

Table S1. Numbers of cysLGIC superfamily gene in characterized insect species.

	Dmel	Amel	Tcas	Nvit	Apis	Aaeg	Atum	Bger	Pame
nAChR									
α type	7	9	11	12	9	10	11	10	9
β type	3	2	1	4	2	4	1	7	10
GABA									
RDL	1	1	1	1	2	1	1	1	1
GRD	1	1	1	1	0	1	1	1	1
LCCH3	1	1	1	1	0	1	0	1	1
Others									
8916	1	1	1	1	0	0	1	2	2
Insect group1 (CLGC)	3	1	3	1	2	1	1	2	2
HisCl	2	2	2	2	2	2	2	2	2
GluCl	1	1	1	1	2	1	1	2	2
pHCl	1	1	1	1	2	1	1	1	1
12344	1	1	1	1	1	0	1	1	1
Total	22	21	24	26	22	22	21	30	32

Species name annotations: Dmel-D. melanogaster; Amel-A. mellifera; Tcas-T. castaneum; Nvit-Nasonia vitripennis; Apis-Acyrtosiphon pisum; Aaeg-Aedes aegypti; Atum-Aethina tumida; Bger-Blattella germanica; Pame-Periplaneta americana.

Table S2. Chromosomal location of CPB cysLGIC subunit genes

Subunit gene	Chromosomal location	CDS start	CDS end
<i>Ldec α1</i>	CM045695.1	67746761	67768291
<i>Ldec α2</i>	CM045696.1	3170836	3138756
<i>Ldec α3</i>	CM045703.1	13858834	13970425
<i>Ldec α4</i>	CM045695.1	41018307	41754641
<i>Ldec α5</i>	CM045698.1	63481056	63396375
<i>Ldec α6</i>	CM045700.1	61872075	62180805
<i>Ldec α7</i>	CM045700.1	49090340	49655565
<i>Ldec α8</i>	CM045702.1	41014845	40916109
<i>Ldec α9</i>	CM045705.1	44836427	44824104
<i>Ldec α10</i>	CM045702.1	7803128	7759587
<i>Ldec β1</i>	CM045700.1	49515645	49597496
<i>Ldec 12344</i>	CM045702.1	23277189	23316372
<i>Ldec 8916</i>	CM045698.1	41150513	41092358
<i>Ldec CLGC</i>	CM045706.1	4769722	4747013
<i>Ldec GluCl</i>	CM045700.1	16070642	16150978
<i>Ldec Grd</i>	CM045702.1	48892562	48874817
<i>Ldec HisCl1</i>	CM045702.1	53884993	53871120
<i>Ldec HisCl2</i>	CM045700.1	51999541	51966458
<i>Ldec Lcch3</i>	CM045698.1	41229977	41274666
<i>Ldec pHCl copy1</i>	CM045712.1	24506564	24420011
<i>Ldec pHCl copy2</i>	CM045712.1	25371690	25459705
<i>Ldec Rdl</i>	CM045709.1	22156746	22310306

Table S3. BLASTn results and the selection of protein sequence for alignment.

Receptor Name	Subunit ID	Number of transcript hits with bit score > 900)	ID of the transcript producing the longest protein isoform*	Completeness of “signature regions” on the longest isoform	Treatment for protein alignment
nicotinic Acetylcholine receptor	Ldec_α1	2	GGNV01022290.1	Complete	No treatment
	Ldec_α2	3	GGNV01160150.1	Complete	No treatment
	Ldec_α3	9	GGNV01295170.1	Complete	N/A
	Ldec_α4	3	GGNV01187321.1	Only TM4 present	Replaced with gDNA-translated protein
	Ldec_α5	1	GGNV01170968.1	Missing Loops D, A, E, and part of Loop B	Replaced with AWC68049.1
	Ldec_α6	6	GGNV01207620.1	Complete	No treatment
	Ldec_α7	1	GGNV01057173.1	Complete	No treatment
	Ldec_α8	1	GGNV01106564.1	Complete	No treatment
	Ldec_α9	1	GGNV01304972.1	Complete	No treatment
	Ldec_α10	4	GGNV01314871.1	Complete	No treatment
	Ldec_β1	1	GGNV01068309.1	TM4 partially missing	Replaced by AKL79440.1
GABA-gated ion channels	Ldec_Rdl	NA	GGNV01394567.1	Missing ligand binding loops and TMs	Replaced by XP_023012203.1
	Ldec_GRD	4	GGNV01130196.1	Complete	No treatment
	Ldec_LCCH3	2	GGNV01135992.1	Complete	No treatment
glutamate-gated chloride channels	Ldec_GluCl	6	GGNV01268739.1	N-terminal loops missing	Replaced by XP_023018980.1
Histamine-gated chloride channels	Ldec_HisCl1	3	GGNV01338955.1	Complete	No treatment
	Ldec_HisCl2	4	GGNV01159996.1	Complete	No treatment
pH-sensitive chloride channel	Ldec_pHCl_copy1 and copy2**	2	GGNV01044124.1	Complete	No treatment
insect group 1 ligand-gated ion channels	Ldec_CLGC	3	GGNV01134008.1	Complete	No treatment
Others	Ldec_8916	1	GGNV01117992.1	Missing Loops D, A, E, B, and F	Replaced by gDNA-translated protein
	Ldec_12344	1	GGNV01048522.1	Complete	No treatment

*All transcripts are found in a transcriptome with the accession number: GGNV00000000.1. XP_023012203.1 and XP_023018980.1 are predicted by automatic computational analysis; AKL79440.1 and AWC68049.1 are deduced from cDNA obtained by mRNA RT-PCR and 5'- & 3'- RACE.

**Protein sequence translated from Ldec_pHCl_copy1 and Ldec_pHCl_copy2 only differ by one amino acid. Transcript GGNV01044124.1 was the best fit for both, with identical amino acid as Ldec_pHCl_copy2 at the polymorphic site. Therefore protein sequence from GGNV01044124.1 was selected to represent both genes and named Ldec_pHCl in the alignment.

Table S4. List of NCBI Genbank (www.ncbi.nlm.nih.gov) accession numbers of sequences used for analysis.

Subunits	Genomic DNA	Transcripts	mRNA	Proteins
Ldec_α1		GGNV01022290.1		
Ldec_α2		GGNV01160150.1*		
Ldec_α3		GGNV01295170.1		
Ldec_α4	CM045695.1**			
Ldec_α5			MF197919.1	AWC68049.1
Ldec_α6		GGNV01207620.1		
Ldec_α7		GGNV01057173.1		
Ldec_α8		GGNV01106564.1		
Ldec_α9		GGNV01304972.1		
Ldec_α10		GGNV01314871.1		
Ldec_β1			KP339870.1	AKL79440.1
Ldec_12344		GGNV01048522.1		
Ldec_8916	CM045709.1**			
Ldec_GRD		GGNV01130196.1		
Ldec_LCCH3		GGNV01135992.1		
Ldec_Rdl				XP_023012203.1
Ldec_HisCl1		GGNV01338955.1		
Ldec_HisCl2		GGNV01159996.1		
Ldec_GluCl				XP_023018980.1
Ldec_pHCl		GGNV01044124.1		
Ldec_CLGC		GGNV01134008.1		
Amel_α1			NM_001098220.1	NP_001091690.1
Amel_α2			NM_001011625.2	NP_001011625.1
Amel_α3			NM_001079561.1	NP_001073029.1
Amel_α4			NM_001098221.1	NP_001091691.1
Amel_α5				AAS75781.1
Amel_α6			DQ026035.1	AAY87894.1
Amel_α7			NM_001011621.1	NP_001011621.1
Amel_α8			NM_001011575.1	NP_001011575.1
Amel_α9			NM_001098224.2	NP_001091694.1
Amel_β1			NM_001079560.1	NP_001073028.1
Amel_β2			NM_001098229.1	NP_001091699.1
Amel_12344			DQ667194.1	ABG75746.1
Amel_6927			XM_006563702.3	XP_006563765.1
Amel_8916			NM_001077822.1	NP_001071290.1
Amel_GRD			DQ667183.1	ABG75735.1
Amel_LCCH3			NM_001077812.1	NP_001071280.1
Amel_Rdl			DQ667182.1	ABG75734.1
Amel_HisCl1			NM_001077811.1	NP_001071279.1
Amel_HisCl2			DQ667188.1	ABG75740.1
Amel_GluCl			NM_001077809.1	NP_001071277.1
Amel_pHCl			NM_001143878.1	NP_001137350.1

Dmel_α1			X07194.1	CAA30172.1
Dmel_α2			NM_079758.3	NP_524482.1
Dmel_α3			Y15593.1	CAA75688.1
Dmel_α4			AJ272159.1	CAB77445.1
Dmel_α5			AF272778.1	AAM13390.1
Dmel_α6			NM_164874.3	NP_723494.2
Dmel_α7			NM_001298507.1	NP_001285436.1
Dmel_β1				P04755.1
Dmel_β2			X55676.1	CAA39211.1
Dmel_β3			NM_080359.4	NP_525098.1
Dmel_12344			NM_001299369.1	NP_001286298.1
Dmel_6927			NM_131966.2	NP_572194.1
Dmel_8916			NM_132862.3	NP_573090.3
Dmel_11340			NM_001316541.1	NP_001303470.1
Dmel_7589			NM_001275057.1	NP_001261986.1
Dmel_GRD			NM_079407.3	NP_524131.1
Dmel_LCCH3			NM_206746.2	NP_996469.1
Dmel_Rdl			M69057.2	AAA28556.1
Dmel_HisCl1			NM_079682.3	NP_524406.1
Dmel_HisCl2			NM_141859.4	NP_650116.2
Dmel_GluCl			AF297500.1	AAG40735.1
Dmel_pHCl			NM_001038936.4	NP_001034025.2
Pame_α1			JQ585634.1	AFJ04793.1
Pame_α2			KP725464.1	AKV94621.1
Pame_α3			KR021292.1	AKR16132.1
Pame_α4			JN390946.1	AFA28129.1
Pame_α5		GFCQ01005211.1		
Pame_α6			JF731243.1	AEA40429.1
Pame_α7			MW201211.1	QQH14653.1
Pame_α8			MW201212.1	QQH14654.1
Pame_α9			MW201214.1	QQH14656.1
Pame_β1			MW201213.1	QQH14655.1
Pame_β2		GFCQ01032711.1		
Pame_β3		GFCQ01027461.1		
Pame_β4		GAWS02039241.1		
Pame_β5		GFCQ01009686.1		
Pame_β6		GFCQ01010089.1		
Pame_β7		GFCQ01012153.1		
Pame_β8		GFCQ01034959.1		
Pame_β9		GBJC01015771.1		
Pame_β10		GFCQ01027794.1		
Pame_12344		GFCQ01007480.1		
Pame_8916		GFCQ01012789.1		
Pame_8916_2			MW206636.1	QQH14658.1

Pame_GRD			MW201215.1	QQH14657.1
Pame_LCCH3		GFCQ01015543.1		
Pame_Rdl			LC171474.1	BAW87781.1
Pame_HisCl1		GFCQ01022534.1		
Pame_HisCl2			MW206637.1	QQH14659.1
Pame_GluCl			LC171470.1	BAW87777.1
Pame_GluCl2		GFCQ01031120.1		
Pame_pHCl		GAWS02040818.1		
Pame_CLGC1		GAWS02050116.1		
Pame_CLGC2		GFCQ01015521.1		
Tcas_α1			NM_001109775.1	NP_001103245.1
Tcas_α2			NM_001109953.2	NP_001103423.1
Tcas_α3			NM_001114298.1	NP_001107770.1
Tcas_α4			NM_001109776.2	NP_001103246.1
Tcas_α5			NM_001109782.1	NP_001103252.1
Tcas_α6			NM_001114301.1	NP_001107773.1
Tcas_α7			NM_001109950.1	NP_001103420.1
Tcas_α8			NM_001109949.1	NP_001103419.1
Tcas_α9			NM_001109954.1	NP_001103424.1
Tcas_α10			NM_001109777.1	NP_001103247.1
Tcas_α11			NM_001114299.1	NP_001107771.1
Tcas_β1			NM_001109948.1	NP_001103418.1
Tcas_12344			NM_001109773.1	NP_001103243.1
Tcas_8916			NM_001109955.1	NP_001103425.1
Tcas_GRD			NM_001114300.1	NP_001107772.1
Tcas_LCCH3			NM_001109781.1	NP_001103251.1
Tcas_Rdl			NM_001114337.1	NP_001107809.1
Tcas_HisCl1			NM_001109952.1	NP_001103422.1
Tcas_HisCl2			NM_001109951.1	NP_001103421.1
Tcas_GluCl			NM_001114303.1	NP_001107775.1
Tcas_pHCl			NM_001114303.1	NP_001107777.1
Tcas_CLGC1			NM_001109780.1	NP_001103250.1
Tcas_CLGC2			NM_001109779.2	NP_001103249.1
Tcas_CLGC3			NM_001109778.1	NP_001103248.1

*Transcript is found in the transcriptome with accession number: GGNV00000000.1

**As indicated in the result section, protein sequences of these two subunits were translated from the manually assembled genomic DNA sequences based on Exonerate results. Accession numbers were linked to the chromosomes from which the genomic DNA was extracted.

Table S5 Exon specific depth of coverage of RNA seq data.

Chromosome ID	Start	End	Exon #	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Average
nAChR $\alpha 4$								
CM045695.1	41018307	41018510	exon1	0.41	0.00	0.68	0.00	0.27
	41206210	41206254	exon2	0.00	0.00	0.00	0.00	0.00
	41240941	41241050	exon3	0.00	0.00	2.61	0.06	0.67
	41445469	41445649	exon4	0.00	1.01	3.92	0.61	1.38
	41550858	41550973	exon5*	0.00	2.26	12.82	0.00	3.77
	41561504	41561672	exon6	0.00	2.75	8.38	0.99	3.03
	41618616	41618752	exon7	0.10	4.74	5.00	1.95	2.95
	41674793	41674939	exon8	1.60	3.69	7.77	2.34	3.85
	41693584	41693718	exon9	3.00	5.50	8.93	1.40	4.71
	41731215	41731452	exon10	0.41	18.21	6.79	2.50	6.98
	41754450	41754638	exon11	1.07	16.47	10.54	1.61	7.42
nAChR $\alpha 1$								
CM045695.1	67746761	67746955	exon1	20.35	3.81	9.39	33.26	16.70
	67749241	67749285	exon2	33.68	4.98	8.80	55.41	25.72
	67749346	67749455	exon3	31.52	3.47	9.58	78.22	30.70
	67751505	67751685	exon4	29.97	4.02	10.97	84.73	32.43
	67754374	67754755	exon5*	23.84	4.99	12.94	96.34	34.53
	67757418	67757712	exon6	29.72	2.15	10.05	121.35	40.82
	67764909	67765149	exon7	17.23	6.30	4.60	107.83	33.99
	67768091	67768291	exon8	27.20	7.22	8.72	132.18	43.83

Exons with an * indicate it encodes the sequence connecting the N-terminal extracellular region and the membrane spanning regions.

```

Tcas_α4  - - - - - MPPSAAETLRAWL L S A L V V H G A V A G N P D A K R L Y D D L L S N Y N K L V R P V V N T S D V L R V C I K L K L S Q L I D V N L K N Q I M T T N L W V E Q S W Y D Y K L R W E P K E Y G G V H M 102
Atum_α4  - - - - - MPPSAAETLRAWL L S A L V V H G A V A G N P D A K R L Y D D L L S N Y N K L V R P V V N T S D V L R V C I K L K L S Q L I D V N L K N Q I M T T N L W V E Q S W Y D Y K L R W E P K E Y G G V H M 102
Pame_α4  - - - - - M V C A G D A L R A W V L S A L V V H G A V A G N P D A K R L Y D D L L S N Y N K L V R P V V N T S D V L R V C I K L K L S Q L I D V N L K N Q I M T T N L W V E Q S W Y D Y K L R W E P K E Y G G V H M 101
Amel_α4  - - - - - M P P I I G E T L R V W F L S A L V V H G A V A G N P D A K R L Y D D L L S N Y N K L V R P V V N T S D V L R V C I K L K L S Q L I D V N L K N Q I M T T N L W V E Q S W Y D Y K L R W E P K E Y G G V K M 102
Dmel_α4  M K F L F E I L R L F I K K P S S G G T L R A W I L S A L M V H G A V A G N P D A K R L Y D D L L S N Y N K L V R P V V N T S D V L K V C I K L K L S Q L I D V N L K N Q I M T T N L W V E Q S W Y D Y K L R W E P K E Y G G V H M 114

Tcas_α4  L H V P S D H I W R P D I V L Y N N A D G N F E V T L A T K A T I Y H Q G L V E W K P P A I Y K S S C E I D V E Y F P F D E Q T C V L K F G S W T Y D G F K V D L R H M D E K A G S N V V D V G V D L S E F Y M S V E W D I L E V P 216
Atum_α4  L H V P S D H I W R P D I V L Y N N A D G N F E V T L A T K A T I Y H Q G L V E W K P P A I Y K S S C E I D V E Y F P F D E Q T C V L K F G S W T Y D G F K V D L R H M D E K A G S N V V E V G V D L S E F Y M S V E W D I L E V P 216
Pame_α4  L H V P S D H I W R P D I V L Y N N A D G S Y E V T I K T K A T V Y Y T G L V V W O P P A V Y K S S C A I D V E F F P Y D V Q T C V L K L G S W T Y D G F K V D L R H M D E K A G S N V V E V G V D L S E F Y M S V E W D I L E V P 215
Amel_α4  L H V P S D H I W R P D I V L Y N N A D G N F E V T L A T K A T I Y H Q G L V E W K P P A I Y K S S C E I D V E Y F P F D E Q T C V L K F G S W T Y D G F K V D L R H M D E K S G S N V V D V G V D L S E F Y M S V E W D I L E V P 216
Dmel_α4  L H V P S D H I W R P D I V L Y N N A D G N F E V T L A T K A T I Y S E G L V E W K P P A I Y K S S C E I D V E Y F P F D E Q T C V L K F G S W T Y D G F K V D L R H M D E Q Q G S N V V A V G V D L S E F Y M S V E W D I L E V P 228

Tcas_α4  A V R N E K F Y T C C D E P Y L D I T F N I T M R R K T L F 246
Atum_α4  A V R N E K F Y T C C D E P Y L D I T F N I T M R R K T L F 246
Pame_α4  A V R N E K F Y T C C D E P Y L D I T F N I T M R R K T L F 245
Amel_α4  A V R N E K F Y T C C D E P Y L D I T F N I T M R R K T L F 246
Dmel_α4  A V R N E K F Y T C C D E P Y L D I T F N I T M R R K T L F 258

```

Figure S1. N-terminal extracellular region protein sequence alignment nAChR $\alpha 4$. NCBI accession number of sequence: Tcas_α4, [NP_001103246.1](#); Dmel_α4, CAB77445.1; Pame_α4, [AFA28128.1](#), and Amel α4, NP_001091691.1