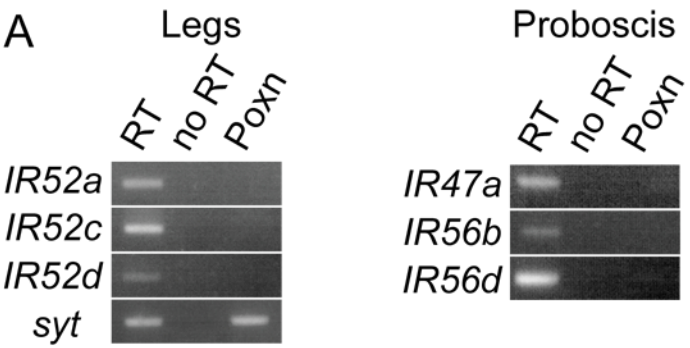
Figure S1
related to Figure 1



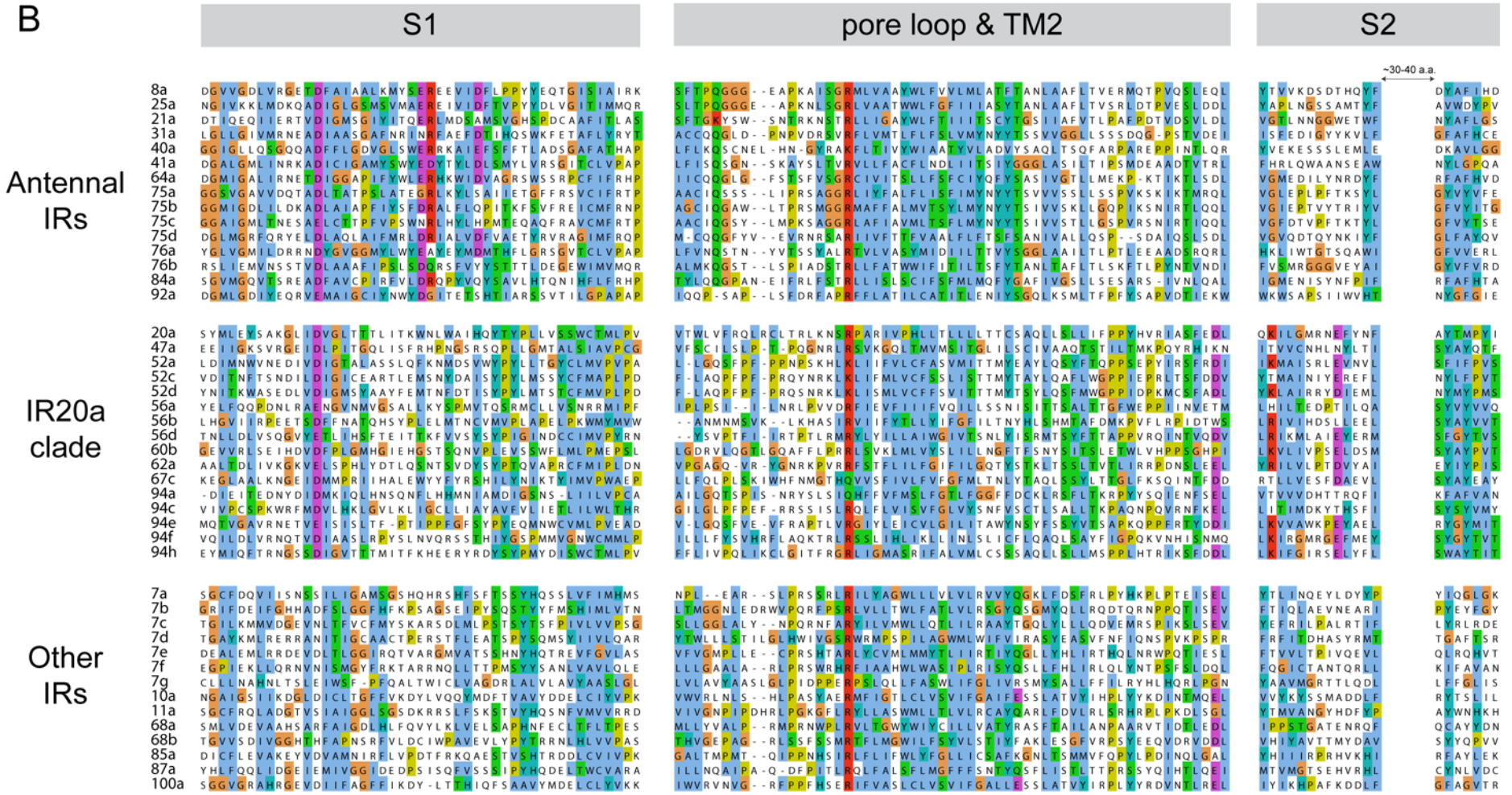


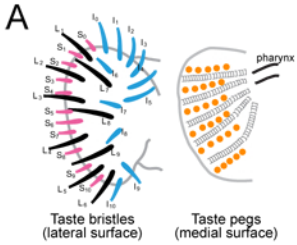
Figure S1. RT-PCR analysis and sequence alignment of members of the *IR20a* clade.
(A) *Left:* In reverse transcriptase (RT)-treated polyA$^+$ RNA extracts from legs, transcripts of *IR52a, IR52c* and *IR52d* were detected in the control genotype, Cantonized w^{1118} (RT lanes), but not in the $Poxn^{70}$ mutant, in which taste sensilla are missing (Awasaki and Kimura, 1997). As a control for RNA quality, the *synaptotagmin* (*syt*) transcript was detected in both control and $Poxn^{70}$ genotypes. In the "no RT" lanes, polyA$^+$ RNA extracts were not treated with RT and the lack of PCR products indicates the absence of genomic DNA contamination.
Right: In RT-treated polyA$^+$ RNA extracts from proboscis, transcripts of *IR47a, IR56b* and *IR56d* were detected in the control, but not in the $Poxn^{70}$ mutant. No genomic DNA contamination was detected in the "no RT" controls.
(B) Alignment of protein sequences from the IR family. IRs expressed in the antenna are grouped at the top (Antennal IRs), IR20a clade members that are expressed in taste organs are grouped in the middle, and non-IR20a clade, non-olfactory IRs are grouped at the bottom ("Other IRs"). Shown are regions predicted to be of particular functional importance. Based on homology with ionotropic glutamate receptors, predicted ligand-binding structures are formed by two non-contiguous domains, S1 and S2 (Armstrong et al., 1998). Core portions of the S1 and S2 domains that are known to contact ligands are shown in these alignments (Armstrong et al., 1998). The S2 domain is composed of two segments of conserved residues separated by ~30-40 non-conserved residues. S1 and S2 flank the pore loop region and second transmembrane domain (TM2), which are predicted to form parts of the ion channel (Benton et al., 2009; Mayer and Armstrong, 2004). The alignment was performed using the TCoffee algorithm (Notredame et al., 2000), manually adjusted using the JalView program and color-coded using the ClustalX color scheme (Waterhouse et al., 2009). The IR20a clade members that are expressed in taste organs appear to differ from the olfactory IRs in the following manners: (i) the absence of a relatively conserved arginine (red) in the S1 domain, (ii) the presence of a moderately conserved basic residue (red; arginine or lysine) and a acidic residue (purple; aspartate or glutamate) in the N-terminal portion of the S2 domain, and (iii) the presence of moderately conserved serines/threonines (green) and the lack of a conserved glycine (orange) in the C-terminal portion of the S2 domain. Although IR25a and IR76b are expressed in both olfactory and taste organs, we have grouped them with the olfactory IRs. Within the group labeled "Other IRs", the *GAL4* drivers of *IR7a*, *IR11a* and *IR100a* were previously shown to be expressed in taste organs (Croset et al., 2010).

Table S1, related to Figure 1

Gene	5' flanking region (kb)	3' flanking region (kb)	Vector	5' flanking region		3' flanking region		Template
				forward primer	reverse primer	forward primer	reverse primer	
IR20a	4	2.2	*pCasper5*^{P}	GTTGGCCAACTAAGTTTAAGTAGC	GTCGCCGGCATCGAAG	GCGTCTGCGATTTTCATAATGG	AACTTACCGGGCTCTCATAG	*BACR23I18*
IR47a	3.5	3.3	*pCasper5*^{P}	GACAAATTACCTCCACATGAGC	TTTTTATGGCCTTTTGAAACTGAACTG	CGAATATGAAAGACGATGGATAGG	CTTTACGATAAATGTGCAATTGG	*BACR20O22*
IR48b	4	4	*pBGRY*$^{\phi C31}$	CGCCATGGGAACTTTACTCGAAACTG	GGTGGTTCACAAAATCTAGC	CGCTTCGATCAGGATTACG	GAATCAGCGACAATTACCC	*BACR35F01*
IR48c	3.6	1	*pCasper5*^{P}	TTTTAGCTGTTTAGCACTCG	TCGAAAGGGTTGTAAGTTAATAAAAG	CCTAATGGGCTTACTCTCAG	CTATTTTTACCCCAACTTATTTGAACC	*BACR26P21*
IR51b	4.5	1.2	*AD1515*^{P}	CCCCCTTGATGTGCAATAATTC	GGGCACAATCAAGCAGTTTGG	CATGGCGATTTATTGCTTAG	CCACCACCTCAATTGCCAATTAAAAAGG	CS-5
IR52a	3.8	none	*pG4PN+*^{P}	GTTTATCTTGGGAACCGAAAGAG	TAGGGTGTACGAGTTCTCTTC	n.a.	n.a.	CS-5
	3.8	none	*pBGRY*$^{\phi C31}$	GTTTATCTTGGGAACCGAAAGAG	TAGGGTGTACGAGTTCTCTTC	SV40 poly A signal		CS-5
IR52b	4.4	8.2	*pBGRY*$^{\phi C31}$	CAGATTTCGGTGGCTCTGG	CTATCCGACACCAGCACTACG	TCTGCGCAGACACCTGC	CTGCCACAGTGGGCATTGC	*BACR48C01*
IR52c	6.6	6	*pBGRY*$^{\phi C31}$	CAGATTTCGGTGGCTCTGG	GTGTCGTAGGACTTCTCG	ACATGCTGCTGCAGAAGG	CTGCCACAGTGGGCATTGC	*BACR48C01*
IR52d	9	3.4	*pBGRY*$^{\phi C31}$	CAGATTTCGGTGGCTCTGG	TGGGGACTTGCAATCTCC	TGGGATTACCAAGTTGGACG	CAGCAACAGCTGTTAAGTCG	*BACR48C01*
IR54a	1.3	3.2	*pCasper5*^{P}	GCAGCGTTATCTGAGCTC	AACAGTAAGAGGGTGACCG	TCATTCACATGTTTATTGGAAAACGTG	GTCGAAGCTGTTTCAATCAATTCC	*BACR20E20*
IR56a	3.3	2.9	*AD1515*^{P}	GCTATTTGTCGGTGCGGAATCC	GGCTGCCTTACCACTTTGACTTC	CGTTGCTGAACTGCTTCACAATGG	GCTCGGTCTAATGATACTGGTCAC	CS-5
IR56b	3.7	11.3	*pBGRY*$^{\phi C31}$	GGTGCCCATACTGGTTTTCG	CCAGGTCCGTGTCAAGC	CGGGAATACCCATTAGTTCG	CGTAACGGCACCACACG	*BACR25K02*
IR56c	5.4	9.2	*pBGRY*$^{\phi C31}$	GGTGCCCATACTGGTTTTCG	AATGGAGCCAAAAGATTGAGC	CCGGGATCGTGTTTGTCC	CGTAACGGCACCACACG	*BACR25K02*
IR56d	4	0.2	*pCasper5*^{P}	TGTTATGGCAGTGGTTTTCC	ATTTGTACGGCGACTGCCAGTG	GGCGCTTCATGAAATTATGATCC	GGAAACTCTTTCTAGGTGAAGG	*BACR25K02*
IR60b	1.2	6.5	*pBDPR*$^{\phi C31}$	CAGCAGGGCGTTCATGG	CTCCGCCTCATTTTCGAGTTG	TCGTGGGCGGAACCATCG	CTACGCCCAGCTCAGTCC	*BACR11C07*
(IR60c)	3.4	4.8	*pBGRY*$^{\phi C31}$	CAGCAGGGCGTTCATGG	CTGCGGCTCGAAGTACC	CTGGCTTGGCTTTCTTGG	CTACGCCCAGCTCAGTCC	*BACR11C07*
IR60d	5.1	2.7	*pBGRY*$^{\phi C31}$	CAGCAGGGCGTTCATGG	CAAGCAGGCAAGTGCAAGC	GTGGCACTTGGGGAACTCC	CTACGCCCAGCTCAGTCC	*BACR11C07*
(IR60e)	2.3	none	*pG4PN+*^{P}	AACGAGAACTGACGAGACC	TGGATTCCGAGTCAGAAGC	n.a.	n.a.	CS-5
IR62a	4.4	3.9	*pBGRY*$^{\phi C31}$	TGGCGATGTGAGAGTGG	TGGCCACGATTTGGTAGC	AACCCCGGGGATTAATGC	GTGGCACAAACCTTTGC	*BACR27O01*
IR67a	1.1	3.6	*AD1515*^{P}	CAGTAGGAAATGAAACACCTGC	GCTAACCATCACCACATACAC	GATCCATTTCCATATGTTGCTTTCCGAG	CCCGCTAGCCGCTAATTTGC	CS-5
	8.3	none	*pBGRY*$^{\phi C31}$	ACTTTGGTCTCTTTGTTCG	GCTGTCTGGCAAACC	SV40 poly A signal		*BACR12K04*
IR67b	4.2	2.7	*pBGRY*$^{\phi C31}$	TTGTTGCCGTTCGTTGTC	CCGAGCCCGGTAGCATTAC	CGCTGCATTTGGACGTCC	CAGAGAAGGCGAGAGC	*BACR05G01*
IR67c	1.6	3	*pCasper5*^{P}	ATTTACATAAATGCTTTGAATGC	GATGCAGTCCTGCCCGAAAAAG	GGAATGGGATGGTGTGTC	CCTTACGACTCATACAAACATATG	CS-5
	1.6	3	*pBGRY*$^{\phi C31}$	ATTTACATAAATGCTTTGAATGC	GATGCAGTCCTGCCCGAAAAAG	GGAATGGGATGGTGTGTC	CCTTACGACTCATACAAACATATG	CS-5
IR94a	3.7	2.4	*pCasper5*^{P}	ATCTCATCAGTTCGAAATTGC	TTTCTACTTTAGCCAACAATCTACTAG	GAAGCGTAAGCATTGTCTAATGC	GGATGTATACTAAACACTTTGTAATCAAAAC	*BACR04F21*
IR94b	6.4	6.1	*pBGRY*$^{\phi C31}$	CAGGGGTATACCGCCTTTGG	CCCATTCCGTGGGCATTCC	TGGACAGGCCGCTGAGG	TTGGTCAGCCAACATCAGC	*BACR04F21*
IR94c	3.9	2.6	*AD1515*^{P}	CATATTTCGCTCTTTCTCTTAATG	TGCCTGTCAGCGTTTAACAAGTG	TCCAGAAAAGTAACCAACAC	CTGTGGGCTGTACTTTGCATTGC	CS-5
	3.3	12	*pBGRY*$^{\phi C31}$	GGGTATACCGCCTTTGG	CGAGCATGTGCCATTGC	AAGCGAAGATGACGATTTCG	CGCGCACATTTCTATGTTGC	*BACR04F21*
IR94d	4.1	none	*pG4PN+*^{P}	CACTGCCGAAAATGAAGAGTAC	CAATCGGTACCTCATAGC	n.a.	n.a.	CS-5
	4.1	5.2	*pBDPR*$^{\phi C31}$	CACTGCCGAAAATGAAGAGTAC	CAATCGGTACCTCATAGC	TTTGTGGGATGGATGCTGG	TAGTGAGCATCGGCAGC	*BACR15B19*
IR94e	5.8	2.3	*pBGRY*$^{\phi C31}$	ACTGCCGAAAATGAAGAGTAC	CCGCTCAGGATCCACTTTGG	TCGCTAGTTTGGCTGCTGG	CCACCTGGAGGAGTGC	*BACR15B19*
IR94f	1.7	8.3	*pBGRY*$^{\phi C31}$	CCACCGCACTTCTGC	TGAAACCTCAGGCAGGAC	GAATGGCAGAATGTGG	GCGGCCGTTTCCGTTTGAG	*BACR15B19*
IR94g	4	6.7	*pBGRY*$^{\phi C31}$	CCACCGCACTTCTGC	CTCCGAGTAATCTGGCATTG	AACAGAAGCGCAGCTC	AGGGAGCGGCCGTTTCC	*BACR15B19*
IR94h	7.5	2.7	*pBGRY*$^{\phi C31}$	CCACCGCACTTCTGC	AGCCTTCTTTATCCTGCACC	GTGGGTCTGGGCACTTCG	AGGGAGCGGCCGTTTCC	*BACR15B19*

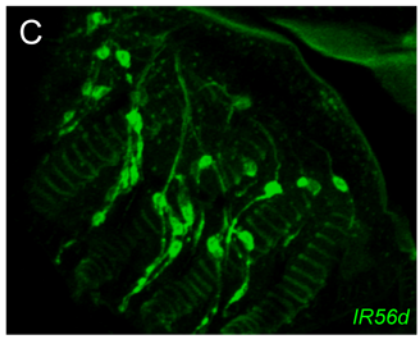
Table S1. *GAL4* drivers for 28 members of the *IR20a* subfamily predicted to encode full-length proteins and 2 pseudogenes (parentheses). To maximize reporter fidelity, constructs for 28 of the 30 genes included both the 5' and 3' flanking regions, and 20 of 30 genes were inserted into matched *phiC31* integration sites in the genome (vectors indicated with superscript ϕC31). Twelve constructs were inserted into random genomic loci by *P* element-mediated transformation (vectors indicated with superscript P). For *IR52a*, *IR67a*, *IR67c*, *IR94c* and *IR94d*, more than one construct was made. Putative promoter/enhancer sequences of *IR-GAL4* constructs were mostly amplified from tiling BACs corresponding to the *Drosophila melanogaster* reference genome, except for seven constructs, which were based on CS-5, a Canton-S strain (Monte et al., 1989). There are no available cDNAs to verify the annotations of *IR20a* family genes, presumably because the genes are expressed at very low levels, and therefore our *GAL4* constructs are based on electronic gene predictions. Vectors are as described in the Extended Experimental Procedures section. In cases where only the 5' flanking region was included in the constructs, a poly(A) signal was provided as either a *hsp70* 3' UTR (in the *pG4PN+* vector) or a *SV40* poly(A) signal (*IR67a-GAL4*). Of the two versions of *IR67a-GAL4*, the first one is based on an earlier version of Flybase annotation (Release 5.12), which contains a premature stop codon in the *IR67a* ORF, and the second one is based on the current version (Release 5.50), which predicts a full-length ORF. Neither version of *IR67a-GAL4* showed expression in central or peripheral neurons. Note also that both versions of *IR94c-GAL4* are based on annotations earlier than Release 5.22, and as such, have 5' flanking regions that begin ~ 0.7 kb upstream of the currently predicted start codon in Release 5.50. The first version of *IR94c-GAL4* did not show expression, but the second one, which is larger, shows expression in the pharynx. We did not make *GAL4* drivers for *IR47b*, *IR48a*, *IR51a*, *IR56e*, *IR60c* and *IR60f*, which contain premature stop codons and are putative pseudogenes. As we have based the *GAL4* drivers for the *IR52* gene cluster on the reference genome, we did not make a construct for *IR52e*, which we discovered in the Canton-S genome at a late phase of this study (see Figure S3A; it is not present in the reference genome).

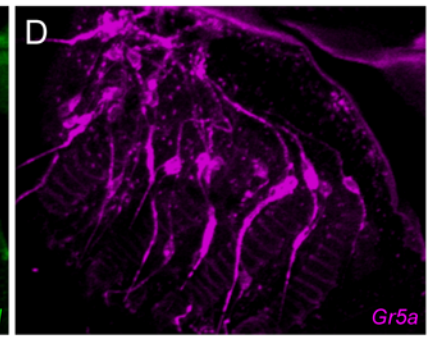
Figure S2, related to Figures 1 and 2

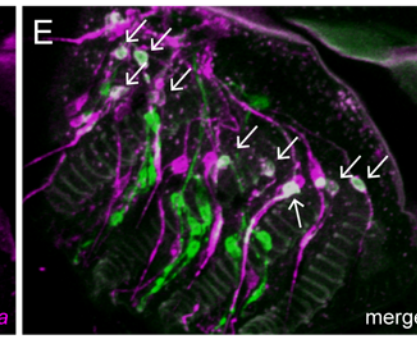


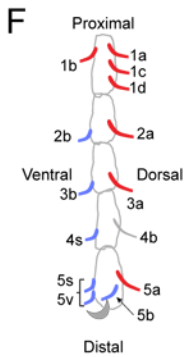
B

	# lines examined	I0	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	L1	L2	L3	L4	L5	L6	L7	L8	L9	pegs
47a	3	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
56a	2	-	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
56b	2	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	+	-	-	+	+	+	+	+	+	+	+	+	-
56d	3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	-	+	-	+	-	+	+	+	+	+	+
94e	2	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	-	+	-	-



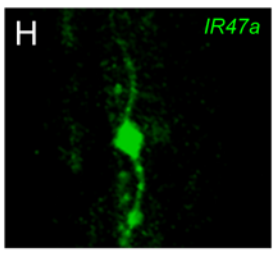


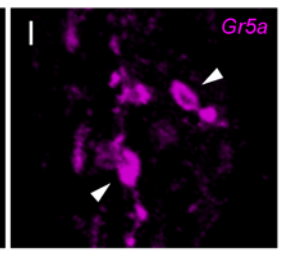


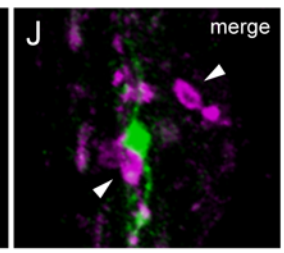


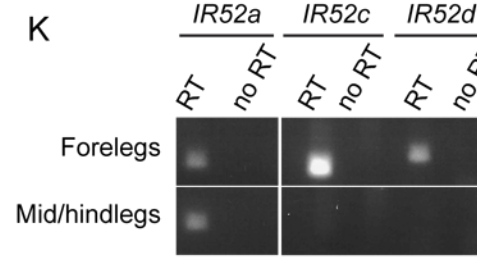
G

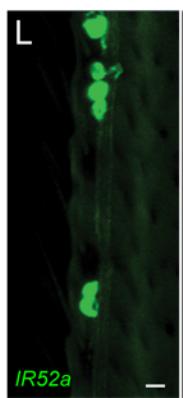
	# lines examined	Foreleg											Midleg										Hindleg									
		Tibia	T1	T2		T3		T4		T5			Tibia	T1	T2		T3	T4		T5			Tibia	T1	T2		T3	T4		T5		
			1a, 1b, 1c, 1d	2a	2b	3a	3b	4b	4s	5a	5b	5v/s		1a, 1c, 1d	2a	2b	3a	4b	4s	5a	5b	5v		1a, 1c, 1d	2a	2b	3a	4b	4s	5a	5b	5v
20a	3	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
47a	4	-	-	-	-	-	-	+	-	+	+	-	-	-	-	-	-	-	+	-	+	+	-	-	-	-	-	-	+	-	+	+	+
52a	2	+	+	+	+	+	+	-	+	-	-	-	+	+	+	+	+	+	-	+	-	-	+	+	+	+	+	+	-	+	-	-	-
52c	2	-	+	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
52d	1	-	+	+	-	+	-	*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
56a	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
56b	2	-	-	-	+	-	+	-	+	-	+	+	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	+	-	+	-	-
56d	2	+	+	+	+	-	-	-	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	+	+	+	+	+	+	+	-	-	-
62a	2	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
94h	2	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

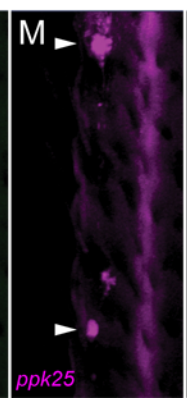


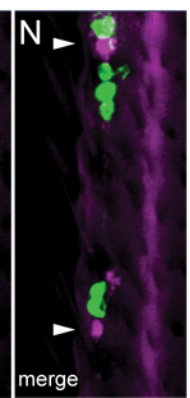












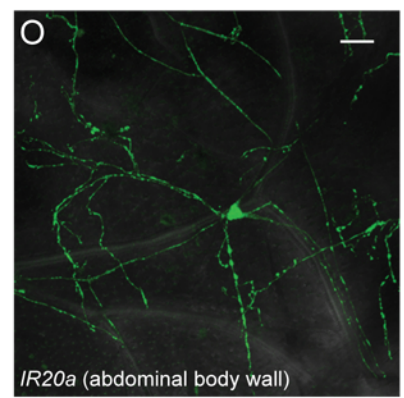
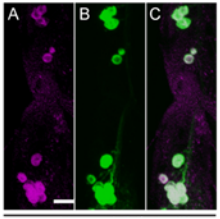
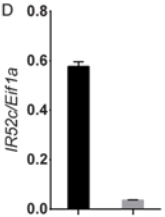


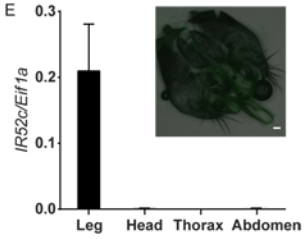
Figure S2. Expression analysis of *GAL4* drivers and transcripts of *IR20a* clade genes.
(A) Illustrations showing taste bristles and taste pegs on the lateral and medial surface of the labellum, respectively. Adapted from Weiss *et al.* (2011) and Falk *et al.* (1976).
(B) Summary of *IR-GAL4* expression in the female labellum. The scoring method is described in Supplemental experimental Procedures. The colors of bars above sensilla subtypes correspond to those of sensilla in (A).
(C-E) Expression of *IR56d-GAL4* (C, E, GFP, green) partially overlaps with that of *Gr5a-LEXA* (D, E, tandem Tomato, magenta) in neurons in labellar bristles. Neurons co-expressing both drivers are labeled in white and indicated with arrows in (E). *IR56d-GAL4* shows additional expression in taste peg neurons, most of which do not express *Gr5a-LEXA*.
(F) Spatially stereotyped taste sensilla on the female foreleg. Blue sensilla respond to food-related tastants, such as sugars, salt and bitter compounds; red sensilla do not (Meunier et al., 2000; Meunier et al., 2003; Miyamoto et al., 2013). Nearly all sensilla depicted exist as symmetrical pairs on both the medial and lateral sides of the foreleg. Sensillum 4b is not colored due to variation in its response to sugars (Ling et al., 2014; Meunier et al., 2000). Sensilla 5v and 5s are highly similar and adjacent to each other; as the present analysis does not provide sufficient resolution to distinguish these two sensilla, we have assigned the drivers expressed in either one to "5v/s". Hence, drivers of *IR56b*, *IR62a* and *IR94h* were expressed in either 5v or 5s, whereas *IR20a* and *IR47a* appear to be expressed in both 5v and 5s. Similarly, in our confocal microscopy preparations we could not reliably visualize 4c, a sensillum that is asymmetrically located only on the lateral side of the foreleg, and hence 4c is not depicted in the cartoon. Male forelegs contain additional sensilla close to 1a-d, 2a, 3a, and 4b. Midlegs and hindlegs are sexually monomorphic.
(G) Summary of *IR-GAL4* expression in female forelegs, midlegs and hindlegs.
(H-J) *IR47a-GAL4* (H, J, GFP, green) and *Gr5a-GAL4* (I, J, RFP, magenta, arrowheads) label distinct populations of cells in legs (the T4 tarsal segment of the male foreleg is shown here; differential labeling is also observed in female legs). In the SOG, *IR47a-GAL4* labels axon projections that resemble those of sugar-sensing neurons, and in the periphery, this driver labels L-type labellar bristles, which respond to sugars but not bitter compounds. Intriguingly, this driver does not show co-expression with a *Gr5a* driver in the legs (H-J); it is possible that *IR47a-GAL4* is expressed in a distinct set of taste neurons mediating attraction, such as those expressing *Gr43a* (Ling et al., 2014; Miyamoto et al., 2012). Attempts to determine whether the axon projections labeled by *IR94e-GAL4* were bona fide sugar-sensing neuron projections were not successful, due to the weak expression of a single copy of *IR94e-GAL4* in our double-labeling experiments. In (I) and (J), the bottom-left arrowhead indicates neurons in sensilla 4s and the top-right arrowhead indicates a neuron in sensilla 4b.
(K) In RT-treated polyA$^+$ RNA extracts from male forelegs, transcripts of *IR52a*, *IR52c* and *IR52d* were detected. Transcripts of *IR52a*, but not *IR52c* and *IR52d*, were detected in the midlegs and hindlegs, which were pooled.
(L-N) *IR52a-LEXA* (L,N, GFP, green) and *ppk25-GAL4* (M, N, RFP, magenta, arrowheads) label distinct populations of cells on the wing margin. Scale bar in L = 5 μm.
(O) Some neurons with wide dendritic arbors on the internal surface of the abdominal wall are labeled by *IR20a-GAL4*. Scale bar represents 20 μm.

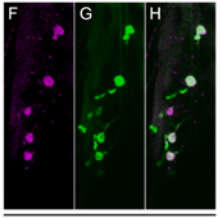
Figure S3, related to Figure 4



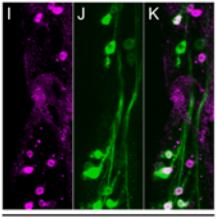
αIR52c v. *IR52c-Gal4*



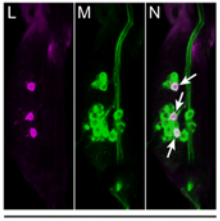




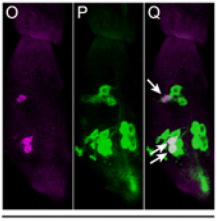
αIR52c v. *IR52d-Gal4*



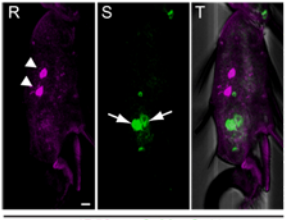
αIR52c v. *IR52a-Gal4*



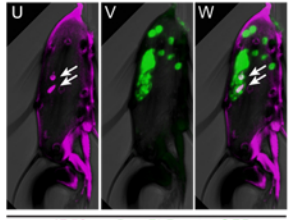
αIR52c v. *IR25a-Gal4*



αIR52c v. *IR76b-Gal4*



αIR52c v. *Gr32a-Gal4*



αIR52c v. *Dsx FLP out > GFP*

Figure S3. Comparison of IR52c protein expression with other markers and *IR52c* transcript levels between sexes and tissues. (A-C) Comparison of anti-IR52c antibody labeling (A, magenta) with GFP (B, green) expression driven by *IR52c-GAL4*. (D-E) Quantitative RT-PCR comparing *IR52c* transcript levels between male and female forelegs (D, n = 3 , $p<0.01$, t-test), and between forelegs, head, thoraces with appendages removed, and abdomen (E, n = 3, $p<0.01$, one-way ANOVA with Dunnet's multiple comparisons test); *IR52c* levels are normalized to *Eif1a*, a ubiquitous transcript that serves as an internal control. Inset in (E) shows the lack of GFP labeling in genitalia of males with *IR52c-GAL4* driving *UAS-mCD8GFP*. (F-W) Comparison of anti-IR52c antibody labeling (magenta in F, I, L, O, R, U) with GFP (green) expression driven by *IR52d-GAL4* (G), *IR52a-GAL4* (J), *IR25a-GAL4* (M), *IR76b-GAL4* (P), *Gr32a-GAL4* (S) and Dsx^{GAL4} *UAS-FLP Act-FRT-CD2-FRT-GAL4* (V). White labeling in merged panels (C, H, K, N, Q, T, W) and arrows in N, Q and W indicate co-labeling by antibodies and GFP. Note that *IR25a-GAL4* and *IR76b-GAL4* are expressed in large numbers of cells in the legs, most or all of which are putative taste neurons. (R-T) anti-IR52c antibodies (R, magenta, arrowheads) and *Gr32a-GAL4* (T, GFP, green, arrows) label distinct cells in the male foreleg tarsal segment 5. (U-W) anti-IR52c antibodies (U, magenta, arrows) label a subset of GFP^+ cells in the foreleg tarsal segment 5 of males of the genotype *UAS-GFP/+; UAS-FLP Act-FRT-CD2-FRT-GAL4/*Dsx^{GAL4}(Rideout et al., 2010); we note that the GFP^+ cells were too numerous in the other leg segments to reliably determine co-labeling of anti-IR52c and GFP. (A-C, F-W) are Z-projections of confocal images. Scale bar in (A) represents 10 μm in (A-C, F-Q), scale bar in inset of (E) represents 25 μm and scale bar in R represents 5 μm in (R-W).

Movie S1. Movie of a male fly tapping the abdomen of a female fly.
Movie is recorded using a Phantom Miro ex4 high speed camera at 1000 frames per second and played back at 1/33th of the original speed. Note that the dorsal surfaces of the male forelegs repeatedly contact the female abdomen, while the male vibrates one of its wings.

Figure S4, related to Figure 7

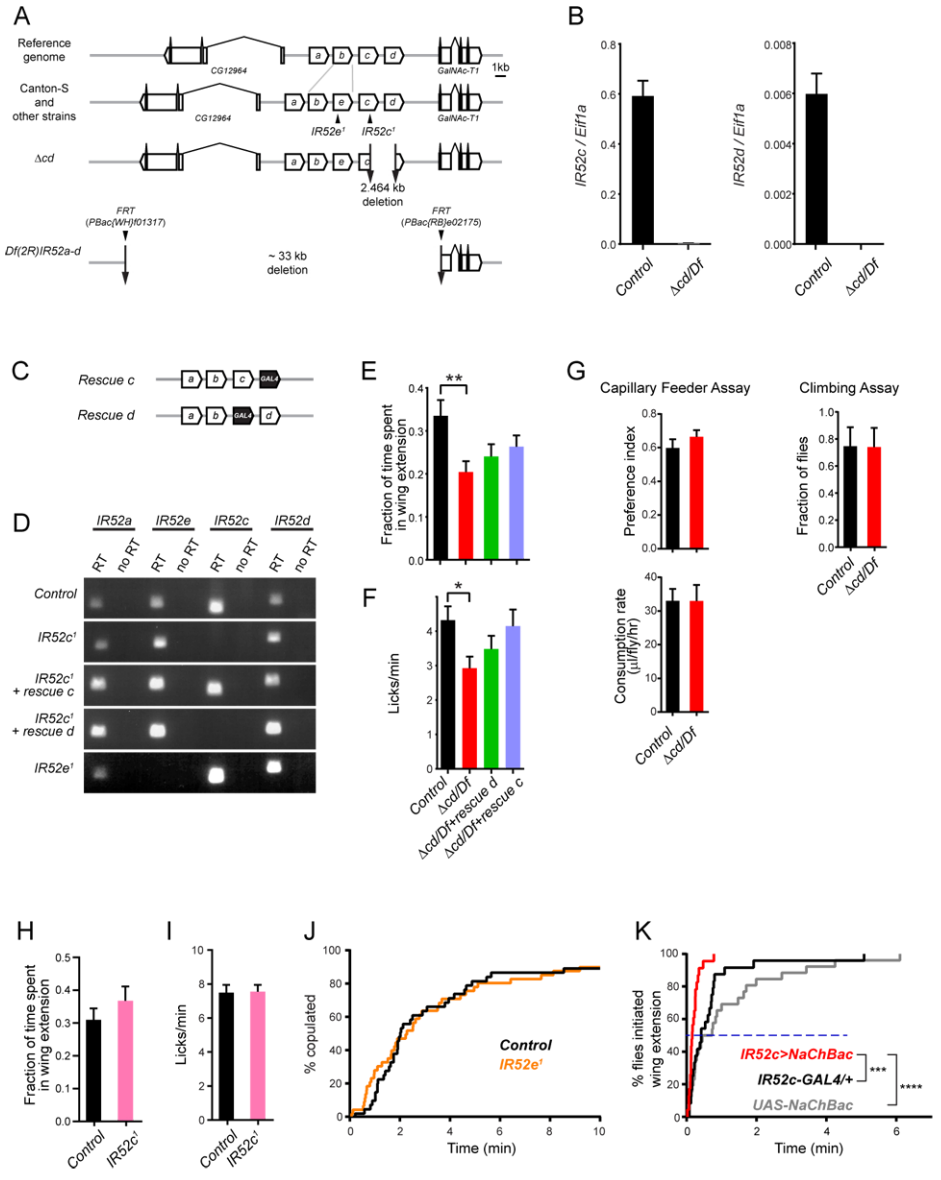


Figure S4. *IR52* cluster organization, expression and function

(A) Genomic region of the *IR52* cluster and two flanking genes. Note that the *D. melanogaster* reference genome contains four genes in the *IR52* cluster (Adams et al., 2000), whereas *Canton-S5*, $IR52c^1$ and $IR52e^1$strains contain five; *IR52b* in the reference genome is an in-frame fusion of the *IR52b* and *IR52e* genes found in *Canton-S5*. $IR52c^1$ and $IR52e^1$are *Minos* transposon insertions, *MB04402* and *MB02231*, respectively. *Δcd* is a deletion generated by the imprecise excision of the *Minos* transposon in $IR52c^1$. *Df(2R)IR52a-d* is a ~ 33kb deficiency generated by *FLP*-mediated site-specific recombination between *FRT* elements in *PiggyBac* transposons, *PBac{WH}f01317* and *PBac{RB}e02175*. The genomic sequence of the *IR52* cluster in the Canton-S line has been deposited in GenBank (KM016699).

(B) Quantitative RT-PCR comparison of *IR52c* (left, n=3, $p<0.001$, t-test) and *IR52d* (right, n=3, $p<0.01$, t-test) transcript levels in male forelegs of controls and *Δcd/Df*. Transcript levels were normalized to *Eif1a*, a ubiquitous transcript that serves as an internal control. Note that transcripts of *IR52c* are approximately a hundred times more abundant than those of *IR52d*.

(C) The genomic constructs *Rescue c* and *Rescue d* are genomic fragments containing the *IR52* cluster corresponding to the *D. melanogaster* reference genome. In *Rescue c*, *IR52d* is replaced by *GAL4*. In *Rescue d*, *IR52c* is replaced by *GAL4*.

(D) RT-PCR analysis of *IR52a*, *IR52e*, *IR52c* and *IR52d* in *Control*, $IR52c^1$, $IR52c^1$*+rescue c*, $IR52c^1$*+rescue d* and $IR52e^1$ foreleg polyA RNA extracts. "RT" denotes poly A RNA extracts that were treated with reverse transcriptase and "no RT" denotes poly A RNA extracts that were not treated with reverse transcriptase; the absence of PCR product in "no RT" controls indicates the absence of genomic DNA contamination.

(E,F) Fraction of time spent in wing extension (E) and frequency of licks (F) in *Control* (black), *Δcd/Df* (red), *Δcd/Df+rescue d* (green) and *Δcd/Df+rescue c* (blue). n=30-36, one-way ANOVA with Dunnet's multiple comparisons test. "*" and "**" denote $p<0.05$ and $p<0.01$, respectively.

(G) *Δcd/Df* males show normal behavioral responses to the taste of sucrose, and normal locomotor function, as judged by CAFE (left, n=28-30, $p>0.05$, t-test, for both top and bottom graphs) and climbing assays (right, n=10-15 trials with 10 flies per trial), respectively.

(H-I) Fraction of time spent in wing extension (H, n = 18-22, $p>0.05$, t-test) and frequency of licks (I, n = 26-27, $p>0.05$, t-test) in *Control* (black) and $IR52c^1$ (pink).

(J) Cumulative plots showing the percentage of males that achieved successful copulation over a 10 min period; the curves for *Control* (black) and $IR52e^1$ (orange) males are not significantly different (n = 35-38, log-rank test).

(K) Activation of $IR52c^+$ neurons by the bacterial sodium channel NaChBac-GFP(Nitabach et al., 2006), led to earlier initiation of unilateral wing extension. Cumulative plots show the percentage of males that has initiated unilateral wing extension. Males carrying three copies of *IR52c-GAL4* and two copies of *UAS-NaChBac-GFP* showed significantly earlier courtship initiation compared to controls containing only *IR52c-GAL4* or *UAS-NaChBac-GFP* (n = 23-26, log-rank test, *** denotes $p < 0.001$ and **** denotes $p < 0.0001$). Blue dashed line indicate the times at which 50% of flies of each genotype have initiated wing extension. Due to the weak expression of the *GAL4* driver, we aged the males for 5-6 weeks to allow accumulation of NaChBac-GFP before performing the experiment.

Figure S5, related to Figure S4

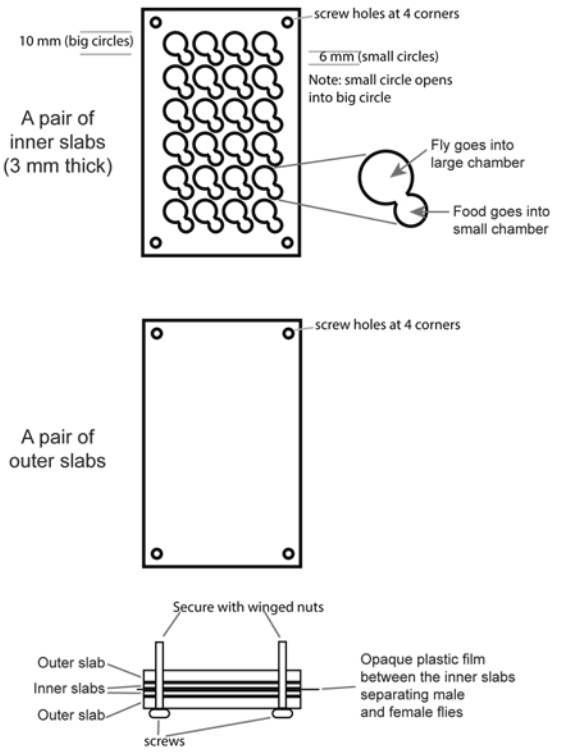


Figure S5. A custom-made multi-mating chamber device.
The device, modified from a previous publication (Drapeau and Long, 2000), consists of a pair of inner Plexiglass slabs with 24 pairs of large and small holes drilled into the plastic (top), and a pair of outer slabs serving as lids over the chambers (middle). The slabs are assembled as depicted at the bottom, with an opaque plastic film between the two inner slabs to separate the male and female flies in each half of the chamber.

Table S2

*Or*s

Name	Sensilla	DoS	
Or47b	Trichoid	0.32	**
Or49a	Basiconic	0.32	**
Or65a	Trichoid	0.20	*
Or65c	Trichoid	0.36	***
Or67a	Basiconic	0.27	**
Or67d	Trichoid	0.18	*

Table S2. Signatures of positive selection among *Ors* (*Odor receptors*).
D. melanogaster Ors compared with their orthologs in *D. simulans* using the McDonald-Kreitman test (MKT) (McDonald and Kreitman, 1991). MKT looks for evidence of natural selection by comparing the amount of variation within species to the divergence between species. DoS stands for Direction of Selection, a value that indicates adaptive evolution when positive and purifying selection when negative (Stoletzki and Eyre-Walker, 2011). For the sake of simplicity, only the *Or*s with signatures of adaptive evolution ($p<0.05$) are shown here. Data were derived from the popDrowser website (Ramia et al., 2011).