < 0.05$) relative to all *L. humile* genes (*SI Appendix S1, Fig. S3*). These included odorant receptors, peptidases that may play a role in venom production, genes associated with lipid activities that could be involved in cuticular hydrocarbon (CHC) synthesis or catabolism, and DNA methylation genes, which may play a role in caste development.

Relative to other species, six cellular location terms were enriched, with most associated with the synapse ($P < 1.96E-04$) or postsynaptic membrane ($P < 4.74E-04$; *SI Appendix S1, Table S7*). Of the 17 genes associated with enriched molecular functions, 6 (all $P < 1.87E-08$ for GO categories) included cation binding and may be involved in neurological or other signal transduction ($P < 1.36E-04$) processes associated with odorant binding and olfaction, learning, memory, and behavior. Lipid catabolism ($P < 1.59E-05$), lipase ($P < 8.88E-03$), and phospholipid function (3; all $P < 0.04$) enrichments may include genes for the synthesis of CHCs involved in social communication. Electron transport, heme, and cation binding functions characteristic of cytochrome P450s were also enriched, consistent with the observed expansion of these genes.

The mean GC content of the *L. humile* genome was 37.7%, and the ratio of observed to expected CpG nucleotides [CpG(O/E)] was 1.55 (Fig. 2); both values are within the ranges reported for other Hymenoptera (*A. mellifera* and *N. vitripennis*; refs. 20 and 21). A comparison of GC compositional-domain lengths among

Table 1. Genome and sequencing statistics for three successive assemblies

	Roche gsAssembler	Celera V0.3	Celera V0.4
Total scaffolded sequence, bp*	199,810,258 (80%)	218,892,451 (87%)	215,552,578 (86%)
No. of scaffolds	3,093	3,180	3,030
No. of contigs	48,934	22,086	18,227
N50 contig size, bp	18,503 [†]	28,104	35,858
N50 scaffold size, bp	453,083	1,427,074	1,386,360
Total contig coverage	25 \times	14.6 \times	23 \times
Total reads used in scaffolds	38,902,428	21,970,969	46,779,980
454			
3-kb paired-end	1,613,759	1,276,042	1,255,603
8-kb paired-end	1,789,614	298,804	292,277
Unpaired	5,040,847	4,440,267	4,346,871
Illumina			
3-kb paired-end	NA	12,088,916	1,226,551
8-kb paired-end	NA	3,866,940	4,047,251
Unpaired	31,222,237	0	24,572,427

NA, not applicable.

*Values in parentheses indicate percent of the total genome size (12).

[†]Contigs > 500 bp; genome size, 250.8 Mb (0.26 pg).

Caste differentiation and division of labor within colonies are key innovations underlying the proliferation of ants, both numerically and taxonomically. Although the processes underlying caste differentiation are not fully understood, CpG DNA methylation is an important process for transcriptional regulation in many animals (27–29) and plays a role in the development of honey bee queens vs. workers (30). *L. humile* possesses a fully intact methylation toolkit with all three *Dnmt* genes (*SI Appendix S1*, Table S2 and Fig. S4). The only other sequenced insect genomes with the de novo methyltransferase *Dnmt3* are the pea aphid, *A. mellifera*, and *N. vitripennis* (20, 31). Interestingly, all three genes occur as single copies in *L. humile*, whereas *A. mellifera* and *N. vitripennis* have two and three copies of *Dnmt1*, respectively. Although inherited and de novo methylation have not yet been tested in *L. humile*, active methylation occurs in at least two other ant subfamilies (32).

To test for a genomic signature of CG methylation, we performed two dinucleotide analyses (*SI Appendix S1*, Figs. S5–S10 and Table S8) that revealed overall CG bias similar to *A. mellifera*, but no indication of exon or intron-specific methylation as seen in *A. mellifera* or *N. vitripennis*, respectively. We also found that dinucleotide transition SNPs associated with DNA methylation (i.e., CG↔TG) were 10-fold more prevalent in the *L. humile* SNP data compared with either of the transversions at this site (i.e., CG↔AG, CG↔GG). We observed 21,623 SNPs at the C position of the CG/TG sites, which should yield an expected 17,428 CG↔TG transition SNPs (80.6%) based on the overall observed genome rate of 4.15 transition mutations per transversion. However, we measured 18,611 CG↔TG transitions (86%), which indicated significant bias of CG↔TG SNPs relative to all other mutations at this site ($\chi^2 = 80.3$, $P < 0.001$). This bias in dinucleotide CG↔TG mutation relative to all other dinucleotide mutations is consistent with the variable distribution seen for the CG dinucleotide in the genome and suggestive of in vivo methylation (*SI Appendix S1*, Figs. S5, S6, and S10). Genes ranked with the largest numbers of CG↔TG SNPs included *major facilitator superfamily transporters*, male sterility proteins, several classes of zinc finger transcription factors, and several Ig superfamily cell adhesion proteins implicated in neuronal development (*SI Appendix S1*, Table S8).

Wing polyphenism and reproductive division of labor between queens and workers are two important features associated with eusociality in ants. The gene networks that underlie wing and reproductive development have evolved to be differentially expressed between winged reproductive castes and wingless sterile worker castes in response to environmental factors (33, 34). We found that the mean CpG(O/E) (35) for genes involved in reproductive system development ($n = 38$; mean = 1.21; $P < 0.05$) and apoptosis ($n = 18$; mean = 1.22; $P < 0.05$) were significantly lower than the mean CpG(O/E) (mean = 1.63) for 50 random genes resampled 10,000 times. (Fig. 2 and *SI Appendix*). Genes in the wing polyphenism network were not significantly different from the average gene CpG(O/E) ($n = 37$; mean = 1.50; $P = 0.4865$). The mean CpG(O/E) of the *Drosophila* orthologs (coding regions) that underlie wing development (mean = 0.95; $P = 0.86$), reproduction (mean = 0.98; $P = 0.98$), and apoptosis (mean = 1.00; $P = 0.99$) were not significantly different (*SI Appendix S1*, Fig. S11) than average gene CpG(O/E) in *Drosophila*. These results indicate that developmental genes in the reproduction and apoptosis networks have distinct germ-line methylation signatures relative to the rest of the genome, whereas genes in the wing polyphenism network do not.

Yellow genes occur in insects as well as some bacteria and fungi, but they are curiously absent in all noninsect metazoans (36, 37). The initially described *yellow-y* gene (*Y-y*) functions in cuticle pigmentation in *D. melanogaster* (38), but these genes have also been implicated in processes such as male courtship behavior (39), follicle cell function, and egg development (40). Proteins from a gene expansion of the ancestral *yellow-e3* gene in *A. mellifera* [the major royal jelly (*mrj*) subfamily] are involved in a regulating reproductive division of labor, but also have age-, sex-, caste-, and brain-specific expression (37, 41).

In the *L. humile* assembly we detected 10 *yellow* genes and 10 major royal jelly protein-like genes (Fig. 3C and *SI Appendix S1*, Fig. S14). The 8 *yellow* genes (*LhumY-y, -b, -c, -e, -e3, -g, -g2, and -h*) are similar to those of *D. melanogaster* and are likely orthologs. The remaining 2 are putative orthologs of the *yellow-x1* and *-x2* genes, which have been found only in the Hymenoptera. Interestingly, the *L. humile* genome contains an independent radiation of major royal jelly protein-like (*mrjpl*) genes similar to those in *A. mellifera* and *N. vitripennis*. The *mrj* and *mrjpl* gene sets of all three focal taxa each form their own strongly supported clade within the monophyletic *mrj* subfamily (Fig. 3C and *SI Appendix S1*, Fig. S14). These independent radiations in different hymenopteran lineages may indicate that the ancestral gene had a tendency to proliferate, allowing *mrjpls* to take on new functions and to respond quickly to new forms of selection (42).

Invasion Biology. Although Argentine ants are a widespread and familiar invasive species, many details of their invasion biology remain shrouded in mystery. To clarify some of the biological

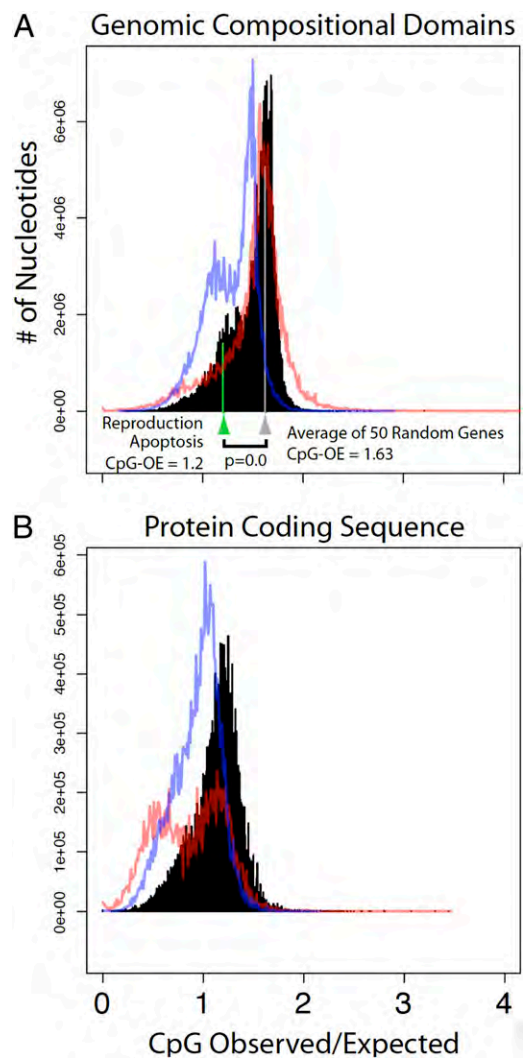


Fig. 2. Distribution of CpG(O/E) predictions in genomic compositional domains (A) and protein coding regions (B) for *L. humile* (black), *A. mellifera* (red line), and *N. vitripennis* (blue line). The combined average CpG(O/E) for all genes involved in either reproductive (mean = 1.21) or apoptosis processes (mean = 1.22) (green arrow and line) is significantly lower than the average CpG(O/E) for 50 random genes resampled 1,000 times as a control (gray arrow and line; $P < 0.05$, statistical randomization test).

processes that may be related to their invasive success, we annotated and analyzed genes that are likely to be involved in Argentine ant immune processes, insecticide resistance, and dietary detoxification. We also identified a large number of candidate SNP markers that will be valuable tools for identifying source populations in the Argentine ant's native range, reconstructing the history of introductions around the world, and identifying current routes of transport.

The Argentine ant's unicolonial social structure allows introduced populations to attain enormous population densities in its introduced range. However, it is not clear how Argentine ants control the proliferation of pathogens and parasites in these high-density populations, particularly in light of the extremely low levels of genetic diversity that typify introduced populations (5, 8). We annotated immune genes primarily from the *Toll*, *Imd*, *JNK*, and *JAK/STAT* signaling pathways. In total, 90 of 202 immune genes had reciprocal best matches (SI Appendix S1, Tables S2 and S16). Of these 202 genes, 152 are present in the *Drosophila* genome and 78 in *Apis*. Thus, *L. humile* appears similar to *Apis* and *Nasonia* in having markedly fewer immune genes than *Drosophila*. Moreover, the many hygienic behaviors and chemical secretions described for the social Hymenoptera may play key roles in controlling pathogens.

Introduced populations of Argentine ants are also frequently targets of heavy insecticidal control, suggesting that genes conferring insecticide resistance may be under strong positive selection. Like many invasive ants, Argentine ants are also extreme dietary generalists, and this plasticity may require an underlying ability to detoxify many different components of their food sources. Cytochrome P450s are a family of heme-thiolate enzymes that catalyze a diverse array of chemical reactions in nearly all organisms (43), and they have been implicated in both insecticide resistance and dietary detoxification.

The *L. humile* genome encodes 111 cytochrome P450s (SI Appendix S1, Table S2 and SI Appendix S2, Fig. S13), substantially more than *A. mellifera* (46 genes) (44) and *N. vitripennis* (92 genes) (45). It has been hypothesized that the paucity of P450 genes in the genome of *A. mellifera* is a product of its eusocial life history (44), but their abundance in *L. humile* suggests that eusociality alone cannot explain their scarcity in *Apis*. The *L. humile* genome encodes 69 CYP3 clan P450s (SI Appendix S2, Fig. S13), the most in any sequenced insect genome. CYP3 P450s are associated with oxidative detoxification of xenobiotics (46), and their abundance

in *L. humile* may be an adaptation to a variety of toxins encountered in the diet of this generalist ant, compared with the rather specialized diet of the honey bee.

To facilitate the development of population genetic markers for future analyses of the origin, history, and movement of Argentine ants, we scanned the genome for SNPs. Because the source material contained a single queen pupa and ~100 workers from a single nest (SI Appendix S1, Table S1), we were able to identify predicted polymorphic sites. We discovered 231,248 SNPs (0.9 SNP/kb) that occurred in at least three reads and at least 10% of overlapping reads. A total of 5,734 genes (36%) had at least 1 SNP, with 14,136 SNPs (6%, 0.7 SNP/kb) in exons, 26,720 (12%, 0.9 SNP/kb) in introns, 66,830 (29%, 2.8 SNP/kb) in annotated repeats, and 123,492 (53%, 1.1 SNP/kb) in nonrepeat intergenic regions. Although 77% of exonic SNPs would be nonsynonymous (23% synonymous) if all substitutions were equally likely, we observed that only 54% of SNPs were nonsynonymous and 46% were synonymous ($P < 0.001$). When each type of SNP was normalized to the total number of possible sites for that type, we observed 4.9×10^{-4} SNPs per nonsynonymous site and 1.9×10^{-3} SNPs per synonymous site. As expected, we saw a 3.87-fold excess of synonymous SNPs in the OGS1.1 gene set.

We also manually investigated the genes with the highest number of SNPs per kb of exon sequence (SI Appendix S1, Table S3). Interestingly, six *L. humile* genes with Interproscan-predicted functions similar to male sterility and gametogenesis genes in *Drosophila* showed a high degree of polymorphism in *L. humile* (24–38 SNPs/kb). These polymorphic genes may be under sexual selection or diversifying selection and could be useful for studying different patriline lineages in native vs. invasive ants. Cytochrome P450s, lipid metabolism genes, and transcription factors were also ranked highly, with >75 SNPs in the exon and intronic regions.

Conclusion. With the sequencing and annotation of the Argentine ant genome and the development of associated genomic resources, this species is well positioned to become a model organism in which powerful genetic approaches can be coupled with a wealth of natural history and behavioral and ecological knowledge. Given the immense financial and ecological costs associated with introductions of Argentine ants, these tools will likely find widespread application and produce tangible benefits for agriculture, societies, and ecosystems. Finally, the evolutionary forces driving some of the unusual and remarkable genomic patterns reported here

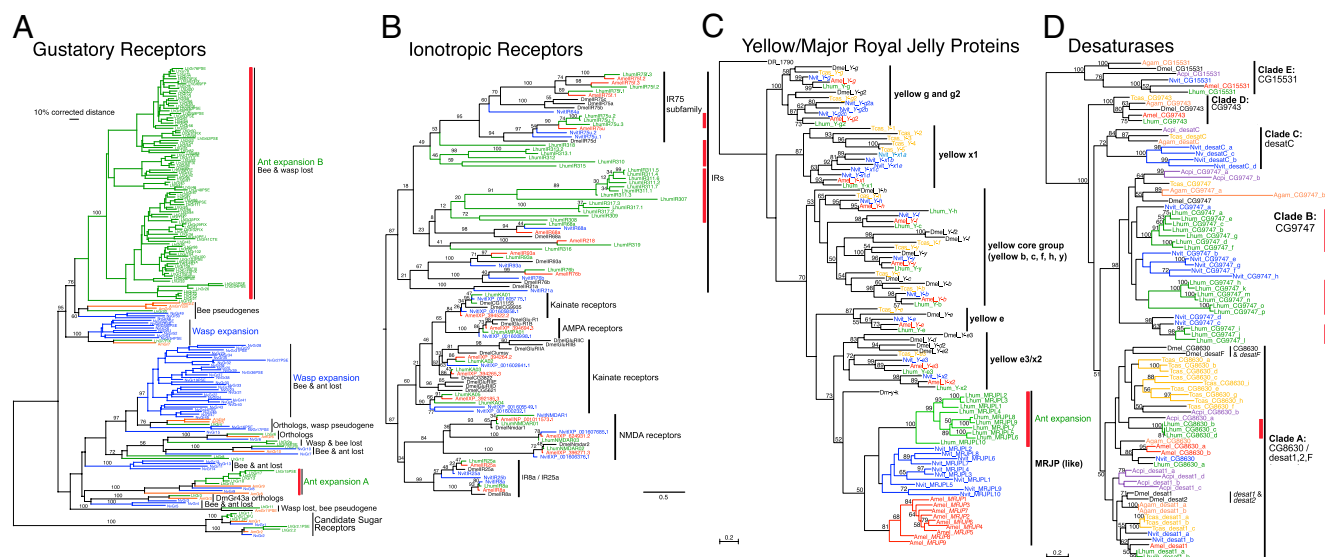


Fig. 3. Gene expansions in the Argentine ant genome. (A) Gustatory receptors. (B) Ionotropic receptors. (C) Yellow/major royal jelly proteins. (D) Desaturases. Vertical red bars indicate *L. humile* gene duplications and expansions. The *L. humile* genome also possesses an enormous expansion of odorant receptors (Ors; SI Appendix S2, Fig. S12). Green, *L. humile*; red, *A. mellifera*; blue, *N. vitripennis*; black, *D. melanogaster*.

remain unknown and will be productive avenues for future research that explores the basis of eusociality and the causes and consequences of biological invasions.

Materials and Methods

Source Material. We collected an Argentine ant colony fragment from a residential orchard in Santa Clara County, CA. We confirmed that these ants belong to the large supercolony that dominates the introduced range in California using behavioral assays, microsatellite genotyping, and analysis of CHC profiles (*SI Appendix S1, Table S14 and Fig. S17*).

Library Preparation and Sequencing. Transcriptome. cDNA was generated from mixed source material and sequenced using Roche 454 Genome Sequencer LR70 FLX technology (Table 1 and *SI Appendix S1, Table S1*). This process yielded ~128 Mb of DNA sequence, which was assembled into 20,070 contigs. **Genomic.** Genomic DNA was extracted and purified from a single queen pupa (Saratoga) and sequenced using seven runs of 454 FLX Titanium sequencing. Additional worker-derived genomic libraries were constructed and sequenced on the 454 and Illumina platforms.

- Hölldobler B, Wilson EO (1990) *The Ants* (Belknap Press, Cambridge, MA), p 733.
- Patek SN, Baio JE, Fisher BL, Suarez AV (2006) Multifunctionality and mechanical origins: Ballistic jaw propulsion in trap-jaw ants. *Proc Natl Acad Sci USA* 103:12787–12792.
- Holway DA, Lach L, Suarez AV, Tsutsui ND, Case TJ (2002) The causes and consequences of ant invasions. *Annu Rev Ecol Syst* 33:181–233.
- Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proc Natl Acad Sci USA* 98:1095–1100.
- Suarez AV, Holway DA, Tsutsui ND (2008) Genetics and behavior of a colonizing species: The invasive Argentine ant. *Am Nat* 172(Suppl 1):S72–S84.
- van Wilgenburg E, Torres CW, Tsutsui ND (2010) The global expansion of a single ant supercolony. *Evolutionary Applications* 3:136–143.
- Holway DA, Suarez AV, Case TJ (1998) Loss of intraspecific aggression in the success of a widespread invasive social insect. *Science* 282:949–952.
- Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proc Natl Acad Sci USA* 97:5948–5953.
- Tsutsui ND, Suarez AV, Holway DA, Case TJ (2001) Relationships among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. *Mol Ecol* 10:2151–2161.
- Weston DP, Holmes RW, You J, Lydy MJ (2005) Aquatic toxicity due to residential use of pyrethroid insecticides. *Environ Sci Technol* 39:9778–9784.
- Smith CR, Dolezal A, Elyahu D, Holbrook CT, Gadau J (2009) Ants (Formicidae): Models for social complexity. *Cold Spring Harb Protoc* 7:pbm125.
- Tsutsui ND, Suarez AV, Spagna JC, Johnston S (2008) The evolution of genome size in ants. *BMC Evol Biol* 8:64.
- Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24:2818–2824.
- Cantarel BL, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18:188–196.
- Parra G, Bradnam K, Korfi I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067.
- Uechi T, Tanaka T, Kenmochi N (2001) A complete map of the human ribosomal protein genes: Assignment of 80 genes to the cytogenetic map and implications for human disorders. *Genomics* 72:223–230.
- Marygold SJ, et al. (2007) The ribosomal protein genes and minute loci of *Drosophila melanogaster*. *Genome Biol* 8:R216.
- Quevillon E, et al. (2005) InterProScan: Protein domains identifier. *Nucleic Acids Res* 33(Web Server issue):W116–W120.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(Web Server issue):W182–W185.
- Werren JH, et al.; Nasonia Genome Working Group (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343–348.
- Weinstock GM, et al.; Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931–949.
- Torres CW, Brandt M, Tsutsui ND (2007) The role of cuticular hydrocarbons as chemical cues for nestmate recognition in the invasive Argentine ant (*Linepithema humile*). *Insectes Soc* 54:363–373.
- Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: Expansion of the odorant, but not gustatory, receptor family. *Genome Res* 16:1395–1403.

Assembly. We created several assemblies using Newbler and CABOG assembler (13), as described in *SI Appendix*.

Annotation. We first used the automatic annotation pipeline MAKER (14) to annotate the genome of *L. humile*. We also manually annotated ~1,000 genes (*SI Appendix S1, Table S2*), including a number that are related to fundamental biological processes (oxidative phosphorylation, ribosomal), complex behaviors (learning, memory and aggression), sensory biology (vision, chemoreception), insecticide resistance, immunity, and developmental networks (wings, reproduction).

ACKNOWLEDGMENTS. We thank G. Anderson for providing the ants used in this project; A. Smith for assistance with transcriptome sequencing; L. Tonkin and the V. Coates Genomic Sequencing Facility and the Center for High Performance Computing for assistance and use of facilities; G. Robinson for support; M. Wong for script support; M. Goodisman for valuable discussion; B. Hunt and S. Yi for generously sharing useful data; and B. Moore for assistance with SNP analysis. Infrastructure for this work was supported in part by National Human Genome Research Institute National Institutes of Health (NIH) Grant 1R01HG004694 (to M.D.Y.), National Institute of Mental Health NIH Grant 5SC2MH086071 (to C.D.S.), and the University of Illinois at Urbana–Champaign Romano Professorial Scholarship (H.M.R.).

- Robertson HM, Gadau J, Wanner KW (2010) The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol Biol* 19(Suppl 1):121–136.
- Crosset V, et al. (2010) Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet* 6:e1001064.
- Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB (2009) Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136:149–162.
- Field LM, Lyko F, Mandrioli M, Prantero G (2004) DNA methylation in insects. *Insect Mol Biol* 13:109–115.
- Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481–514.
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet* 33(Suppl):245–254.
- Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827–1830.
- Wang J, et al. (2007) An annotated cDNA library and microarray for large-scale gene-expression studies in the ant *Solenopsis invicta*. *Genome Biol* 8:R9.
- Kronforst MR, Gilley DC, Strassmann JE, Queller DC (2008) DNA methylation is widespread across social Hymenoptera. *Curr Biol* 18:R287–R288.
- Abouheif E, Wray GA (2002) Evolution of the gene network underlying wing polyphenism in ants. *Science* 297:249–252.
- Khila A, Abouheif E (2010) Evaluating the role of reproductive constraints in ant social evolution. *Philos Trans R Soc Lond B Biol Sci* 365:617–630.
- Elango N, Hunt BG, Goodisman MAD, Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci USA* 106:11206–11211.
- Arakane Y, et al. (2010) Identification, mRNA expression and functional analysis of several yellow family genes in *Tribolium castaneum*. *Insect Biochem Mol Biol* 40:259–266.
- Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R (2006) Evolution of the *Yellow/Major Royal Jelly Protein* family and the emergence of social behavior in honey bees. *Genome Res* 16:1385–1394.
- Wittkopp PJ, True JR, Carroll SB (2002) Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* 129:1849–1858.
- Drapeau MD, Radovic A, Wittkopp PJ, Long AD (2003) A gene necessary for normal male courtship, *yellow*, acts downstream of *fruitless* in the *Drosophila melanogaster* larval brain. *J Neurobiol* 55:53–72.
- Claycomb JM, Benasutti M, Bosco G, Fenger DD, Orr-Weaver TL (2004) Gene amplification as a developmental strategy: Isolation of two developmental amplicons in *Drosophila*. *Dev Cell* 6:145–155.
- Schmitzová J, et al. (1998) A family of major royal jelly proteins of the honeybee *Apis mellifera* L. *Cell Mol Life Sci* 54:1020–1030.
- Johnson BR, Linksvayer TA (2010) Deconstructing the superorganism: Social physiology, groundplans, and sociogenomics. *Q Rev Biol* 85:57–79.
- Isin EM, Guengerich FP (2007) Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim Biophys Acta* 1770:314–329.
- Claudianos C, et al. (2006) A deficit of detoxification enzymes: Pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol* 15:615–636.
- Oakeshott JG, et al. (2010) Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. *Insect Mol Biol* 19(Suppl 1):147–163.
- Li XC, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol* 52:231–253.

Chapter 3: A map of IR expression in the larval chemosensory system

Introduction

Among the 60 intact IRs present in the genome of *Drosophila melanogaster*, only a small fraction (16) are expressed in antennae and function as olfactory receptors (Benton et al., 2009). This suggests that the 45 other IRs, referred to as non-olfactory IRs, might be expressed in other types of tissues or at different life stages. In order to understand the function of these non-olfactory IRs, it is essential to know in which cells or organs they are expressed. In this chapter, I describe the production of flies expressing transgenic reporters for all non-olfactory IRs, and the experiments that I performed to generate a comprehensive map of IR expression in the larval stage of *Drosophila melanogaster*.

The larva of Drosophila melanogaster: a model for the study of taste perception

While most studies on chemosensory perception in *Drosophila* have been conducted on adults, the larval stage has also been used as a model for olfactory and gustatory experiments (Gerber and Stocker, 2007). Larvae present several advantages compared to adults. Although the general architecture of their chemosensory system is conserved, the larval one is much simpler compared to adults. In addition, larval stages represent an incredible period of growth: the length of the animal increases from 0.5 mm

to 7-8 mm (Markow et al., 2009) in five days, which represents an increase in size of almost 75% per day. Therefore, larvae spend most of their time eating or foraging and are thus particularly good models to study questions related to taste and feeding.

Olfaction in Drosophila larvae

Although very similar to the adult one, the larval olfactory system has a reduced number of cells. The larval olfactory organ is called the dorsal organ (DO) (Python and Stocker, 2002) and is located at the tip of the head. It is composed of a dome surrounded by six peripheral sensilla (Gerber and Stocker, 2007). The DO hosts 21 olfactory sensory neurons (OSN) that express ORs like adult OSNs. Olfactory stimuli are thus detected by single OSNs, whose activation is sufficient to drive appropriate behaviours (Fishilevich et al., 2005). Each OSN projects to one of the 21 glomeruli of the larval antennal lobe (Figure 2C), whose function is similar to the adult one (Kreher et al., 2005; Ramaekers et al., 2005; Thum et al., 2011). From there, only 21 PNs connect to the larval mushroom body and lateral horn, which are less complex but have similar functions than the adult ones (Marin et al., 2005; Schleyer et al., 2011; Schroll et al., 2006). There is thus a 1:1:1 ratio between the OSNs, AL glomeruli and PNs (and in a lesser extent of MB calyx glomeruli) in larvae.

Taste perception in Drosophila larvae

Like adults, larvae only have peripheral and internal gustatory organs in the head dedicated to taste perception, as well as putative gustatory sensory neurons along their body segments. As for the olfactory system, the larval gustatory system is much simpler than the adult one (Gerber and Stocker, 2007). The terminal (TO) and ventral (VO) organs are the two major cephalic peripheral gustatory organs, although six GSNs also innervate the sensilla around the dome of the DO (Python and Stocker, 2002). 37 and 7 neurons innervate the TO and VO, respectively. Most of them are gustatory however some are likely to be mechano- or thermosensory (Gerber and Stocker, 2007; Liu et al., 2003b; Python and Stocker, 2002). Like the DO, the TO and the VO have their own ganglion; however, a few TO neurons have their somata in the DO (Stocker, 1994). Larvae have three internal gustatory organs called the dorsal (DPS), ventral (VPS) and posterior (PPS) pharyngeal sensilla (Figure 2C), containing 17-18, 16 and 6 neurons, respectively (Colomb et al., 2007; Singh and Singh, 1984).

According to GR transgenic reporters, many GRs are expressed in larval GSNs. Although several GRs were found only in single neurons, most GRs are expressed in several neurons (up to 12), suggesting a more complex combinatorial code than in the olfactory system (Kwon et al., 2011). 27 of the GRs expressed in adults are also found in larvae, where they are likely to play similar roles. In addition, 10 GRs are specific to larvae (Kwon et al., 2011; Weiss et al., 2011). Importantly, although larvae are able to sense sugars (Heimbeck et al., 1999; Rohwedder et al., 2012), no

reporters for the known sugar-sensing GRs (GR5a, GR61a and the GR64 cluster) are expressed in larvae (Kwon et al., 2011), suggesting that different sugar receptors may be used in the larval stage.

Larval GSNs project to the SOG through the maxillary (TO and VO) and labial (VPS) nerves that connect to the posterior side of the SOG, or through the labral nerve that connects more anteriorly (Figure 2C) (Gendre et al., 2004; Python and Stocker, 2002). Neurons from the peripheral organs project more laterally whereas neurons from internal organs project more medially, and sometimes even join each other at the centre (Colomb et al., 2007; Kwon et al., 2011), although they remain exclusively ipsilateral (Gerber and Stocker, 2007). The quality of the tastants detected by particular GSNs is also likely to be reflected in their projection patterns, although this is less clear than in adults (Colomb et al., 2007). Second order neurons have not been identified in larvae.

Production of IR-Gal4 lines

Chemoreceptor genes are usually very lowly expressed (Graveley et al., 2011). In order to study the expression pattern of IRs, *in situ* hybridisation techniques or RT-PCR thus turned out to be inappropriate due to their low detection level (unpublished data). I therefore used the Gal4-UAS system (Brand and Perrimon, 1993), which provides a sensitive method to detect receptor expression. In order to generate IR-Gal4 lines, the intergenic region up to 3 kb upstream of the start codon of each IR was selected as a promoter. These sequences were cloned into the *pGal4-attB* vector. These were used to transform flies bearing *attP* sites at the suitable

location (Markstein et al., 2008). Expressing phi-C31 integrase in germ cells allows recombination of the *attB* and *attP* sites (Bischof et al., 2007), thus the integration of the IR-Gal4 transgene into a defined site in the genome. The use of defined integration sites allows standardisation of genomics effects between transgenes. We used these IR-Gal4 lines to drive the expression of a UAS-mCD8:GFP fluorescent reporter transgene in cells that normally express each particular IR.

Expression of IRs in larval chemosensory organs

Expression of all 45 IR-Gal4 lines that were produced, as well as reporters for the 16 olfactory IRs, were tested for expression in 3rd instar larvae. In total, 33 IR-Gal4 lines drive expression in small, bilaterally symmetric populations of neurons in larval cephalic gustatory organs, including 9 peripherally (Figure 4), 15 internally (Figure 5), and 8 in both types of organs. In addition to gustatory organs in the head, seven IRs (IR7d, IR7g, IR10a, IR25a, IR68b, IR76b and IR85a) were also detected in putative chemosensory cells along the abdomen (Figure 4). Among the 11 non-olfactory IRs that are expressed in more than one cell, 9 are present in two or more organs. Among these, four belong to the IR7 subfamily of eight receptors (IR7a-IR7g and IR11a). Amongst these, IR7d is also expressed in one multi-dendritic non-chemosensory neuron near the mouth hooks (Figure 4). The function of this neuron is unknown but is supposed to be thermosensory.

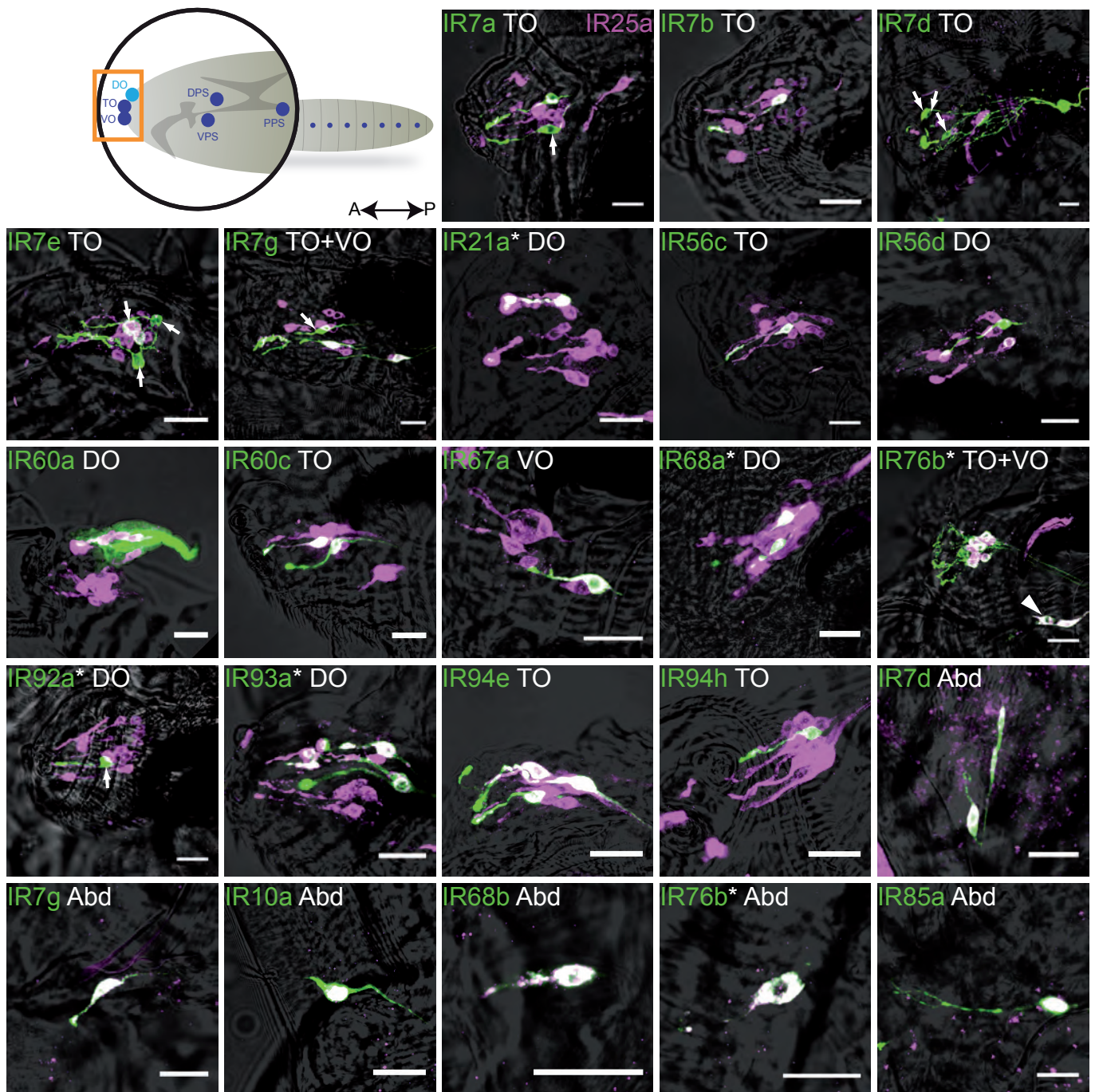


Figure 4 | Expression of IRs in *Drosophila* larval external gustatory organs. Top left panel: scheme of *Drosophila* larval anatomy, with a zoom on the head, and chemosensory organs shown in blue. Other panels: Immunofluorescence with anti-GFP (green) and anti-IR25a (magenta) antibodies (overlaid on bright-field images) on whole-mount tissues from animals expressing a membrane targeted GFP reporter transgene (UAS-mCD8:GFP) under the control of the indicated IR promoter GAL4-driver transgenes. The scale bars represent 20 μ m. The organs stained with anti-GFP are indicated on each picture. Names followed by a star are IRs also found in the adult 3rd antennal segment. Small arrows indicate IR neurons that do not co-express IR25a. Arrowheads point out VO cells on pictures showing both TO and VO expression. DO: dorsal organ, TO: terminal organ, VO: ventral organ, Abd: abdominal segments.

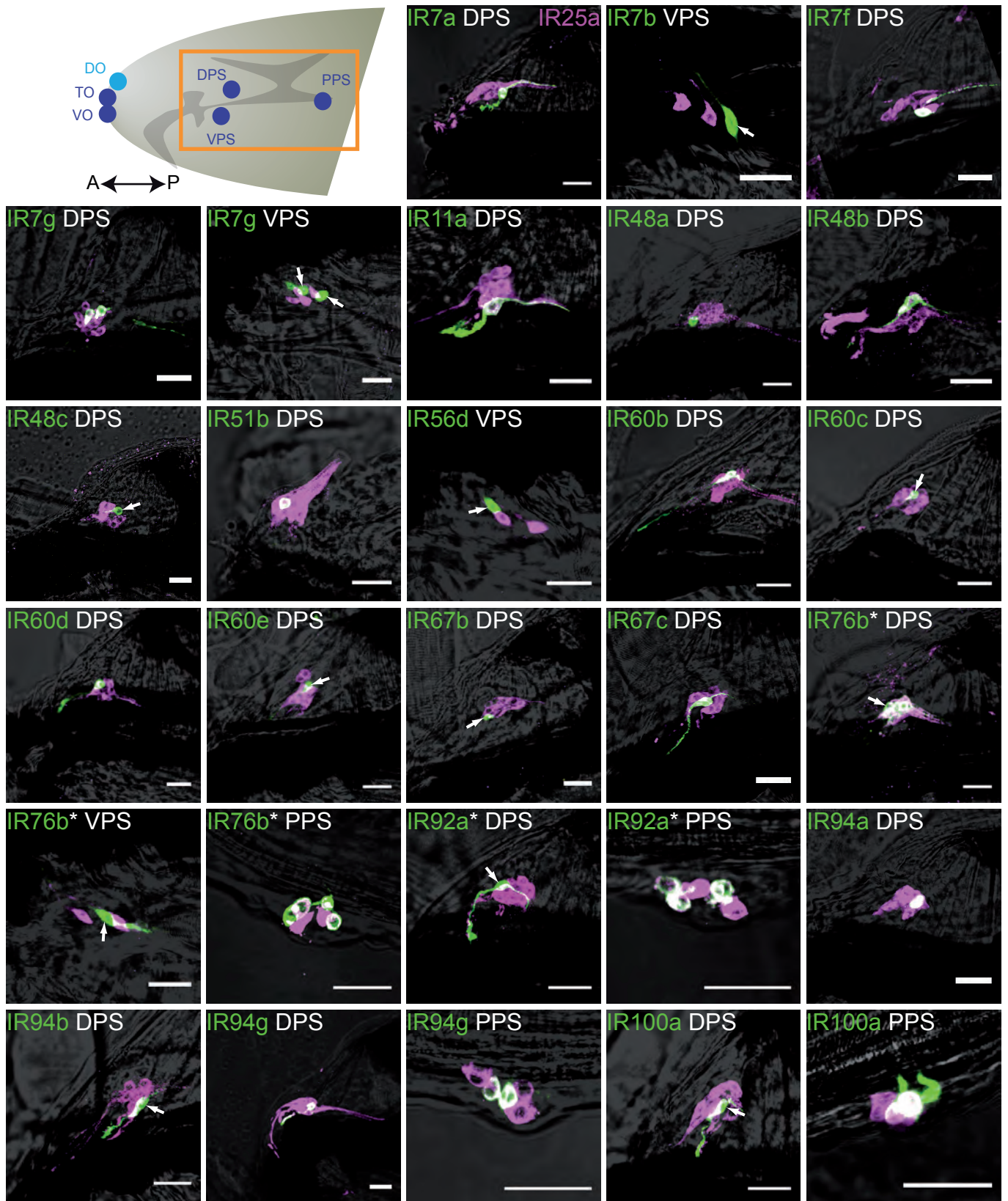


Figure 5 | Expression of IRs in *Drosophila* larval internal gustatory organs. Top left panel: scheme of *Drosophila* larval head anatomy, with chemosensory organs shown in blue. Other panels: Immunofluorescence with anti-GFP (green) and anti-IR25a (magenta) antibodies (overlaid on bright-field images) on whole-mount tissues from animals expressing a membrane targeted GFP reporter transgene (UAS-mCD8:GFP) under the control of the indicated IR promoter GAL4-driver transgenes. The scale bars represent 20 μm. The organs stained with anti-GFP are indicated on each picture. Names followed by a star are IRs also found in the adult 3rd antennal segment. DPS: dorsal pharyngeal sensilla, VPS: ventral pharyngeal sensilla, DPO: dorsal pharyngeal organ, PPS: posterior pharyngeal sensilla. Small arrows indicate IR neurons that do not co-express IR25a.

Only five of the 16 olfactory IR-Gal4 lines (IR21a, IR25a, IR68a, IR92a, IR93a) are expressed in one or more neurons of the DO (Figure 4). In addition, IR92a, a receptor for ammonia (Benton et al., 2009), was also found in one neuron of the DPS and five neurons of the PPS, suggesting that ammonia is detected at both olfactory and gustatory levels. The IR60a Gal4-line also drives expression in the DO (Figure 5), although fluorescence seems to be spread around all neurons, questioning the faithfulness of this Gal4-line. Another transgene was thus produced that included the 3' region of the gene as well. However, this transgene no longer drives expression in the larva (data not shown). Finally, IR76b was not found in larval olfactory neurons but is expressed in many gustatory neurons. Similarly to its function in the olfactory system (Abuin et al., 2011), this IR is likely to function as a co-receptor in taste perception.

In parallel with this analysis in the larva, another PhD student in the Benton lab, Anantha Krishna Sivasubramaniam, has investigated the expression of IR-Gal4 lines in adults and found that 23 of them drive expression to gustatory neurons. Thus, IRs are likely to be involved in taste perception in adults as well, whereas their olfactory function seems almost entirely restricted to the adult stage. Among all IRs that are expressed in chemosensory organs, 15 are larval-specific, 18 are adult-specific, and 19 are expressed at both life stages. In total, almost two thirds (63%) of them are life-stage specific. This suggests that many compounds are detected by adults and not by larvae, and *vice versa*. I analysed whether the phylogeny of IRs could be reflected in their expression patterns. However, no

correlation was found between these parameters, and even IRs belonging to the same tandem cluster have rather divergent expression patterns (Figure 7).

Central projections of IR neurons

The use of the mCD8:GFP reporter allows visualisation of the axons of labelled neurons, making it possible to trace their projections from peripheral chemosensory organs to the central nervous system. For all IR Gal4-lines driving expression in gustatory neurons, labelled axons were detected in the suboesophageal ganglion (SOG), confirming their potential gustatory function. As for GRs (Colomb et al., 2007; Kwon et al., 2011), the projection patterns of IR axons in different areas of the SOG depend on which organ their somata are located in and which nerve they follow. Gal4 lines that drive expression to both peripheral and internal organs thus show generally more complex projection patterns. In general, projection patterns are much less likely to reflect neuronal function in larvae than in adults (Colomb et al., 2007; Kwon et al., 2011). In a few cases however, additional variability was observed. For example, both IR94e and IR94h are expressed in the TO (Figure 4), but IR94h neurons project more laterally than IR94e (Figure 6) suggesting that these neurons detect tastants of different value.

Similar to OR neurons, IR neurons from the DO project to the larval AL. Consistent with additional expression in gustatory cells, IR76b and IR92a neurons also project to the SOG. Finally, IR neurons from thoracic and



Figure 6 | Projections of IR-neurons to the larval CNS. Top left panel: scheme of *Drosophila* larval brain. LAL: larval antennal lobe, SOG: subesophageal ganglion, VNC: Ventral Nerve Cord. Other panels: Immunofluorescence with anti-GFP (green) and nc82/Brp (magenta) antibodies on whole-mount brains from animals expressing a membrane targeted GFP reporter transgene (UAS-mCD8:GFP) under the control of the indicated IR promoter GAL4-driver transgenes. Names followed by a star are IRs also found in the adult 3rd antennal segment. The scale bars represent 20 μ m.

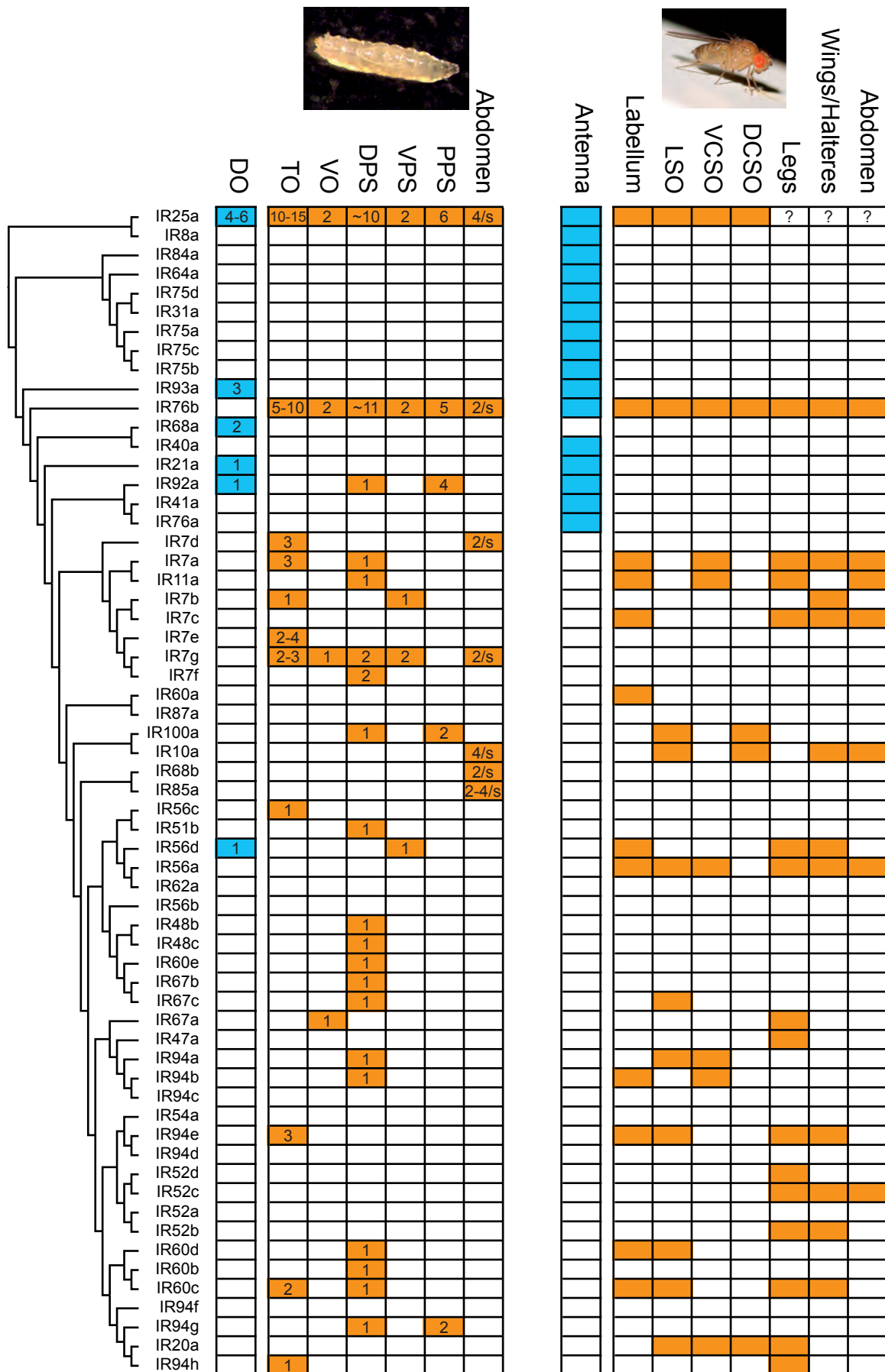


Figure 7 | Summary of expression data, and comparison with phylogeny. The tree on the left shows the phylogenetic relationships between *Drosophila melanogaster* IRs. Sequences were aligned with MUSCLE, and the tree was made with RaxML under the WAG model of substitution, with 1000 bootstrap replicates. The tables show the number of cells expressing each IR in all olfactory (blue) and gustatory (orange) organs. See text and Figures 3-4 for the meaning of the abbreviations. The expression data on adults was obtained from A.K. Sivasubramaniam.

abdominal segments do not project to the SOG, but to the ventral nerve cord (VNC), where they are likely to connect to the SOG through second-order neurons (Bader et al., 2007).

A few IR-Gal4 lines drive expression to other regions of the central nervous system. Cells around the VNC were visible for 6 IR-Gal4 lines, and two other lines drove expression in various areas of the protocerebrum. In addition, the pseudogenic IR48a-Gal4 line drives expression to a variable number of cells in the Mushroom Body (data not shown). All these expression patterns are likely to be due to the unfaithfulness of the Gal4 lines and do not reflect endogenous cerebral expression. The fact that IR84a-Gal4 but not IR8a-Gal4 marks neurons in the VNC supports this hypothesis, as IR84a localisation and function in the antenna are absolutely depending upon IR8a (Abuin et al., 2011).

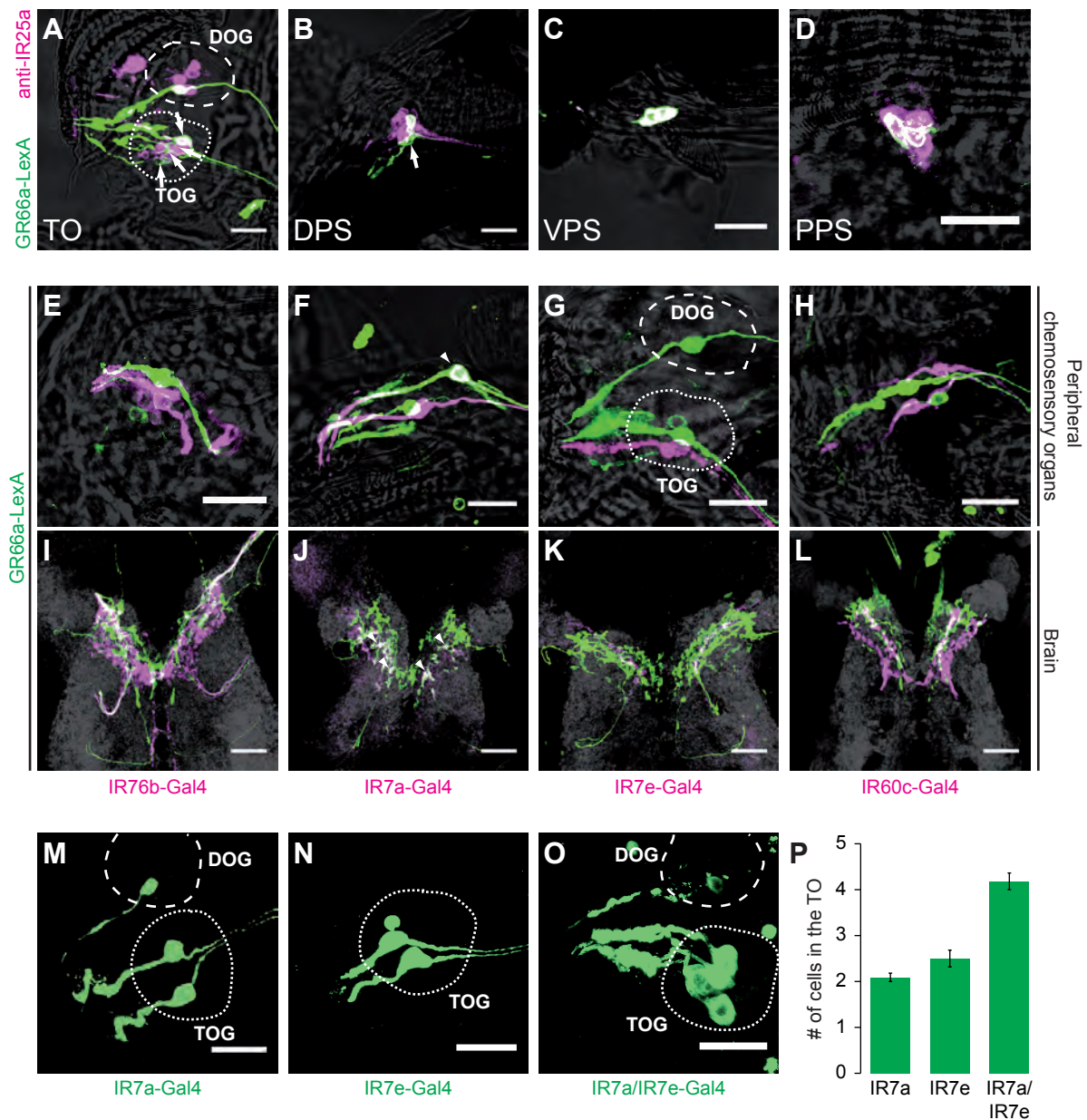
Receptor co-expression in gustatory neurons

Receptors that are expressed in the same neuron are likely to be activated by similar cues or even to assemble to form a functional channel. In order to better understand the function of IRs, I determined whether they are co-expressed with other IRs or other types of taste receptors. Here, I present a limited study of the co-expression of some IRs of particular interest with the bitter-sensing GR66a and with other IRs in larvae.

IR25a is a co-receptor in adult antennae and is likely to be co-expressed with many other IRs. The availability of an antibody against this

protein allowed labelling of neurons that express IR25a in all IR-Gal4 lines that were tested above (Figure 4-5). This showed that IR25a is broadly expressed across all larval gustatory organs and that it co-localises with many IR reporters. However, for 13 IR-Gal4 lines, at least one neuron does not express IR25a (Figure 4-5). This supports the presumed function of IR25a as a gustatory IR co-receptor, but also suggests that some IRs may function alone or with other co-receptors. In this experiment, the IR25a-Gal4 line did not faithfully recapitulate the labelling of the IR25a antibody (data not shown), suggesting that this Gal4 line is not well suitable for performing experiments targeted in IR25a neurons.

GR66a is a receptor that is expressed in most if not all bitter-sensing neurons in adults (Lee et al., 2009; Marella et al., 2006; Moon et al., 2006). In order to test, whether IR25a is expressed in GR66a neurons in larvae, immunofluorescence was performed with the anti-IR25a antibody on larvae expressing rCD2:GFP under the control of GR66a-LexA (Lai and Lee, 2006). In the TO, GR66a is expressed in 4 neurons from the distal group (with their somata in the TOG) and 1-2 neurons from the dorsolateral group (with their somata in the DOG). In addition, it is also expressed in two neurons in each of the pharyngeal sensing organs (Kwon et al., 2011). No neuron from the TO-distal subgroup expresses both receptors, and only one neuron from the TO-dorsolateral subgroup is marked by both GR66a-LexA and anti-IR25a (Figure 8A). In the DPS, only one of the two GR66a neurons expresses IR25a (Figure 8B), and in the VPS and the PPS, both GR66a neurons also express IR25a (Figure 8C-D). These results show that IR25a



is expressed indifferently in neurons that detect appetitive or aversive tastants. This IR may assemble with IRs of both functions and be involved in the detection of several classes of tastants.

In addition to IR25a, IR76b is also a putative co-receptor in larval gustatory neurons. Most IR76b neurons also express IR25a (Figure 4-5). In order to test, whether IR76b and GR66a are co-expressed, two different reporters were expressed together through the Gal4 and LexA:VP16 expression systems and detected by immunofluorescence. GR66a-LexA was used to mark bitter sensing neurons, while IR76b expression was visualised with IR76b-Gal4. In the TO, no overlap was observed between these two lines (Figure 8E). Because the projections of IR76b neurons are also not overlapping with those of GR66a neurons (Figure 8I), I concluded that these two receptors are not co-expressed. This shows that unlike IR25a, IR76b is not co-expressed with GR66a. This IR is thus more likely to be involved in the detection of appetitive compounds.

IR7a is expressed in three neurons in the TO that are not co-expressed with IR25a, and the projection pattern of these neurons into the adult SOG looks very similar to the one of GR66a (data not shown). This gene is thus likely to be expressed in neurons that sense aversive compounds. Larvae expressing reporters under the control of both IR7a-Gal4 and GR66a-LexA were thus analysed. Both receptors are co-expressed in one TO neuron (Figure 8F), and their SOG projections partially overlap (Figure 8J). This suggests that IR7a can be expressed in bitter-sensing neurons and may play a role in the detection of aversive tastants. I also tested co-expression

of two other cell-type specific IR drivers with GR66a-LexA. Both IR7e and IR60c were not co-expressed in the TO (Figure 8G-H) and their SOG projections did not overlap (Figure 8K-L). This suggests that these IRs function in sensing tastants that are not aversive.

The use of two IR-Gal4 lines together in the same animal also allows testing the colocalisation of two IRs. This was used to confirm that IR7a and IR7e are not co-expressed. Both lines usually mark two neurons that have their somas in the TOG (Figure 8M-N), although IR7e sometimes weakly labels one or two additional neurons (Figure 4). At least four neurons are labelled in larvae bearing both Gal4 lines (Figure 8O). This pattern was observed in several animals (Figure 8P), showing that these two receptors are not co-expressed.

Discussion

Non-olfactory IRs are candidate gustatory receptors

The results of this section provide a comprehensive picture of the expression of IRs in *Drosophila* larva. The fact that many IRs are expressed across gustatory organs supports the hypothesis that they function as gustatory receptors. Most IR reporters co-localise with the co-receptor IR25a, which suggests that the typical co-receptor/ligand-specific receptor assembly observed in olfactory IRs (Abuin et al., 2011) is also true in the gustatory system, although the other major antennal co-receptor IR8a is absent from larvae.

It is reasonable to speculate that IRs expressed at both life stages keep their function during metamorphosis and detect similar cues in adults than in larvae. Receptors expressed exclusively in adults are good candidates for being involved in adult-specific behaviours such as courtship or oviposition. Meanwhile, larval specific IRs are more likely to influence foraging or feeding behaviour. Finally, it is also possible that certain cues are detected by different systems in adults than in larvae. For example, sugar-sensing GRs are only expressed in adults, suggesting that other receptors, potentially IRs, fulfil this function in larvae.

The central projections of IR neurons to the main gustatory centre in the brain look very similar to the ones of GR neurons (Kwon et al., 2011). In contrast to adults, it is difficult to distinguish appetitive from aversive projections in larvae, and most differences observed are likely to be due to the organ that is innervated by each neuron.

The two most broadly expressed IRs do not co-localise with the bitter sensing GR66a (with one exception for IR25a). IRs that are expressed together with these co-receptors are thus not co-expressed in GR66a neurons. However, several IRs are not co-expressed with IR25a and at least one IR-Gal4 line (IR7a) overlaps with GR66a neurons. This suggests that in contrast to the olfactory system where OR and IR neurons are completely segregated, IRs and GRs can be co-expressed. Gustatory neurons are thus likely to use multiple receptors to perceive broader ranges of tastants that are eventually not discriminated from each other. Current models indicate that taste perception relies upon a limited number of “labeled line” sensory pathways tuned to specific classes of tastants. By

contrast, my results suggest a richer coding capacity in the *Drosophila* gustatory system than previously appreciated, and identify molecular candidates for novel classes of gustatory receptors.

Perspectives

One significant challenge that remains is the classification of the diverse IR and GR projection patterns. The availability of Gal4 lines for all IRs and GRs will probably allow more accurate mapping of gustatory projections, both in larvae and adults. The registration of projection patterns visualised in different animals onto a common template brain (Jefferis et al., 2007) would be a first step. Furthermore, comprehensive co-expression analysis is necessary to understand more precisely the molecular diversity of distinct taste neurons.

The accuracy of the expression data obtained with Gal4-lines relies on the faithfulness of the Gal4-lines themselves. This parameter is to take into consideration before drawing any definitive conclusion on IR expression. In order to confirm the expression of gustatory receptors, other methods should be envisaged (e.g. improved *in situ* RNA detection or single-cell transcriptomics (Miller et al., 2009; Nagoshi et al., 2010; Yang et al., 2005))

Material and Methods

IR Gal4 transgenes

Primers were designed to amplify putative promoter regions from Oregon-R *D. melanogaster* genomic DNA with flanking restriction sites. The

length of each IR promoter was arbitrarily fixed to 2.5 kb, but was sometimes reduced to the distance to the 3' end of the next gene upstream. Gel purified PCR products were cloned in the pGEM-T Easy vector (Promega), end-sequenced, and sub-cloned into the pGAL4-attB vector, comprising the GAL4 ORF-hsp70-3'UTR in the pattB vector (Bischof et al., 2007). These constructs were integrated into the defined attp2 landing site at cytological location 68A4 (Markstein et al., 2008), by standard transformation procedures (Genetic Services, Inc.). IR-Gal4 transgenic flies were double-balanced and crossed with flies bearing a UAS-mCD8:GFP (Lee and Luo, 2001) or UAS-mCD8 transgenes to reveal driver expression.

GR66a-LexA flies were obtained from Kristin Scott and crossed with UAS-rCD2:GFP flies to obtain w , P{GR66a-LexA}; P{UAS-rCD2:GFP}; TM2/TM6B flies that were crossed with w ; P{UAS-mCD8:GFP}; P{IRXX-Gal4} flies to perform co-expression studies.

Histology

Third instar larvae were placed in a Petri dish containing 1xPBS/0.1% Triton (P/T) and the anterior part containing head chemosensory organs was removed, or the abdomen was gutted and opened with forceps. Dissected tissues were placed in a 1.5 ml microcentrifuge tube and fixed in 4% PFA in 1xPBS for >1 hour at 4°C, washed 3 x 10 minutes in P/T, blocked for 1 hour in 5% heat-inactivated goat serum in P/T (P/T/S) and incubated 24 hours at 4°C with appropriate dilutions of mouse anti-GFP (Invitrogen), rat anti-mCD8 (Jackson ImmunoResearch) or rabbit anti-IR25a (Benton et al., 2009) antibodies in P/T/S. Tissues were washed and blocked

as above and incubated with Alexa488-anti mouse, Cy3-anti mouse, Alexa488-anti rabbit, Cy3-anti rabbit, Cy3 anti-rat or Cy5 anti-rat secondary antibodies (Milan Analytica AG), all diluted to 1:500 in P/T/S for at least 24 hours at 4°C. Samples were mounted on glass slides with 100 µl Vectashield (Vector Laboratories, Inc.). Images were collected with a Zeiss LSM 510 Meta upright confocal microscope (Zeiss, Oberkochen, Germany), using a Plan-APOCHROMAT 63x/1,40 Oil DIC objective. Images were further processed with NIH ImageJ.

Dissections and stainings of larval CNS were performed as described above, with the same antibodies, in addition to nc82/Brp (Wagh et al., 2006) that was used to mark neuropil. All these were diluted to 1:500 in P/T/S, and nc82/Brp, which was diluted to 1:10.

Chapter 4: Behavioural characterisation of IR25a mutants

Introduction

Several classes of tastants are detected by the gustatory system of *Drosophila*. Amongst these, the perception of sugars and bitter compounds are the most widely studied, but it has also been shown that flies can taste salt (Liu et al., 2003a), carbonation (Fischler et al., 2007), pheromones (Bray and Amrein, 2003; Miyamoto and Amrein, 2008), water (Cameron et al., 2010) and amino acids (Toshima and Tanimura, 2012).

In adult flies, Gustatory Receptors (GRs) have been identified that detect sugars, bitter compounds and pheromones (Montell, 2009), and members of the *pickpocket* family of DEG/ENaC channels detect salts and water (Cameron et al., 2010; Liu et al., 2003a). However, no receptor for other classes of tastants has been identified. Because IRs are expressed in gustatory neurons, they are good candidates for fulfilling these functions.

The expression of GRs has been well described in larvae (Colomb et al., 2007; Kwon et al., 2011), and receptors for bitter compounds and pheromones are present at both life stages and are believed to keep the same function in larvae and in adults. However, amongst GRs that are involved in sugar perception, only GR43a is expressed at the larval stage (Kwon et al., 2011; Miyamoto et al., 2012), but GR5a and receptors from the GR64a-f cluster are absent. Two hypotheses can explain this: either the Gal4 lines made from the promoters of these receptors are missing key

regulatory elements that prevent them from being expressed in larvae, or these receptors are indeed absent, which would imply different receptors (such as IRs) fulfil this function.

A behavioural assay for testing gustatory preference

To assess the role of IRs in gustation, I first established a taste-choice preference assay, based upon a previous design (Heimbeck et al., 1999). In this assay, 20-30 larvae are placed into a four-compartment Petri dish filled with agarose, in which two opposite quadrants contain a defined concentration of a specific tastant (Figure 9A). Animals are starved for one hour and then allowed to wander freely on the plate. *Wild type* larvae will naturally spread with a higher density on the compartments containing the most attractive or less repulsive tastant. Larvae on each compartment are counted after 20 minutes, and a response index that illustrates the preference of larvae for one substance compared to the other or to agarose alone is calculated (Figure 9A).

IR25a mutant: a mutant lacking most IR functions?

IR25a is broadly expressed in all larval gustatory organs and may act as a co-receptor, similar to its function in the olfactory system (Abuin et al., 2011). I hypothesised that the disruption of *IR25a* would inactivate all IR-related taste perception. The strategy here was to screen *IR25a* null mutants (*IR25a²*) (Benton et al., 2009) for behavioural defects towards several classes of tastants. Tastants that are no longer preferred by these

animals are likely to be those detected by IRs. Rescue animals were generated that carry a bacterial artificial chromosome construct containing a 20kb genomic region spanning the *IR25a* locus that is integrated on the *IR25a* mutant chromosome. A UAS-*IR25a* construct could not be used here because no suitable driver was available, as *IR25a-Gal4* does not faithfully recapitulate endogenous *IR25a* expression. Immunohistochemistry with anti-*IR25a* antibody revealed that *IR25a* is not expressed in any chemosensory organ in *IR25a*² larvae but present in *wild type* and *rescue* larvae (Figure 9B).

***IR25a* is not involved in the detection of sugars and bitter compounds**

Bitter compounds are detected by GRs (Lee et al., 2009; Moon et al., 2006; Weiss et al., 2011), and disruption of IRs should not influence preference behaviour towards these compounds. To check this, a taste-choice preference assay where larvae were given the choice between sucrose mixed with a bitter compound or sucrose alone was performed. *Wild type* larvae always avoided the quadrant containing the bitter alkaloids quinine, caffeine, strychnine, lobeline or berberine (Figure 9C). Compared to these, *IR25a*² larvae showed no significant behavioural defect. Surprisingly however, there was a tendency for *rescue* larvae to be less repulsed by these compounds (Figure 9C), which was significant for caffeine, lobeline and berberine. This phenotype may be due to the presence of other genes in the *IR25a*-BAC construct. As expected, these results suggest that *IR25a* is not involved in tasting bitter compounds.

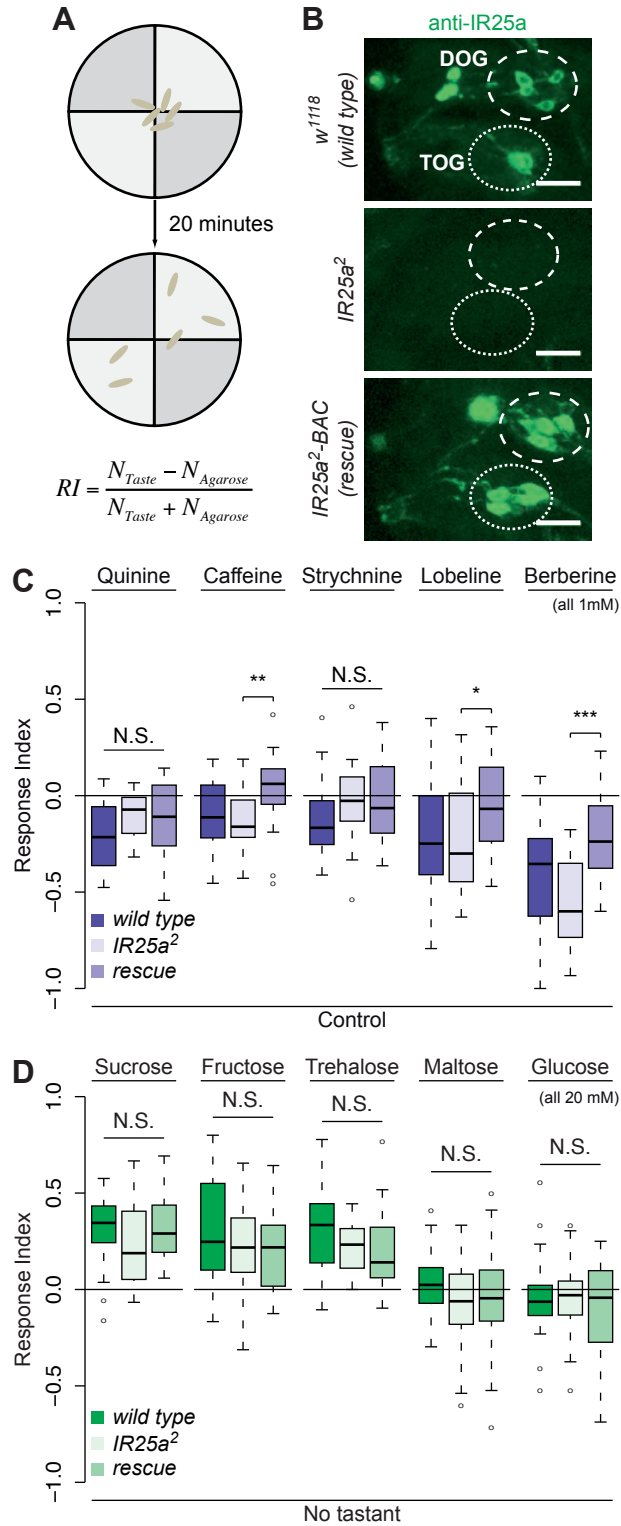


Figure 9 | Preference for sugars and bitter compounds is not reduced in *IR25a* mutants. (A) Schematic representation of the larval taste-choice assay and the formula used to calculate the Response Index. (B) *IR25a²* totally suppresses the expression of IR25a protein in larval chemosensory organs. The expression is rescued by the insertion of the 32C20 BAC construct containing IR25a. (C) Boxplots showing the response indexes obtained with *wild type*, *IR25a²* and *rescue* larvae in a taste-choice assay between 5mM Sucrose (Control) or 5mM sucrose with 1mM bitter compound. *rescue* larvae are less repulsed by some of these compounds. N=20 for each plot. (D) Boxplots showing the response indexes obtained with *wild type*, *IR25a²* and *rescue* larvae towards various sugars (20mM each). No significant difference is observed between the three genotypes. N=20 for each plot. *: p<0.05, **: p<0.01, ***: p<0.001 in Wilcoxon signed rank test or Kruskal-Wallis rank sum test.

In order to test, whether IR25a is involved in the detection of sugars, the preference of *wild type*, *IR25a²* and *rescue* larvae to 20 mM of five common sugars (sucrose, fructose, trehalose, maltose and glucose) was tested. Larvae of all three genotypes were attracted to sucrose, fructose and trehalose, but not to maltose or glucose. These two sugars may need to be present in a higher concentration to induce preference behaviour. For all sugars, no difference was observed between these three genotypes (Figure 9D), showing that IR25a has no function in behavioural sugar preference.

Reduced preference for acidic media in IR25a mutants

In the olfactory system, some IRs are involved in detection of carboxylic acids (Ai et al., 2010). Whereas the detection of carboxylic acids seems to be restricted to the olfactory system, I hypothesised that IRs in the gustatory system could sense low pH. I therefore assessed the preference of larvae towards sulphuric, hydrobromic and hydrochloric acids. *Wild type* larvae were strongly attracted to each acid at a concentration of 5mM, suggesting that larvae generally prefer acidic environments (Figure 10A). Preference behaviour was also tested with different concentrations of HCl and showed to be concentration-dependent (Figure 10B). Interestingly, *IR25a²* mutants showed significantly decreased preference behaviour towards all three acids, suggesting that IR25a plays a role in sensing low pH (Figure 10A). This decrease was also true for other concentrations of HCl (Figure 10B). Except for sulphuric acid, the rescue construct was able to entirely rescue

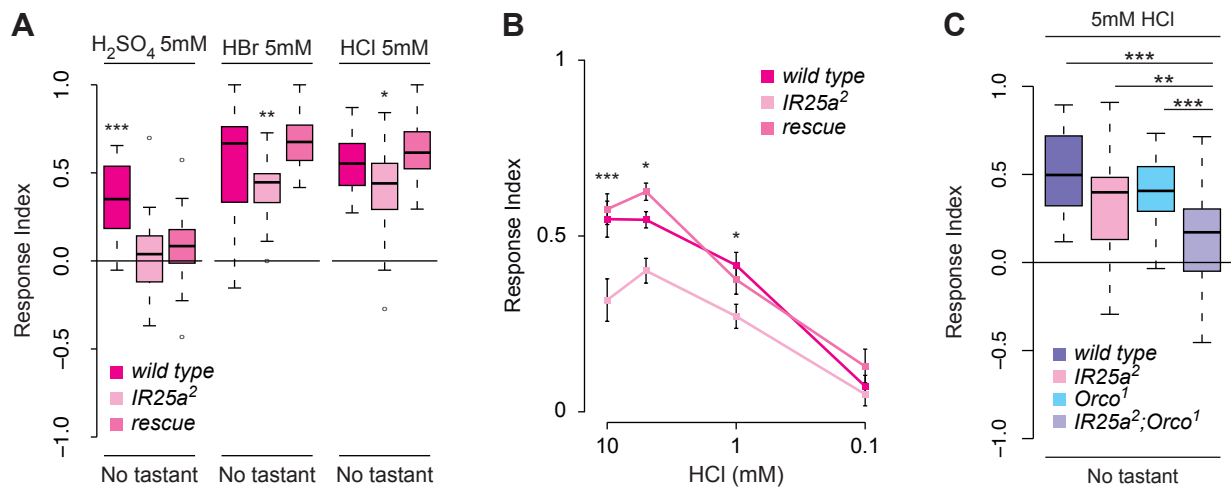


Figure 10 | IR25a is involved in preference behaviour towards low pH. (A) Boxplots showing the response indexes obtained with *wild type*, *IR25a²* and *rescue* larvae towards various strong acids. *IR25a²* larvae show a decreased preference for these acids. N=20 for each plot. (B) Preference behaviour towards HCl is concentration-dependent. At higher concentrations, the preference of *IR25a²* larvae is reduced. N=20 for each plot, error bars show the standard error of the mean. (C) Boxplots showing the response indexes obtained with *wild type*, *IR25a²*, *Orco¹* and *IR25a²;Orco¹* double mutants larvae towards HCl. The *IR25a²;Orco¹* double mutation strongly reduces the preference behaviour. N=20 for each plot. *: p<0.05, **: p<0.01, ***: p<0.001 in Wilcoxon signed rank test or Kruskal-Wallis rank sum test.

the loss of acid preference in *IR25a*² larvae. These results suggest that IR25a plays a role in sensing low pH.

Although significantly reduced, the response index of *IR25a*² larvae towards acids remains positive, which suggests that additional mechanisms are involved in sensing low pH. All acids tested are volatile, and to assess the role of the olfactory system, double mutants lacking both IR25a and the olfactory co-receptor Orco were used in a taste-choice assay (acid-sensing olfactory IRs are not expressed at the larval stage (Figure 7)). The preference of these mutants towards acids was significantly lower than the one of *IR25a*² or *Orco*¹ single mutants and of *wild type* larvae (Figure 10C). This demonstrates that olfaction plays contributes to acid sensing and that both ORs and IRs are involved in sensing low pH. This suggests that IR25a together with other IRs can form a channel that responds to low pH. However, because IR25a is expressed in both olfactory and gustatory organs, it is difficult to define the precise role of the gustatory system in acid sensing based on this behavioural assay. Furthermore, the involvement of two chemosensory systems prevents from further investigating this question with behavioural assays only.

IR25a plays a minor role in modulating amino acid sensing

Amino acids are interesting candidates for being detected by IRs in the gustatory system. In the olfactory system, the main IR ligands belong to the amine and acid chemical classes (Silbering et al., 2011). In addition, IRs

derive from iGluRs (Croset et al., 2010), which bind the amino acid glutamate, and in some case serine (REF). Previous research has demonstrated that substantial amino acid deprivation in adult flies leads to enhanced preference towards amino acids, in particular cysteine, histidine, leucine, phenylalanine, threonine and tyrosine (Toshima and Tanimura, 2012). In order to test, whether larvae are also able to taste amino acids, the preference of larvae for 21 L-amino acids was assessed in a taste-choice assay. Eight of these induced significant preference behaviour. Whereas threonine, arginine, asparagine, phenylalanine, cysteine, glutamic acid and aspartic acid are perceived as appetitive, leucine induces repulsive behaviour (Figure 11A). This experiment shows that these amino acids trigger the activation of chemosensory pathways that induce a behavioural response. It is possible that other amino acids are detected but do not induce any behaviour in these conditions. Although no correlation emerges as to particular chemical properties of these appetitive amino acids, the general pattern suggests that polar and charged amino acids are more likely to trigger preference behaviour. Preference was also not biased towards essential amino acids, as four of these appetitive amino acids are non-essential.

IR25a-defective larvae were tested for their preference towards some amino acids. First, their response to leucine was tested. Surprisingly, these animals no longer avoided this compound. However, *wild type* larvae that were tested at the same time also showed no avoidance, questioning the reproducibility of this behaviour (Figure 11B). Although the preference

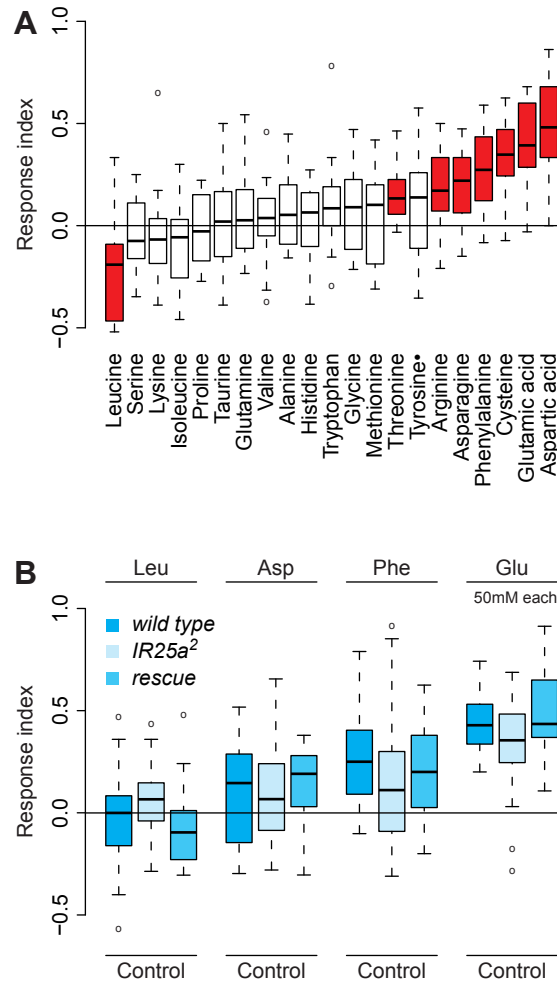


Figure 11 | Preference behaviour towards amino acids. (A) Boxplots showing the response index of *wild type* larvae towards 21 amino acids in a taste-choice assay. 50 mM of each amino acid was used except for tyrosine (2 mM). The values in red are significantly different from 0 ($p < 0.05$ in Wilcoxon signed rank test). $N = 20$ for each plot. (B) Boxplots showing the response indexes obtained with *wild type*, *IR25a²* and *rescue* larvae towards various strong acids. There is no significant difference ($p > 0.05$ in Wilcoxon signed rank test) between the three genotypes. $N > 20$ for each plot.

towards asparagine, phenylalanine and glutamic acid was slightly reduced, no significant difference was observed compared to *wild type* or *rescue* larvae (Figure 11B). IR25a may thus play a minor role in the detection of these amino acids, although behavioural experiments are probably not sensitive enough to attest it as this role might be only modulatory. The fact that *IR25a*² larvae remain significantly attracted by these compounds suggests that mechanisms independent of IR25a are major players in amino acid detection.

Discussion

Together these results show that IR25a has no or little function in gustatory behaviour towards several main classes of tastants. Two hypotheses can explain this lack of function of one of the most broadly expressed receptor in the gustatory system. IR25a is a co-receptor in the adult antenna, and it was hypothesised that it has the same function in the gustatory system. Deletion of IR25a was thus supposed to inactivate all IRs, or at least those that are dependent on this co-receptor. However, no evidence exists that can confirm this hypothesis, and this cannot be provided until the function of gustatory IRs is established. Nevertheless, the fact that IR25a deletion mutants show reduced preference for acidic media as well as slightly reduced preference for amino acids suggests that instead of functioning as a co-receptor in the same way than in the antenna, IR25a is more likely to only modulate gustatory response. This is however impossible to test with behavioural assays, in which natural variations

between individuals are likely to mask subtle modulatory differences. Hence, other IR co-receptors may be present in the gustatory system that allow the maintenance of IR function in the absence of IR25a. One of these could be IR76b that is also an olfactory co-receptor and is also broadly expressed in gustatory tissues.

The fact that IR25a deletion does not abolish any behavioural function also suggests that the tastants that are actually detected by IR25a-dependent mechanisms were not tested. Indeed, only a reduced panel of putative gustatory ligands was used here, and the hypothesis that IR25a is involved in detection of other types of chemicals cannot be excluded. However, the behavioural paradigm that was used here is not ideal for screening hundreds of putative tastants.

Material and methods

Fly strains

All flies were maintained on standard cornmeal-agar medium under a 12 hour light : 12 hour dark cycle at 25°C. *Wild* type flies were w^{1118} , IR25a² flies are from (Benton et al., 2009). Rescue flies (IR25a-BAC) were obtained by recombination of IR25a2 flies with flies having CH322-32C20 BAC inserted in attP16 (53C4). *Orco*¹ mutants were obtained from the Vosshall Lab *via* Richard Benton.

Larval taste-choice assays

0.8% agarose solution was boiled and cooled down to $<50^{\circ}\text{C}$ before adding selected amount of tastant when necessary. All chemicals tested were from Sigma-Aldrich or Applichem. 12.5 ml of appropriate solution were poured in each of the four compartments of Star™ Dish 90x15 petri dishes left for polymerisation for >1 hour. Larvae were raised in tubes containing standard fly food at 22°C . They were collected in 15% sucrose solution, rinsed with tap water and starved for approximately 1 hour on 0.8% agarose gel. Groups of 30 larvae of the same genotype were placed at the centre of each Petri dish under red light. Larvae were counted manually on each quadrant after 20 minutes. Statistics were performed with R (R-Development-Core-Team, 2010).

Chapter 5: Identification of IR neurons that detect sugars and amino acids

Introduction

The experiments described above have demonstrated that deletion of IR25a does not significantly alter preference behaviour towards most tested compounds. One major issue that emerged is that the sensitivity of behavioural assays is too low to bring out mechanisms that could be subtler than expected. Hence, it is necessary to choose another strategy that allows direct measurement of the physiological effects of tastant recognition by IR-expressing neurons.

Two methods for measuring neuronal activity in vivo

Two methods are generally used to measure the physiological response of a neuron to a stimulus in *Drosophila*. Electrophysiological recordings are used for measuring the electrical activity of a whole organ, for example the antenna (Alcorta, 1991; Park et al., 2002), of single olfactory or gustatory sensilla (Benton and Dahanukar, 2011a, b), and even of individual neurons in the brain (Ruta et al., 2010). Although it is a rather precise and straightforward method to measure the activity of olfactory sensory neurons, it has the disadvantage of being difficult to perform on larvae because their chemosensory sensilla are much smaller than the adult ones and grouped in ganglia (Hoare et al., 2011).

The second method takes advantage of the Gal4-UAS system to drive the expression of genetically-encoded calcium reporter GCaMP3, a calmodulin subunit coupled to two halves of GFP into the cells of interest (Tian et al., 2009). Action potentials in neurons induce rapid increases in intracellular calcium concentration, which is bound by calmodulin in a concentration-dependent fashion. This induces conformational changes that increase the GFP fluorescent signal by bringing its two halves closer to each other. These variations in fluorescence intensity can be visualised and quantified (Mank and Griesbeck, 2008; McCombs and Palmer, 2008; Tian et al., 2009). Calcium imaging has been used on adult flies for various purposes. In olfactory studies, the activation of sensory neurons is usually recorded in the antennal lobe, where neurons expressing the same olfactory receptors project to distinct glomeruli (Fiala et al., 2002; Silbering et al., 2012; Wang et al., 2003). Similar, though less extensive, studies have been performed on projections of gustatory neurons in the SOG (Cameron et al., 2010; Fischler et al., 2007; Marella et al., 2006).

In larvae, calcium imaging has rarely been used outside of a few studies on the motor system (Guerrero et al., 2005; Karunanithi et al., 1997) or injury (Ghannad-Rezaie et al., 2012), although one study reported measurement of odour-evoked signals in larval OSNs and PNs (Asahina et al., 2009). In this chapter, I describe my establishment of an imaging preparation in the *Drosophila* gustatory system and use of this to demonstrate taste-evoked response in IR-expressing neurons.

Establishment of calcium imaging methods for adults and larvae

IR76b neurons respond to sucrose in adults

In order to test whether calcium imaging can be used on IR neurons in the gustatory system, a method was adapted with the help of Ana F. Silbering, a post-doc from the Benton group. In this method, flies are held in a collar and their antennal plate is dissected out in order to directly visualise the SOG (Figure 12A, D and see Materials and Methods for a more precise description of the preparation). Preliminary assays were performed on adult flies expressing GCaMP3 under the control of the IR76b-Gal4 line. IR76b functions as a co-receptor in the olfactory system (Abuin et al., 2011) and its broad expression throughout both adult and larval gustatory organs (Figure 7) suggests that it plays a similar role in taste perception. Because the projection pattern of IR76b neurons resembles the one of the GR5a, which marks sweet-sensing neurons (Figure 12C) (Marella et al., 2006; Thorne et al., 2004; Wang et al., 2004), the response of IR76b neurons to sucrose was tested. When sucrose was presented to the proboscis, a strong increase in fluorescence was observed from these neurons (Figure 12E), showing that sucrose is able to activate them. However, this activation is likely to be due to sugar-sensing GRs that might be co-expressed in these cells. Nevertheless, this experiment confirms that IR76b neurons drive gustatory signals and also suggests that IR76b is involved in the perception of appetitive cues.

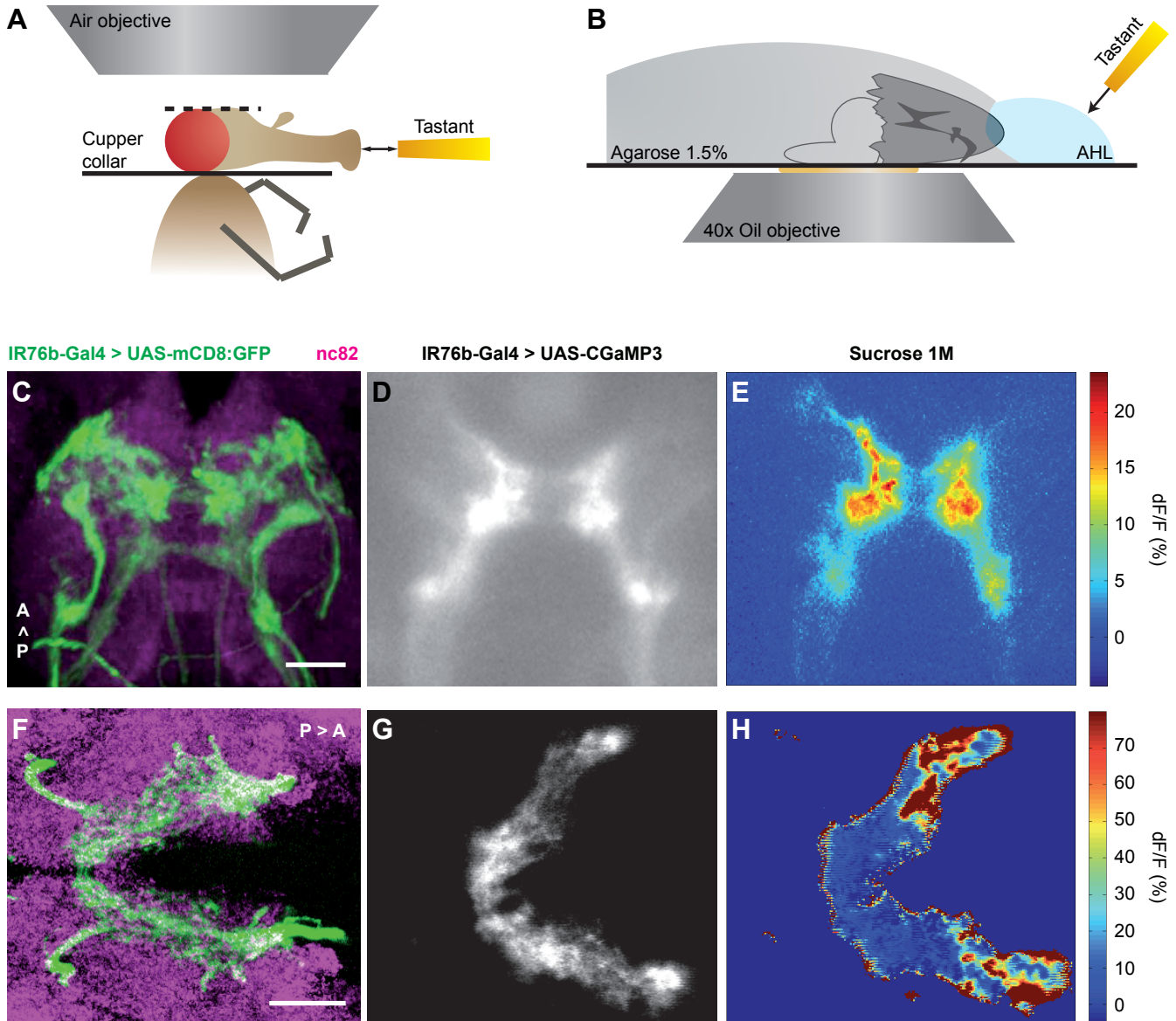


Figure 12 | Calcium imaging in adult and larval gustatory centres. (A) Scheme of the adult calcium imaging assay. The fly's head is held in a copper collar and fixed with n-eicosane. The antennal plate is dissected out in order to be able to image the SOG. Tastants are applied directly to the proboscis. (B) Scheme of the larval calcium imaging assay. A dissected larva brain attached to the head is embedded in 1.5% agarose on a glass coverslide. Agarose is removed from the tip of the head and replaced by a drop of adult-hemolymph-like ringer. Tastants are applied directly on that drop and the fluorescence is recorded on an inverted microscope. (C) Confocal picture of the projections of IR76b-Gal4 labelled neurons to the adult SOG. (Image by A.K. Sivasubramaniam). (D) Raw fluorescence recorded from the projections of IR76b>GCaMP3 neurons in the adult SOG. (E) Activation of the neurons shown in D by 1M sucrose. (F) Confocal picture of the projections of IR76b-Gal4 labelled neurons to the larval SOG. (G) Raw fluorescence recorded from the projections of IR76b>GCaMP3 neurons in the larval SOG. (H) Activation of the neurons shown in G by 1M sucrose.

Calcium imaging on larval SOG

This method was then adapted to larvae (Figure 12F), in collaboration with Lena van Giesen, Abud Farca and Simon Sprecher from the University of Fribourg. The major difficulty when imaging from living larvae is to minimise the movements due to their strong abdominal muscles and their cylindrical-like body. Larvae are thus dissected to isolate their mouthparts and brain from the rest of their body (Figure 12B). Dissected larvae are immobilised on glass coverslips into agarose gel that protects them from drying up. The tip of the mouthparts is kept outside of the gel and covered with a saline solution (Figure 12B). A drop of tastant is added directly to the solution while registering time-lapse images of the SOG on an inverted microscope (Figure 12B, G). While adding 1M sucrose solution, a strong increase in fluorescence was measured from IR76b neurons (Figure 12H), showing that these neurons respond to this tastant. To our knowledge, this is the first time that calcium imaging was successfully used to measure the physiological response of gustatory neurons in *Drosophila* larvae. In addition, while larvae show robust behavioural responses to sugars, none of the known sugar receptors appear to be expressed there. This data therefore reveals the existence of specific sugar-sensing neurons.

Sugar-sensing IR neurons

A subset of IR76b neurons selectively respond to sucrose and fructose

In order to identify sugars that activate IR76b neurons, their response to five of those (sucrose, fructose, glucose, maltose and trehalose) was tested by calcium imaging. Surprisingly, although a strong increase in fluorescence was observed in response to sucrose and fructose, no activation was visible for the three other sugars (Figure 13A). In order to test, whether the response observed is concentration-dependent, sucrose solutions of different molarities were tested. Whereas high concentrations (> 100 mM) induced activation of IR76b neurons, no response was observed for lower concentrations (Figure 13B), suggesting that these neurons only detect rather high concentrations of sugars, and that different mechanisms are involved in the detection of lower concentrations. This concentration-dependent activation was not observed with trehalose, confirming that this sugar is not detected by IR76b neurons (Figure 13B).

To test, whether IR76b neurons are necessary for sucrose and fructose detection, the light chain of tetanus toxin (TNT) (Sweeney et al., 1995) was used to selectively inactivate these neurons. TNT blocks neuroexocytosis by cleaving synaptobrevin (Schiavo et al., 1992). Behavioural responses of larvae bearing TNT or its mutationally inactivated form (IMPTNT) was analysed towards sugars on a taste-choice preference assay. Whereas IR76b>IMPTNT larvae prefer sucrose in a concentration-dependent fashion, IR76b>TNT larvae show a reduced preference (Figure 13C) for

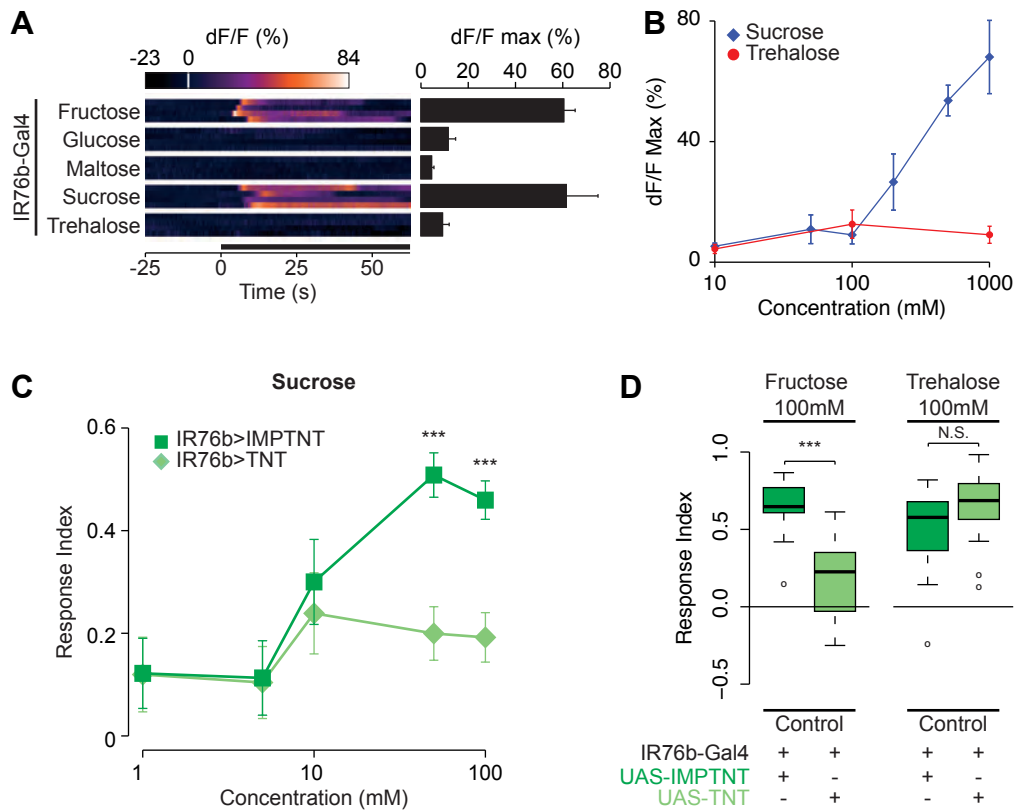


Figure 13 | IR76b neurons respond to sucrose and fructose. (A) Left: heatmaps showing the increase of fluorescence measured from IR76b neurons activated by different sugars. Right: plots showing the maximum dF/F measured from these neurons in response to each sugar. $N \geq 4$, error bars show the standard error of the mean. (B) Maximum dF/F measured from IR76b neurons activated by different concentrations of sucrose (blue) or trehalose (red). $N \geq 4$ for each point, error bars show the standard error of the mean. (C) Preference behaviour (see Figure 8A) of larvae expressing the light chain of tetanus toxin (TNT; light green) or its inactive version (IMPTNT; dark green) in IR76b neurons towards different concentrations of sucrose. ***: $p < 0.001$ in Wilcoxon signed rank test. (D) Preference behaviour towards fructose (left) and trehalose (right) of larvae expressing TNT (light green) or IMPTNT (dark green) in IR76b neurons. TNT reduces preference to fructose but not to trehalose. ***: $p < 0.001$ in Wilcoxon signed rank test.

higher concentrations of sucrose (≥ 50 mM), but behave normally at lower concentrations. This demonstrates the necessity of IR76b neurons for sensing high concentrations of sucrose, but confirms the hypothesis that mechanisms independent of IR76b neurons are involved in sensing lower concentrations. A similar reduction in preference was observed with fructose, but not with trehalose (Figure 13D), confirming the fact that IR76b neurons do not detect this sugar. These results indicate that IR76b neurons are necessary for proper preference behaviour towards sucrose and fructose, but not towards other sugars.

The fact that IR76b neurons respond to sucrose and fructose but not to the other sugars tested also suggests that there are at least two pathways underlying sugar detection. This might allow larvae to discriminate between different classes of sugar.

IR76b neurons innervate a broad area of the SOG (Figures 6, 14A), and the increase in fluorescence observed when activating IR76b neurons with sucrose is not uniform and seems to be higher in the anterior than in the posterior part of the projection (Figure 14B). In order to identify, which region of the IR76b projection responds to sugars, variations in fluorescence were measured from these two regions. Sucrose strongly activates the anterior region, whereas no activation is visible in the posterior one (Figure 14C-D). The same pattern was observed for fructose (Figure 14D). This confirms the hypothesis that only a subset of IR76b neurons is sensitive to sugars. Because these neurons are the ones projecting more

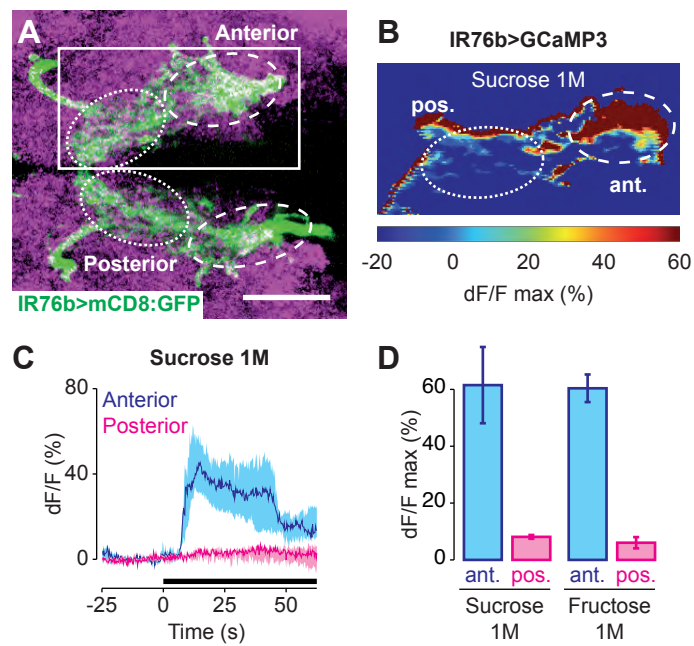


Figure 14 | Only the anterior region of IR76b projection is activated by sugars. (A) Projections of IR76b neurons to the SOG. Anterior and posterior regions are indicated with long and short dashes, respectively. (B) Activation of IR76b neurons by sucrose increases calcium levels mostly in the head region. The image only shows the right side of the SOG (white frame in A). (C) Activation of the anterior and posterior regions by sucrose. The darker lines show the median of ≥ 9 replicates. The lighter zones define the 25th and 75th percentiles for each time point. (D) Plots showing the maximum dF/F for the anterior and posterior regions activated by sucrose or fructose. $N \geq 4$, error bars show the standard error of the mean.

frontally into the SOG, they are more likely to innervate internal organs than peripheral ones, although this remains to be formally tested.

Identification of IR7e neurons that are sucrose-sensing IR neurons

The fact that IR76b neurons respond to some sugars implies that IRs are interesting candidates for fulfilling this function. However, IR76b is likely to be a co-receptor and individual IR neurons that detect sugars remain to be determined. To identify these neurons, all IR-Gal4 lines expressed in larval gustatory organs were used for calcium imaging assays on their central projections. Gal4-driven GCaMP3 signals were clearly visible for most lines, except five of them (IR7d, IR48b, IR67a, IR94a, IR94b) where the intensity of fluorescence did not allow reliable recordings (data not shown). A mix of five sugars (sucrose, fructose, trehalose, glucose and maltose; 200 mM each) was tested. Most IR neurons showed no activation to this mixture (Figure 15A). However, IR7e neurons were strongly activated by the sugar solution (Figure 15A). IR7e-Gal4 drives expression to two neurons of the terminal organ (TO) (Figure 15B), although two additional TO neurons are occasionally weakly labelled. Next, the response of IR7e neurons was tested towards 1M concentrations of each of these five sugars individually. Interestingly, only sucrose was able activate these neurons while no significant change in fluorescence was measured for the four other sugars (Figure 15C). Thus, these experiments revealed the existence of neurons that are sensitive to sugars in the larval terminal organ, in which

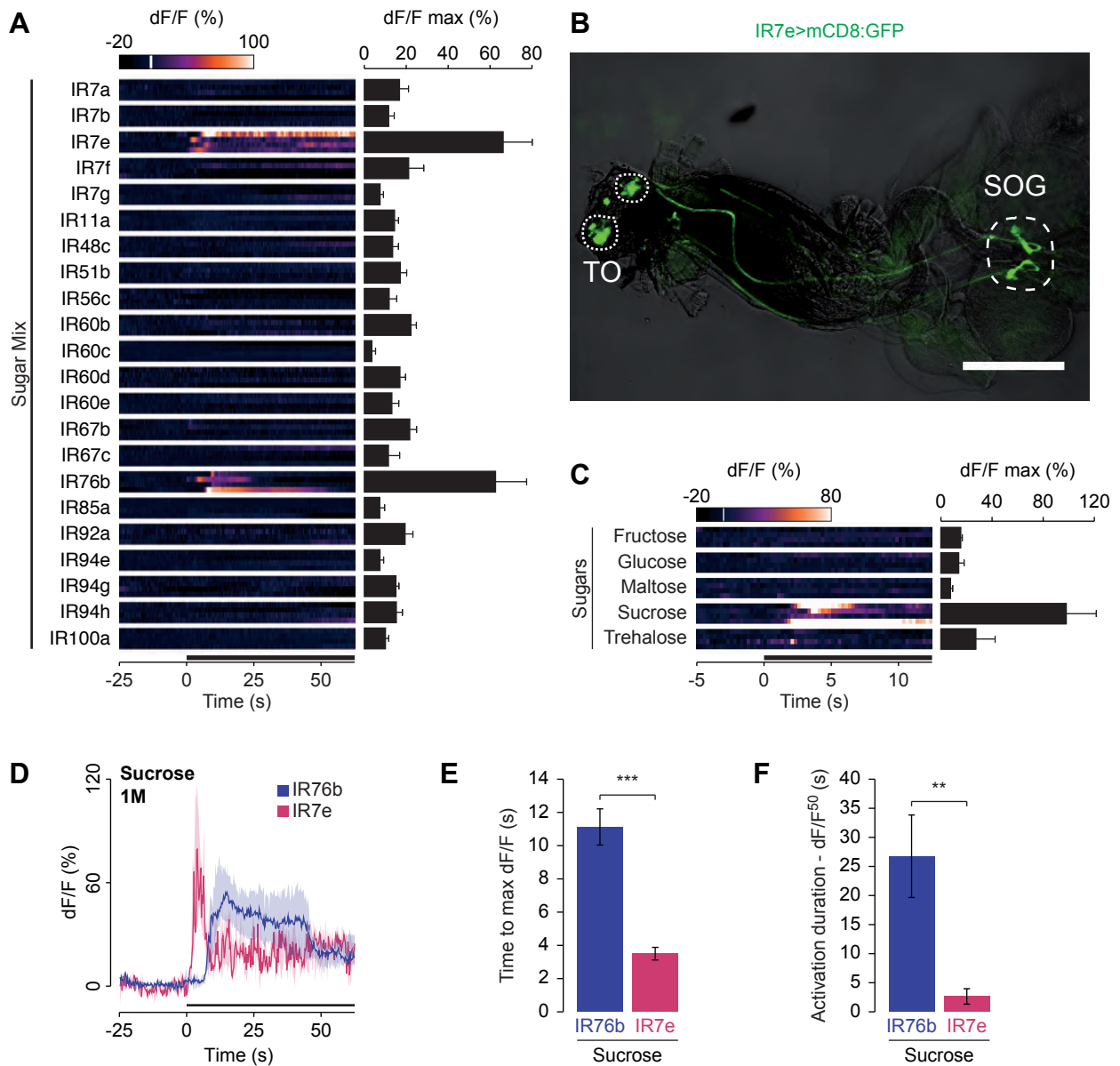


Figure 15 | A screen for sugar-sensing IR neurons (A) Left: heatmaps showing the response of different IR neurons to a mix of sucrose, fructose, trehalose, maltose and glucose (200mM each). Right: plots showing the maximum dF/F of these neurons in response to the sugar mix. Only IR76b and IR7e neurons respond to these sugars. $N \geq 4$, error bars show the standard error of the mean. (B) Immunofluorescence on larvae expressing mCD8:GFP reporter in IR7e neurons. IR7e is expressed in neurons from the TO that project to the SOG. Scale bar is 100 μm . (C) Left: heatmaps showing the increase of fluorescence measured from IR7e neurons towards different sugars. Right: plots showing the maximum dF/F of IR7e neurons in response to the sugars shown on the left. Only sucrose induces a significant increase in fluorescence in these neurons. $N \geq 4$, error bars show the standard error of the mean. (D) Fluorescence measured over time for IR76b and IR7e neurons activated with 1M sucrose. (E) Average time for reaching maximum dF/F after 1M sucrose presentation in IR76b and IR7e neurons. (F) Activation duration measured as the time required for fluorescence to decrease to half its value after activation peak (dF/F^{50}). Error bars show the standard error of the mean. ***: $p < 0.001$ and **: $p < 0.01$ in Student's t-test.

IR7e is a candidate sucrose receptor. The fact that these neurons are not expressing the bitter sensing GR66a (Figure 8G) confirms their role in sensing appetitive tastants.

IR76b and IR7e neurons define two different sugar-sensing pathways

Because no antibody is available against IR76b and because IR76b-Gal4 labels too many neurons to use two Gal4 lines in parallel to assess whether these two receptors are co-expressed, it was not really possible to directly test, whether IR7e neurons co-localise with IR76b ones. Nevertheless, the response patterns of IR7e and IR76b neurons to sucrose were compared, in order to identify putative differences between these lines that could imply that these are not the same neurons (Figure 15D). Indeed, temporal response differences between neurons were observed in the olfactory system (Getahun et al., 2012) and are likely to happen in the gustatory system as well. First, the time to reach the maximum dF/F after sucrose activation was measured. This showed that IR76b neurons need much more time to reach this maximum than IR7e neurons (Figure 15E). Second, after activation, the time for fluorescence to decrease to 50% of the peak value was also measured. This showed that IR76b neurons remain activated approximately ten times longer than IR7e neuron (Figure 15F). These results highlight important timing differences in the activation of IR76b and IR7e neurons. These could be due to dissimilar properties between the receptors involved in sucrose response in these neurons, but also to the fact that tastants probably need more time to reach the IR76b

neurons from the DPS that are likely to be involved in this response. This demonstrates that IR7e- and IR76b-Gal4 lines do not label the same neurons, suggesting that these receptors are not co-expressed. Importantly, the comparison of these two Gal4 lines allowed highlighting a novel phenomenon, in which the same tastant is being sensed by two independent mechanisms of dissimilar kinetics.

Amino acids activate IR neurons

IR76b neurons respond to amino acids

As mentioned in chapter 4, amino acids are interesting candidate IR ligands because IRs derive from amino acid (glutamate) receptors and because olfactory IRs mostly sense acids and amines. In order to test whether amino acids activate IR neurons, calcium imaging was performed with 50 mM dilutions of individual amino acids on IR76b neuronal projections in the SOG. Amongst the 21 amino acids tested, 9 induced a strong increase in fluorescence: alanine, arginine, aspartic acid, cysteine, glutamine, glycine, methionine, serine and threonine. Other amino acids did not induce any measurable response in these neurons (Figure 16A).

Interestingly, by mapping which area of the IR76b projection responds to glutamine, an area in the lateral and posterior region was identified (Figure 16B-C), suggesting that these neurons come from the TO. The same area was activated by other amino acids (data not shown). This shows that the neurons activated by amino acids are different from those activated by sugars.

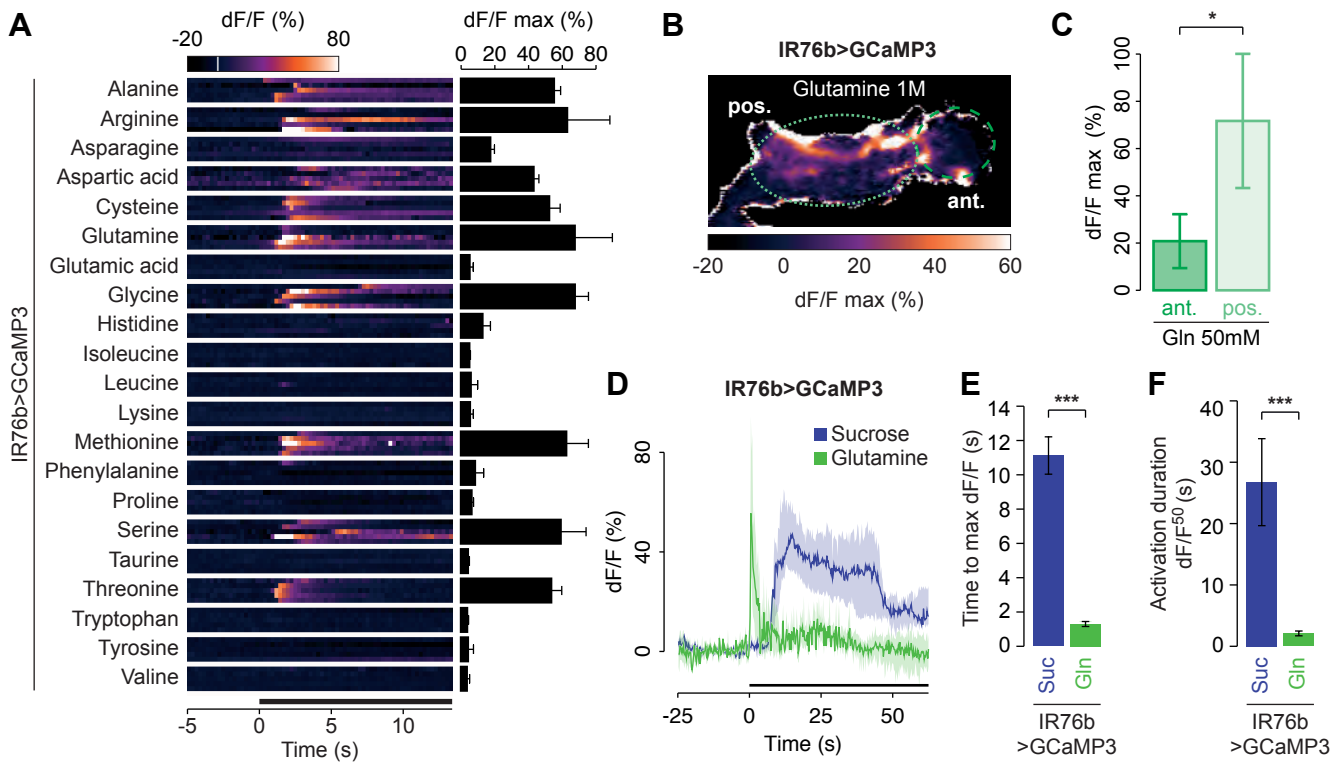


Figure 16 | Amino acid-dependent activation of IR76b neurons. (A) Left: heatmaps showing the response of IR76b neurons to individual amino acids (50mM each). Right: plots showing the maximum dF/F of these neurons in response to amino acids. Alanine, arginine, aspartic acid, cysteine, glutamine, glycine, methionine, serine and threonine activate IR76b neurons. $N \geq 5$, error bars show the standard error of the mean. (B) Activation of IR76b neurons by glutamine increases calcium levels mostly in the posterior region of the projection. (C) Plots showing the maximum dF/F for the anterior and posterior regions of IR76b neurons activated by glutamine. $N \geq 5$. (D) Fluorescence measured over time for IR76b neurons activated with 1M sucrose (blue) or 50 mM glutamine (green). (E) Average time for reaching maximum dF/F after 1M sucrose or 50 mM glutamine presentation in IR76b neurons. (F) Activation duration measured as the time required for fluorescence to decrease to half its value after activation peak (dF/F^{50}). Error bars show the standard error of the mean. ***: $p < 0.001$, **: $p < 0.01$ and *: $p < 0.05$ in Student's t-test.

The timing of the response of IR76b neurons to sucrose and glutamine was also compared. As shown earlier, sugar-sensing IR76b neurons respond rather slowly. In contrast, glutamine induces a rapid increase in fluorescence in IR76b neurons (Figure 16D-E) that also decreases very rapidly (Figure 16F), while sucrose-sensing neurons remain activated for a longer time. Similar fast activation patterns were observed with other amino acids than glutamine (data not shown). This confirms that these two classes of compounds are sensed by separate pathways within the larger IR76b population.

IR60c neurons are activated by the same amino acids as IR76b neurons

In order to identify individual IR neurons that respond to amino acids, the response to a 200 mM glutamine solution was tested in neurons labelled by eight IR-Gal4 lines that drive expression to the TO. A higher concentration was used here compared to the previous experiment to avoid missing potential responses from neurons that would detect only higher concentrations of amino acid. As expected, IR76b neurons are activated by this stimulus (Figure 17A). In addition, IR60c neurons, but no other IR neuron, also responded to glutamine (Figure 17A,C).

IR60c-Gal4 labels two neurons in the TO and one in the DPS (Figure 17B), and although the rapid increase in fluorescence observed while applying glutamine to the sample suggests that TO neurons are the ones being activated, it is difficult to formally test, whether the IR60c DPS neuron also plays a role in this response.

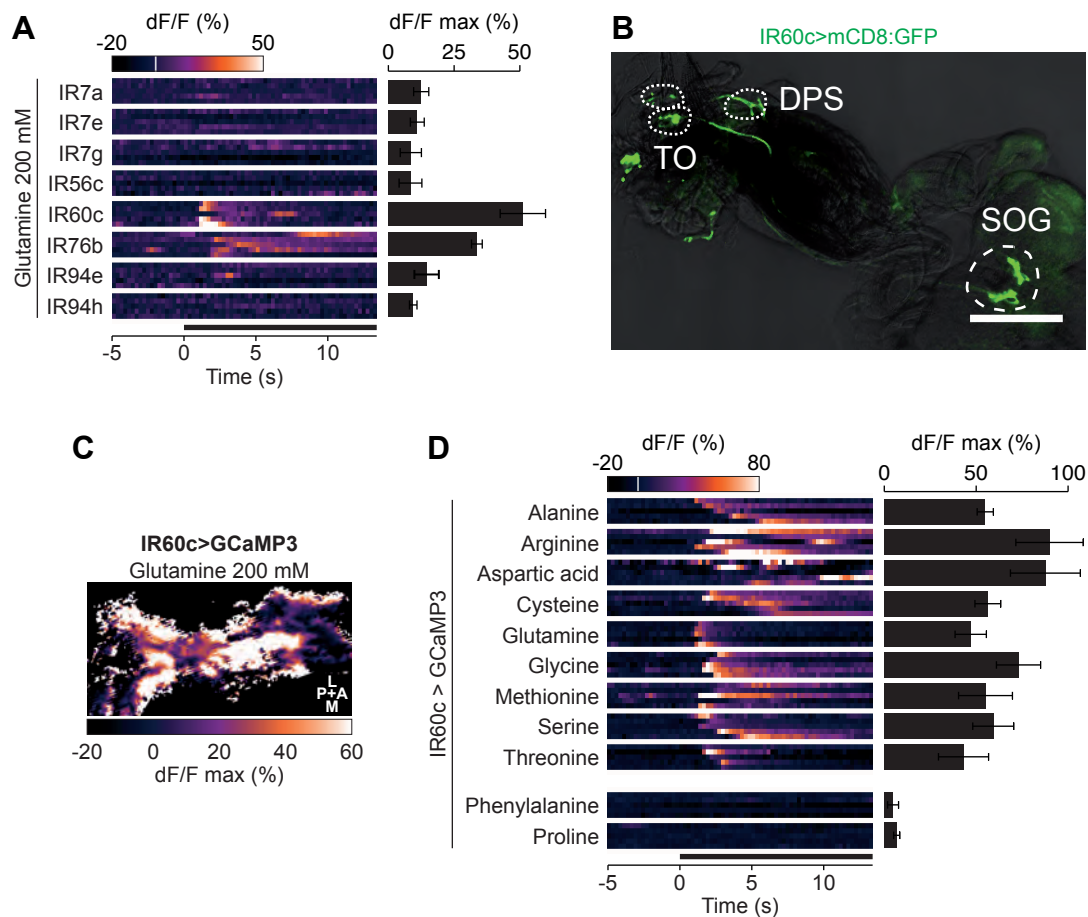


Figure 17 | IR60c neurons respond to amino acids. (A) Left: heatmaps showing the response of IR neurons from the TO to 200 mM glutamine. Right: plots showing the maximum dF/F of these neurons in response to glutamine. Only IR60c and IR76b neurons respond to glutamine. $N \geq 5$, error bars show the standard error of the mean. (B) Immunofluorescence on larvae expressing mCD8:GFP reporter in IR60c neurons. IR60c is expressed in neurons from the TO and DPS that project to the SOG. Scale bar is 100 μm . (C) Activation of GR60c neurons by glutamine increases calcium levels in most areas of the projection. (D) Left: heatmaps showing the response of IR60c neurons to the amino acids (200mM each, except aspartic acid: 50 mM) that activate IR76b neurons (top) or to phenylalanine and proline that do not activate IR76b neurons. Right: plots showing the maximum dF/F of these neurons in response to these amino acids. $N \geq 5$, error bars show the standard error of the mean.

The response of IR60c neurons to different amino acids was then tested. Importantly, these neurons responded to the same subset of amino acids that stimulate IR76b neurons, but not to phenylalanine and proline that do not activate IR76b neurons (Figure 17C). The simplest explanation of this result is that IR60c neurons comprise the only subset of amino acid-sensing IR76b neurons.

Discussion and perspectives

Optical imaging in the Drosophila larval gustatory system

In this chapter, I introduce a novel calcium imaging method to study neuronal activation in the larval SOG. This technique has proven to be very efficient because it allowed identifying neurons that respond to two classes of tastants (sugars and amino acids).

The preparation is rather quick, and an average of 6 to 7 individuals per hour can be tested. However, the major disadvantage of this method is that only one stimulus (along with a positive control) can be performed per larva because of the impossibility of removing the tastant from the sample. An improved stimulus method using a microfluidics chamber is currently being developed by the groups of Simon Sprecher from the University of Fribourg, and of Philippe Renaud from the EPFL, which should provide a tool for screening a larger number of chemicals on the same individual.

In the actual state of research, it was difficult to interpret negative results obtained here. Indeed, no positive control was available for most IR-

Gal4 lines that were tested. Thus, I only relied on the fact that some neurons would repeatedly not respond to a particular tastant to conclude that this tastant is not detected by these neurons. In addition, lower GCaMP3 expression in IR76b neurons in larvae bearing only one copy of the UAS-GCaMP3 transgene resulted in defects in visualising proper fluorescence increase in response to sucrose and glutamine (data not shown), indicating that activation of IR-Gal4 lines that drive expression of low protein levels may also have been missed because of this technical issue. The use of a version of GCaMP whose activation is visible even at low expression levels, or of multiple IR-Gal4 copies could help resolving this issue in the future.

Are IRs involved in sugar perception?

Using this new assay, I identified two populations of IR neurons that respond to sugars. This is the first physiological identification of sugar-sensing gustatory pathways, and raises the question of whether the IRs are the « missing » sugar receptors. However, several factors suggest this is not the case. First, these two populations do not overlap, suggesting that IR7e and IR76b are not co-expressed. However, it has been shown that the function of individual IRs is impaired in absence of a co-receptor (Abuin et al., 2011). Second, because GRs are sugar sensors in adults, it is more likely that they are also involved in sugar perception in larvae, especially as these have been shown to detect all types of sugars (Chyb et al., 2003; Dahanukar et al., 2007; Jiao et al., 2007; Jiao et al., 2008; Wisotsky et al., 2011). Third, no IR neuron was shown to respond to glucose, maltose or

trehalose, suggesting that they are detected by another mechanism. However, it is more parsimonious to hypothesise that only one family of receptors detects all sugars.

In the current state of research, it is thus not possible to assert with certainty what is the exact function of IRs and GRs in sugar detection in larvae. Further research on IR76b and IR7e neurons in larvae should provide the answer to this question, notably by testing the physiological effects of mutating sugar-sensing GRs in these neurons.

An unexpected feature of the larval gustatory system was revealed in these experiments: two different neuronal subsets located in different organs are able to sense sucrose, though with a very different response timing (fast in the TO, long-lasting in the DPS). Due to the experimental setup, there is a prolonged contact with the stimulus and the speed of fluorescence decrease probably reflects the rapidity of neural adaptation. Although the purpose of these dual sensing pathways remains to be established, it may help explaining the function of the internal versus peripheral gustatory organs. Larvae probably use their peripheral organs to seek for food. In order to be able to sense several substrates one after the other, neurons in these organs should be very transiently activated, even in case of prolonged contact with a tastant. In contrary, neurons from the DPS may remain activated for longer in order to carefully assess the quality of the food. Further confirmation of this hypothesis by analysing other types of neurons should be performed in the future.

IRs as candidate amino acid receptors

In addition to sugar-sensing neurons, these experiments have identified the neurons labelled by the IR60c-Gal4 line as neurons that respond to nine amino acids. This represents the first described sensory pathway for amino acids in *Drosophila*. In the two-choice behavioural assay described in Chapter 4 (Figure 11A), these amino acids are not particularly attractive to the larvae, nor are they essential amino acids (apart from methionine and threonine) or do they share particular physico-chemical properties. Thus, the question of what these nine amino acids have in common remains open. One hypothesis is that they are important components of major proteins that are synthesized during larval growth (e.g. cuticular proteins). At least some of these amino acids are of the most abundant amino acids found in fruits (Hall et al., 1980; Pilipenko et al., 1999) or in yeast (Martini et al., 1979), suggesting that these are relevant chemical cues in the substrates that larvae feed on.

Drosophila melanogaster IR60c is annotated as a pseudogene in the genome-sequenced strain because of a deletion in its 5' region compared to other Drosophilid species (Croset et al., 2010). Re-sequencing this region in the w^{1118} strain showed that the 278 bp region missing from the genome strain is present in this line. Because the animals tested in my experiments have a w^{1118} genetic background, they probably express a functional IR60c protein.

The region that was chosen as a promoter to produce the IR60c-Gal4 line is short (274 bp), and so is the distance between IR60b and IR60c. It is thus possible that this sequence is involved in expression regulation of

other members of the IR60b-e cluster. Thus, it is probably wiser to consider all four genes instead of IR60c only as candidate amino acid receptors. Because the sequences of their ligand-binding domains are relatively similar, it is possible that each member of the cluster detects a particular subset of amino acids.

In the future, several experiments remain to be performed, in order to formally identify members from the IR60b-e cluster as amino acid receptors. The response of IR60c neurons should be analysed in mutants for these genes to show their necessity, and mis-expression assays should be conducted in order to test their sufficiency for amino acid sensing.

Material and Methods

Fly strains

All flies were maintained on standard cornmeal-agar medium under a 12 hour light:12 hour dark cycle at 25°C. IR-Gal4 lines that were used here are the same than in chapter 3. The genotype of flies used for calcium imaging was $w[*]; P\{y[+t7.7] w[+mC]=UAS-GCaMP3.T\} attP40; P\{IRXX-Gal4\} / (TM6B)$. UAS-GCaMP3 flies were obtained from Bloomington *Drosophila* Stock Center at Indiana University (Stock n°25813). Flies used in the experiments with tetanus toxin were $w[*]; P\{UAS-TeTxLC.tnt\} E2; P\{IR76b-Gal4\}$ and $w[*]; P\{UAS(FRT.w[+mW.hs]) TeTxLC.IMPTNT\} 14A; P\{IR76b-Gal4\}$.

Calcium imaging in adults

The method for calcium imaging on adult SOG was adapted from (Silbering et al., 2012). A copper collar was fixed on a Plexiglas mounting block and surrounded by a layer of beeswax forming a “U” shape of approximately the height of a fly head. Female flies were held in the collar and their head were immobilised with melted n-eicosane. The antennal plate was dissected out with a surgery blade and a drop of *Drosophila* AHL Ringer's saline (130 mM NaCl, 5 mM KCl, 2 mM MgCl₂, 2 mM CaCl₂, 5 mM HEPES, pH 7.3) was added. Tracheas and the oesophagus were removed with a thin forceps. The proboscis was manually extended and stuck on a tungsten thread with beeswax. The head cavity was subsequently filled with 1.5% low melting point agarose (PeqLab) in *Drosophila* Ringer's saline. A coverslip was placed on top of the preparation and maintained on the beeswax “U”-shape. The tip of the proboscis was kept outside of the coverslip. Tastants were presented to the labellum with a yellow pipette-tip (Figure 11A).

Preparations were visualised with an upright fixed Stage Zeiss Axio Examiner D1 microscope using a Zeiss W "Plan-Apochromat" 20x/1,0 M27 DIC objective. A filter block with the following properties was used: 450-490 nm excitation filter, dichroic mirror (T495LP) and 500-550 nm emission filter (Chroma ET). Images were acquired with a CoolSNAP-HQ2 Digital Camera System using Metafluor acquisition software. Typically, recordings were made for 25 seconds with 4 images per second and tastant was added after 7.5 seconds (frame 30).

Calcium imaging in larvae

Larvae heads were pulled apart from the body so that the brain remains attached to the mouthparts. Imaginal discs and salivary glands were removed but most of the cuticle around the chemosensory organs was left on the animal. Heads were placed on a 50 µl drop of 1.5% agarose in *Drosophila* AHL Ringer's saline on a 25x60 mm glass coverslip and gently pulled away from the drop so that the brain directly lays on the coverslip. Another 50 µl drop of 1.5% agarose was then added to the sample. Once polymerised, a small piece of agarose was removed from the tip of the head with a sharp blade and replaced with 5 µl of AHL Ringer's saline.

Samples were visualised with a Zeiss LSM 510 Meta inverted microscope using a EC Plan-NEOFLUAR 40x/1,30 Oil DIC objective. Images were acquired with an AxioCam MRm camera using LSM 3.5 Software. Typically, recordings were made for 87.5 seconds with 4 images per second (350 frames) and tastant was added after 25 seconds (frame 100).

Calcium imaging data analysis

Sample movement was corrected using the StackReg plugin (bigwww.epfl.ch/thevenaz/stackreg) in NIH ImageJ. Image stacks were further processed using a custom MATLAB (MathWorks) script adapted from Pavan Ramdya and Ana Silbering (Silbering et al., 2011). The relative change in fluorescence (dF/F) was measured for each animal as $(F_i - F_0) / F_0 \times 100$, where F_0 is the mean fluorescence value of frames 6-28 in adults

or 20-95 in larvae (before odor presentation), and F_i is the fluorescence value for the i th frame of the measurement. The average dF/F from circular regions of interest were calculated. Heatmaps and activity traces were produced with custom scripts on R and ImageJ.

Acknowledgements

I would like to thank all the people without whom I would never have been able to complete this thesis.

My first thanks go to my supervisor Richard Benton, first for having me in his lab, and also for all his help during these five great years, for his availability, his patience and for all his wise advices all along my Masters thesis and my PhD. I learned a lot and I was very lucky to do my thesis under his outstanding supervision.

I also would like to thank the members of my thesis committee: Antoine Guisan, Bruno Lemaitre and Bertram Gerber as well as Henrik Kaessmann for their time and all their relevant comments on my work. Many thanks also to my academic mentor Christian Fankhauser for the very interesting discussions about academic life and for his time.

I would also like to thank the people who have provided me with some help all along my PhD. First, Liliane Abuin and Jaime Reina for all their work in managing the lab, their help, their technical advices and their friendship; Ana Silbering and Pavan Ramdya for their help with behavioural experiments and calcium imaging; Raphael Rytz for his help in annotating IRs; Anand Sivasubramaniam for his help in mapping their expression; Ruichen Sun and Romain Groux for their contribution in behavioural experiments; Yael Grosjean for his advices on confocal imaging, and all other past and present members of the Benton lab for their input, discussions about my project and the fun times! Also, a big thank you to Annick Crevoisier for all her help with administrative issues.

Many thanks to Toby Gibson, Aidan Budd and the members of the Gibson team in the EMBL Heidelberg for having me in their group for 3 months and teaching me bioinformatics, to Lena van Giesen, Abud Farca and Simon Sprecher for their invaluable help in larval calcium imaging and for the interesting discussions about larval taste, to Kristin Scott for her constructive comments about my work and for the GR-LexA flies, to Scott Cummins and Bernie Degnan for their contribution in analysing IRs in molluscs and to Jason Pitts for sharing his results about IRs in mosquitoes.

I am grateful to Boehringer Ingelheim Fonds and its entire staff for providing me with a PhD fellowship, for their support and for inviting me to the nice scientific events in Hirschegg and Cold Spring Harbour.

I am truly grateful to all my family and friends for their unconditional support although they do not always really understand what I am doing... Especially, I would like to thank my parents for always being there for me and encouraging me to do what I like.

Finally, all my thoughts go to Rébecca, who was here all the time when I needed. Thank you for everything that you do for me and for your love!

References

- Abuin, L., Bargeton, B., Ulbrich, M.H., Isacoff, E.Y., Kellenberger, S., and Benton, R. (2011). Functional architecture of olfactory ionotropic glutamate receptors. *Neuron* **69**, 44-60.
- Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J., and Zuker, C.S. (2000). A novel family of mammalian taste receptors. *Cell* **100**, 693-702.
- Ai, M., Min, S., Grosjean, Y., Leblanc, C., Bell, R., Benton, R., and Suh, G.S. (2010). Acid sensing by the *Drosophila* olfactory system. *Nature* **468**, 691-695.
- Alcorta, E. (1991). Characterization of the electroantennogram in *Drosophila melanogaster* and its use for identifying olfactory capture and transduction mutants. *J Neurophysiol* **65**, 702-714.
- Amrein, H., and Thorne, N. (2005). Gustatory perception and behavior in *Drosophila melanogaster*. *Curr Biol* **15**, R673-684.
- Armstrong, N., Sun, Y., Chen, G.Q., and Gouaux, E. (1998). Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913-917.
- Asahina, K., Louis, M., Piccinotti, S., and Vosshall, L.B. (2009). A circuit supporting concentration-invariant odor perception in *Drosophila*. *J Biol* **8**, 9.
- Ayalon, G., Segev, E., Elgavish, S., and Stern-Bach, Y. (2005). Two regions in the N-terminal domain of ionotropic glutamate receptor 3 form the subunit oligomerization interfaces that control subtype-specific receptor assembly. *J Biol Chem* **280**, 15053-15060.
- Bader, R., Colomb, J., Pankratz, B., Schrock, A., Stocker, R.F., and Pankratz, M.J. (2007). Genetic dissection of neural circuit anatomy underlying feeding behavior in *Drosophila*: distinct classes of hugin-expressing neurons. *J Comp Neurol* **502**, 848-856.
- Bargmann, C.I. (2006). Chemosensation in *C. elegans*. *WormBook*, 1-29.
- Benton, R., and Dahanukar, A. (2011a). Electrophysiological recording from *Drosophila* olfactory sensilla. *Cold Spring Harb Protoc* **2011**, 824-838.
- Benton, R., and Dahanukar, A. (2011b). Electrophysiological recording from *Drosophila* taste sensilla. *Cold Spring Harb Protoc* **2011**, 839-850.
- Benton, R., Sachse, S., Michnick, S.W., and Vosshall, L.B. (2006). Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol* **4**, e20.
- Benton, R., Vannice, K.S., Gomez-Diaz, C., and Vosshall, L.B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* **136**, 149-162.
- Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* **104**, 3312-3317.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.
- Bray, S., and Amrein, H. (2003). A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019-1029.
- Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175-187.
- Cameron, P., Hiroi, M., Ngai, J., and Scott, K. (2010). The molecular basis for water taste in *Drosophila*. *Nature* **465**, 91-95.
- Chou, Y.H., Spletter, M.L., Yaksi, E., Leong, J.C., Wilson, R.I., and Luo, L. (2010). Diversity and wiring variability of olfactory local interneurons in the *Drosophila* antennal lobe. *Nat Neurosci* **13**, 439-449.

- Chyb, S., Dahanukar, A., Wickens, A., and Carlson, J.R. (2003). *Drosophila Gr5a* encodes a taste receptor tuned to trehalose. *Proc Natl Acad Sci U S A* *100 Suppl 2*, 14526-14530.
- Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J., and Carlson, J.R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* *22*, 327-338.
- Colomb, J., Grillenzoni, N., Ramaekers, A., and Stocker, R.F. (2007). Architecture of the primary taste center of *Drosophila melanogaster* larvae. *J Comp Neurol* *502*, 834-847.
- Couto, A., Alenius, M., and Dickson, B.J. (2005). Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr Biol* *15*, 1535-1547.
- Croset, V., Rytz, R., Cummins, S.F., Budd, A., Brawand, D., Kaessmann, H., Gibson, T.J., and Benton, R. (2010). Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet* *6*, e1001064.
- Dahanukar, A., Foster, K., van der Goes van Naters, W.M., and Carlson, J.R. (2001). A Gr receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. *Nat Neurosci* *4*, 1182-1186.
- Dahanukar, A., Lei, Y.T., Kwon, J.Y., and Carlson, J.R. (2007). Two *Gr* genes underlie sugar reception in *Drosophila*. *Neuron* *56*, 503-516.
- de Bruyne, M., Foster, K., and Carlson, J.R. (2001). Odor coding in the *Drosophila* antenna. *Neuron* *30*, 537-552.
- Derby, C.D., and Sorensen, P.W. (2008). Neural processing, perception, and behavioral responses to natural chemical stimuli by fish and crustaceans. *J Chem Ecol* *34*, 898-914.
- Dulac, C., and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* *83*, 195-206.
- Dunipace, L., Meister, S., McNealy, C., and Amrein, H. (2001). Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. *Curr Biol* *11*, 822-835.
- Falk, R., and Atidia, J. (1975). Mutation affecting taste perception in *Drosophila melanogaster*. *Nature* *254*, 325-326.
- Fiala, A., Spall, T., Diegelmann, S., Eisermann, B., Sachse, S., Devaud, J.M., Buchner, E., and Galizia, C.G. (2002). Genetically expressed cameleon in *Drosophila melanogaster* is used to visualize olfactory information in projection neurons. *Curr Biol* *12*, 1877-1884.
- Fischler, W., Kong, P., Marella, S., and Scott, K. (2007). The detection of carbonation by the *Drosophila* gustatory system. *Nature* *448*, 1054-1057.
- Fishilevich, E., Domingos, A.I., Asahina, K., Naef, F., Vosshall, L.B., and Louis, M. (2005). Chemotaxis Behavior Mediated by Single Larval Olfactory Neurons in *Drosophila*. *Curr Biol* *15*, 2086-2096.
- Fishilevich, E., and Vosshall, L.B. (2005). Genetic and functional subdivision of the *Drosophila* antennal lobe. *Curr Biol* *15*, 1548-1553.
- Gao, Q., and Chess, A. (1999). Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* *60*, 31-39.
- Gao, Q., Yuan, B., and Chess, A. (2000). Convergent projections of *Drosophila* olfactory neurons to specific glomeruli in the antennal lobe. *Nat Neurosci* *3*, 780-785.
- Gendre, N., Luer, K., Friche, S., Grillenzoni, N., Ramaekers, A., Technau, G.M., and Stocker, R.F. (2004). Integration of complex larval chemosensory organs into the adult nervous system of *Drosophila*. *Development* *131*, 83-92.
- Gerber, B., and Stocker, R.F. (2007). The *Drosophila* larva as a model for studying chemosensation and chemosensory learning: a review. *Chem Senses* *32*, 65-89.
- Getahun, M.N., Wicher, D., Hansson, B.S., and Olsson, S.B. (2012). Temporal response dynamics of *Drosophila* olfactory sensory neurons depends on receptor type and response polarity. *Front Cell Neurosci* *6*, 54.

- Ghannad-Rezaie, M., Wang, X., Mishra, B., Collins, C., and Chronis, N. (2012). Microfluidic chips for in vivo imaging of cellular responses to neural injury in *Drosophila* larvae. *PLoS One* *7*, e29869.
- Gordon, M.D., and Scott, K. (2009). Motor control in a *Drosophila* taste circuit. *Neuron* *61*, 373-384.
- Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., *et al.* (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* *471*, 473-479.
- Grosjean, Y., Rytz, R., Farine, J.P., Abuin, L., Cortot, J., Jefferis, G.S., and Benton, R. (2011). An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature* *478*, 236-240.
- Guerrero, G., Reiff, D.F., Agarwal, G., Ball, R.W., Borst, A., Goodman, C.S., and Isacoff, E.Y. (2005). Heterogeneity in synaptic transmission along a *Drosophila* larval motor axon. *Nat Neurosci* *8*, 1188-1196.
- Hall, N.T., Smoot, J.M., Knight, R.J., Jr., and Nagy, S. (1980). Protein and amino acid compositions of ten tropical fruits by gas-liquid chromatography. *J Agric Food Chem* *28*, 1217-1221.
- Hallem, E.A., and Carlson, J.R. (2006). Coding of odors by a receptor repertoire. *Cell* *125*, 143-160.
- Hallem, E.A., Ho, M.G., and Carlson, J.R. (2004). The molecular basis of odor coding in the *Drosophila* antenna. *Cell* *117*, 965-979.
- Heimbeck, G., Bugnon, V., Gendre, N., Haberin, C., and Stocker, R.F. (1999). Smell and taste perception in *Drosophila melanogaster* larva: toxin expression studies in chemosensory neurons. *J Neurosci* *19*, 6599-6609.
- Heisenberg, M., Borst, A., Wagner, S., and Byers, D. (1985). *Drosophila* mushroom body mutants are deficient in olfactory learning. *J Neurogenet* *2*, 1-30.
- Herrada, G., and Dulac, C. (1997). A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* *90*, 763-773.
- Hiroi, M., Marion-Poll, F., and Tanimura, T. (2002). Differentiated response to sugars among labellar chemosensilla in *Drosophila*. *Zool J Linn Soc* *159*, 1009-1018.
- Hoare, D.J., Humble, J., Jin, D., Gilding, N., Petersen, R., Cobb, M., and McCrohan, C. (2011). Modeling peripheral olfactory coding in *Drosophila* larvae. *PLoS One* *6*, e22996.
- Inglis, P.N., Ou, G., Leroux, M.R., and Scholey, J.M. (2007). The sensory cilia of *Caenorhabditis elegans*. *WormBook*, 1-22.
- Ishimoto, H., and Tanimura, T. (2004). Molecular neurophysiology of taste in *Drosophila*. *Cell Mol Life Sci* *61*, 10-18.
- Jefferis, G.S., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfs, T., Maurer, C.R., Jr., and Luo, L. (2007). Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* *128*, 1187-1203.
- Jiao, Y., Moon, S.J., and Montell, C. (2007). A *Drosophila* gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. *Proc Natl Acad Sci U S A* *104*, 14110-14115.
- Jiao, Y., Moon, S.J., Wang, X., Ren, Q., and Montell, C. (2008). Gr64f is required in combination with other gustatory receptors for sugar detection in *Drosophila*. *Curr Biol* *18*, 1797-1801.
- Jones, W.D., Cayirlioglu, P., Grunwald Kadow, I., and Vosshall, L.B. (2007). Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* *445*, 86-90.

- Jones, W.D., Nguyen, T.A., Kloss, B., Lee, K.J., and Vosshall, L.B. (2005). Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr Biol* 15, R119-121.
- Karunanithi, S., Georgiou, J., Charlton, M.P., and Atwood, H.L. (1997). Imaging of calcium in *Drosophila* larval motor nerve terminals. *J Neurophysiol* 78, 3465-3467.
- Keene, A.C., and Waddell, S. (2007). *Drosophila* olfactory memory: single genes to complex neural circuits. *Nat Rev Neurosci* 8, 341-354.
- Koganezawa, M., Haba, D., Matsuo, T., and Yamamoto, D. (2010). The shaping of male courtship posture by lateralized gustatory inputs to male-specific interneurons. *Curr Biol* 20, 1-8.
- Kreher, S.A., Kwon, J.Y., and Carlson, J.R. (2005). The molecular basis of odor coding in the *Drosophila* larva. *Neuron* 46, 445-456.
- Kuner, T., Seeburg, P.H., and Guy, H.R. (2003). A common architecture for K⁺ channels and ionotropic glutamate receptors? *Trends Neurosci* 26, 27-32.
- Kwon, J.Y., Dahanukar, A., Weiss, L.A., and Carlson, J.R. (2007). The molecular basis of CO₂ reception in *Drosophila*. *Proc Natl Acad Sci U S A* 104, 3574-3578.
- Kwon, J.Y., Dahanukar, A., Weiss, L.A., and Carlson, J.R. (2011). Molecular and cellular organization of the taste system in the *Drosophila* larva. *J Neurosci* 31, 15300-15309.
- Lai, S.L., and Lee, T. (2006). Genetic mosaic with dual binary transcriptional systems in *Drosophila*. *Nat Neurosci* 9, 703-709.
- Larsson, M.C., Domingos, A.I., Jones, W.D., Chiappe, M.E., Amrein, H., and Vosshall, L.B. (2004). *Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* 43, 703-714.
- Lee, T., and Luo, L. (2001). Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends Neurosci* 24, 251-254.
- Lee, Y., Moon, S.J., and Montell, C. (2009). Multiple gustatory receptors required for the caffeine response in *Drosophila*. *Proc Natl Acad Sci U S A* 106, 4495-4500.
- Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., and Adler, E. (2002). Human receptors for sweet and umami taste. *Proc Natl Acad Sci U S A* 99, 4692-4696.
- Liberles, S.D., and Buck, L.B. (2006). A second class of chemosensory receptors in the olfactory epithelium. *Nature* 442, 645-650.
- Liu, L., Leonard, A.S., Motto, D.G., Feller, M.A., Price, M.P., Johnson, W.A., and Welsh, M.J. (2003a). Contribution of *Drosophila* DEG/ENaC genes to salt taste. *Neuron* 39, 133-146.
- Liu, L., Yermolaieva, O., Johnson, W.A., Abboud, F.M., and Welsh, M.J. (2003b). Identification and function of thermosensory neurons in *Drosophila* larvae. *Nat Neurosci* 6, 267-273.
- Lledo, P.M., Gheusi, G., and Vincent, J.D. (2005). Information processing in the mammalian olfactory system. *Physiol Rev* 85, 281-317.
- Mank, M., and Griesbeck, O. (2008). Genetically encoded calcium indicators. *Chem Rev* 108, 1550-1564.
- Manoli, D.S., Foss, M., Vilella, A., Taylor, B.J., Hall, J.C., and Baker, B.S. (2005). Male-specific fruitless specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* 436, 395-400.
- Marella, S., Fischler, W., Kong, P., Asgarian, S., Rueckert, E., and Scott, K. (2006). Imaging taste responses in the fly brain reveals a functional map of taste category and behavior. *Neuron* 49, 285-295.
- Marella, S., Mann, K., and Scott, K. (2012). Dopaminergic modulation of sucrose acceptance behavior in *Drosophila*. *Neuron* 73, 941-950.

- Marin, E.C., Watts, R.J., Tanaka, N.K., Ito, K., and Luo, L. (2005). Developmentally programmed remodeling of the *Drosophila* olfactory circuit. *Development* 132, 725-737.
- Markow, T.A., Beall, S., and Matzkin, L.M. (2009). Egg size, embryonic development time and ovoviviparity in *Drosophila* species. *J Evol Biol* 22, 430-434.
- Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40, 476-483.
- Martini, A.E., Miller, M.W., and Martini, A. (1979). Amino acid composition of whole cells of different yeasts. *J Agric Food Chem* 27, 982-984.
- Masuko, T., Kuno, T., Kashiwagi, K., Kusama, T., Williams, K., and Igarashi, K. (1999). Stimulatory and inhibitory properties of aminoglycoside antibiotics at N-methyl-D-aspartate receptors. *J Pharmacol Exp Ther* 290, 1026-1033.
- Matsunami, H., and Buck, L.B. (1997). A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* 90, 775-784.
- Matsunami, H., Montmayeur, J.P., and Buck, L.B. (2000). A family of candidate taste receptors in human and mouse. *Nature* 404, 601-604.
- Mayer, M.L. (2006). Glutamate receptors at atomic resolution. *Nature* 440, 456-462.
- Mayer, M.L., and Armstrong, N. (2004). Structure and function of glutamate receptor ion channels. *Annu Rev Physiol* 66, 161-181.
- McCombs, J.E., and Palmer, A.E. (2008). Measuring calcium dynamics in living cells with genetically encodable calcium indicators. *Methods* 46, 152-159.
- Miller, M.R., Robinson, K.J., Cleary, M.D., and Doe, C.Q. (2009). TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Methods* 6, 439-441.
- Mitri, C., Soustelle, L., Framery, B., Bockaert, J., Parmentier, M.L., and Grau, Y. (2009). Plant insecticide L-canavanine repels *Drosophila* via the insect orphan GPCR DmX. *PLoS Biol* 7, e1000147.
- Miyamoto, T., and Amrein, H. (2008). Suppression of male courtship by a *Drosophila* pheromone receptor. *Nat Neurosci* 11, 874-876.
- Miyamoto, T., Fujiyama, R., Okada, Y., and Sato, T. (1998). Salty and sour transduction. Multiple mechanisms and strain differences. *Ann N Y Acad Sci* 855, 128-133.
- Miyamoto, T., Slone, J., Song, X., and Amrein, H. (2012). A fructose receptor functions as a nutrient sensor in the *Drosophila* brain. *Cell* 151, 1113-1125.
- Miyazaki, T., and Ito, K. (2010). Neural architecture of the primary gustatory center of *Drosophila melanogaster* visualized with GAL4 and LexA enhancer-trap systems. *J Comp Neurol* 518, 4147-4181.
- Montell, C. (2009). A taste of the *Drosophila* gustatory receptors. *Curr Opin Neurobiol* 19, 345-353.
- Moon, S.J., Kottgen, M., Jiao, Y., Xu, H., and Montell, C. (2006). A taste receptor required for the caffeine response *in vivo*. *Curr Biol* 16, 1812-1817.
- Moon, S.J., Lee, Y., Jiao, Y., and Montell, C. (2009). A *Drosophila* gustatory receptor essential for aversive taste and inhibiting male-to-male courtship. *Curr Biol* 19, 1623-1627.
- Nagoshi, E., Sugino, K., Kula, E., Okazaki, E., Tachibana, T., Nelson, S., and Rosbash, M. (2010). Dissecting differential gene expression within the circadian neuronal circuit of *Drosophila*. *Nat Neurosci* 13, 60-68.
- Nei, M., Niimura, Y., and Nozawa, M. (2008). The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9, 951-963.
- Ng, M., Roorda, R.D., Lima, S.Q., Zemelman, B.V., Morcillo, P., and Miesenbock, G. (2002). Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly. *Neuron* 36, 463-474.

- Panchenko, V.A., Glasser, C.R., and Mayer, M.L. (2001). Structural similarities between glutamate receptor channels and K(+) channels examined by scanning mutagenesis. *J Gen Physiol* 117, 345-360.
- Paoletti, P., Ascher, P., and Neyton, J. (1997). High-affinity zinc inhibition of NMDA NR1-NR2A receptors. *J Neurosci* 17, 5711-5725.
- Park, J.H., and Kwon, J.Y. (2011). A systematic analysis of *Drosophila* gustatory receptor gene expression in abdominal neurons which project to the central nervous system. *Mol Cells* 32, 375-381.
- Park, K.C., Ochieng, S.A., Zhu, J., and Baker, T.C. (2002). Odor discrimination using insect electroantennogram responses from an insect antennal array. *Chem Senses* 27, 343-352.
- Pilipenko, L.N., Kalinkov, A.Y., and V., S.A. (1999). Amino acid composition of fruit in the manufacture of sedimentation-stabilized dispersed products. *Chemistry of Natural Compounds* 35, 208-211.
- Python, F., and Stocker, R.F. (2002). Adult-like complexity of the larval antennal lobe of *D. melanogaster* despite markedly low numbers of odorant receptor neurons. *J Comp Neurol* 445, 374-387.
- R-Development-Core-Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramaekers, A., Magnenat, E., Marin, E.C., Gendre, N., Jefferis, G.S., Luo, L., and Stocker, R.F. (2005). Glomerular maps without cellular redundancy at successive levels of the *Drosophila* larval olfactory circuit. *Curr Biol* 15, 982-992.
- Robertson, H.M., and Thomas, J.H. (2006). The putative chemoreceptor families of *C. elegans*. *WormBook*, 1-12.
- Robertson, H.M., Warr, C.G., and Carlson, J.R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 100 Suppl 2, 14537-14542.
- Rohwedder, A., Pfitzenmaier, J.E., Ramsperger, N., Apostolopoulou, A.A., Widmann, A., and Thum, A.S. (2012). Nutritional value-dependent and nutritional value-independent effects on *Drosophila melanogaster* larval behavior. *Chem Senses* 37, 711-721.
- Ruta, V., Datta, S.R., Vasconcelos, M.L., Freeland, J., Looger, L.L., and Axel, R. (2010). A dimorphic pheromone circuit in *Drosophila* from sensory input to descending output. *Nature* 468, 686-690.
- Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., Baker, B.S., Hall, J.C., Taylor, B.J., and Wasserman, S.A. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. *Cell* 87, 1079-1089.
- Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L.B., and Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* 452, 1002-1006.
- Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino de Laureto, P., DasGupta, B.R., and Montecucco, C. (1992). Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* 359, 832-835.
- Schleyer, M., Saumweber, T., Nahrendorf, W., Fischer, B., von Alpen, D., Pauls, D., Thum, A., and Gerber, B. (2011). A behavior-based circuit model of how outcome expectations organize learned behavior in larval *Drosophila*. *Learn Mem* 18, 639-653.
- Schroll, C., Riemensperger, T., Bucher, D., Ehmer, J., Voller, T., Erbguth, K., Gerber, B., Hendel, T., Nagel, G., Buchner, E., *et al.* (2006). Light-induced activation of distinct modulatory neurons triggers appetitive or aversive learning in *Drosophila* larvae. *Curr Biol* 16, 1741-1747.
- Schroter, U., Malun, D., and Menzel, R. (2007). Innervation pattern of suboesophageal ventral unpaired median neurones in the honeybee brain. *Cell Tissue Res* 327, 647-667.

- Shanbhag, S.R., Park, S.K., Pikielny, C.W., and Steinbrecht, R.A. (2001). Gustatory organs of *Drosophila melanogaster*: fine structure and expression of the putative odorant-binding protein PBPRP2. *Cell Tissue Res* 304, 423-437.
- Silbering, A.F., Bell, R., Galizia, C.G., and Benton, R. (2012). Calcium imaging of odor-evoked responses in the *Drosophila* antennal lobe. *J Vis Exp*.
- Silbering, A.F., Rytz, R., Grosjean, Y., Abuin, L., Ramdya, P., Jefferis, G.S., and Benton, R. (2011). Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *J Neurosci* 31, 13357-13375.
- Singh, R.N., and Singh, K. (1984). Fine structure of the sensory organs of *Drosophila melanogaster* Meigen larva (Diptera : Drosophilidae). *International Journal of Insect Morphology and Embryology* 13, 18.
- Slone, J., Daniels, J., and Amrein, H. (2007). Sugar receptors in *Drosophila*. *Curr Biol* 17, 1809-1816.
- Smith, C.D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C.R., Elhaik, E., Elsik, C.G., *et al.* (2011a). Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci U S A* 108.
- Smith, C.R., Smith, C.D., Robertson, H.M., Helmkamp, M., Zimin, A., Yandell, M., Holt, C., Hu, H., Abouheif, E., Benton, R., *et al.* (2011b). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci U S A* 108.
- Stern-Bach, Y., Bettler, B., Hartley, M., Sheppard, P.O., O'Hara, P.J., and Heinemann, S.F. (1994). Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. *Neuron* 13, 1345-1357.
- Stocker, R.F. (1994). The organization of the chemosensory system in *Drosophila melanogaster*: a review. *Cell Tissue Res* 275, 3-26.
- Stocker, R.F., and Schorderet, M. (1981). Cobalt filling of sensory projections from internal and external mouthparts in *Drosophila*. *Cell Tissue Res* 216, 513-523.
- Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L., and Dickson, B.J. (2005). Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* 121, 795-807.
- Sweeney, S.T., Broadie, K., Keane, J., Niemann, H., and O'Kane, C.J. (1995). Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* 14, 341-351.
- Tanaka, N.K., Awasaki, T., Shimada, T., and Ito, K. (2004). Integration of chemosensory pathways in the *Drosophila* second-order olfactory centers. *Curr Biol* 14, 449-457.
- Thorne, N., Chromey, C., Bray, S., and Amrein, H. (2004). Taste perception and coding in *Drosophila*. *Curr Biol* 14, 1065-1079.
- Thum, A.S., Leisibach, B., Gendre, N., Selcho, M., and Stocker, R.F. (2011). Diversity, variability, and suboesophageal connectivity of antennal lobe neurons in *D. melanogaster* larvae. *J Comp Neurol* 519, 3415-3432.
- Tian, L., Hires, S.A., Mao, T., Huber, D., Chiappe, M.E., Chalasani, S.H., Petreanu, L., Akerboom, J., McKinney, S.A., Schreiter, E.R., *et al.* (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Methods* 6, 875-881.
- Toshima, N., and Tanimura, T. (2012). Taste preference for amino acids is dependent on internal nutritional state in *Drosophila melanogaster*. *J Exp Biol* 215, 2827-2832.
- Turner, S.L., and Ray, A. (2009). Modification of CO₂ avoidance behaviour in *Drosophila* by inhibitory odorants. *Nature* 461, 277-281.
- Ueno, K., Ohta, M., Morita, H., Mikuni, Y., Nakajima, S., Yamamoto, K., and Isono, K. (2001). Trehalose sensitivity in *Drosophila* correlates with mutations in and expression of the gustatory receptor gene Gr5a. *Curr Biol* 11, 1451-1455.
- Ulbrich, M.H., and Isacoff, E.Y. (2007). Subunit counting in membrane-bound proteins. *Nat Methods* 4, 319-321.

- Ulbrich, M.H., and Isacoff, E.Y. (2008). Rules of engagement for NMDA receptor subunits. *Proc Natl Acad Sci U S A* *105*, 14163-14168.
- Vosshall, L.B., Amrein, H., Morozov, P.S., Rzhetsky, A., and Axel, R. (1999). A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* *96*, 725-736.
- Vosshall, L.B., and Stocker, R.F. (2007). Molecular Architecture of Smell and Taste in *Drosophila*. *Annu Rev Neurosci*.
- Wagh, D.A., Rasse, T.M., Asan, E., Hofbauer, A., Schwenkert, I., Durrbeck, H., Buchner, S., Dabauvalle, M.C., Schmidt, M., Qin, G., *et al.* (2006). Bruchpilot, a protein with homology to ELKS/CAST, is required for structural integrity and function of synaptic active zones in *Drosophila*. *Neuron* *49*, 833-844.
- Wang, J.W., Wong, A.M., Flores, J., Vosshall, L.B., and Axel, R. (2003). Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* *112*, 271-282.
- Wang, Z., Singhvi, A., Kong, P., and Scott, K. (2004). Taste representations in the *Drosophila* brain. *Cell* *117*, 981-991.
- Ware, R.W., D., C., K., C., and R.L., R. (1975). The Nerve Ring of the Nematode *Caenorhabditis elegans*: Sensory Input and Motor Output. *J Comp Neurol* *162*, 71-110.
- Weiss, L.A., Dahanukar, A., Kwon, J.Y., Banerjee, D., and Carlson, J.R. (2011). The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron* *69*, 258-272.
- Wicher, D., Schafer, R., Bauernfeind, R., Stensmyr, M.C., Heller, R., Heinemann, S.H., and Hansson, B.S. (2008). *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* *452*, 1007-1011.
- Wisotsky, Z., Medina, A., Freeman, E., and Dahanukar, A. (2011). Evolutionary differences in food preference rely on Gr64e, a receptor for glycerol. *Nat Neurosci* *14*, 1534-1541.
- Yang, C.H., Belawat, P., Hafen, E., Jan, L.Y., and Jan, Y.N. (2008). *Drosophila* egg-laying site selection as a system to study simple decision-making processes. *Science* *319*, 1679-1683.
- Yang, Z., Edenberg, H.J., and Davis, R.L. (2005). Isolation of mRNA from specific tissues of *Drosophila* by mRNA tagging. *Nucleic Acids Res* *33*, e148.
- Yao, C.A., Ignell, R., and Carlson, J.R. (2005). Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *J Neurosci* *25*, 8359-8367.