SUPPLEMENTRAY MATERIALS

U12-type Spliceosomal Introns Of Insecta

Jessin Janice, et al.

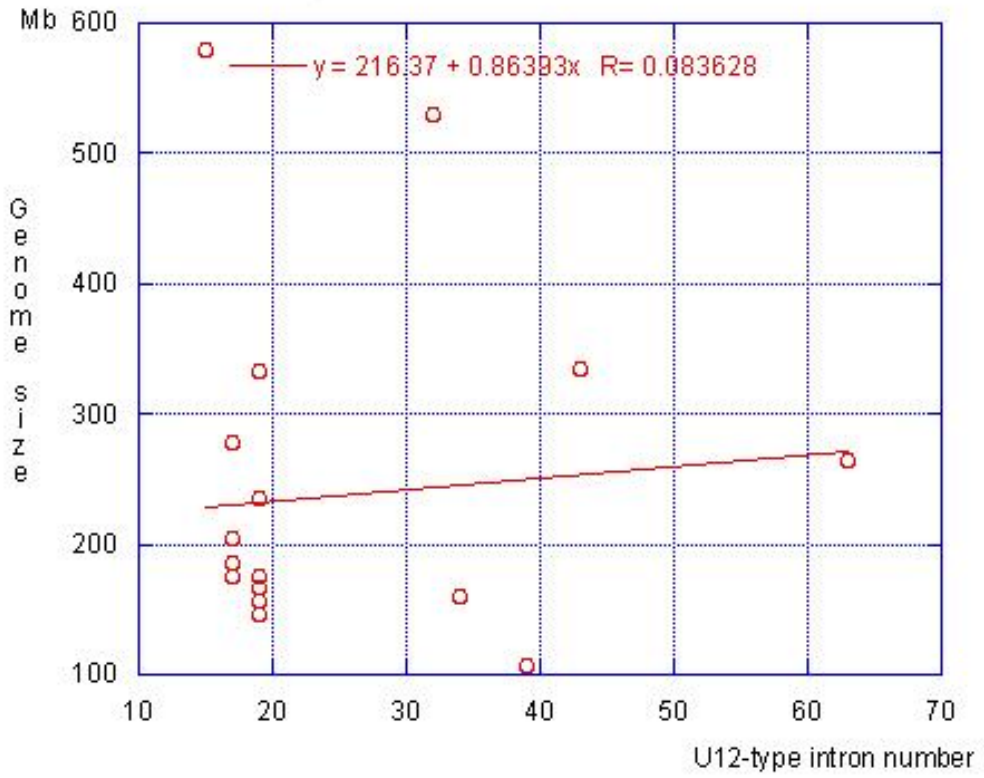
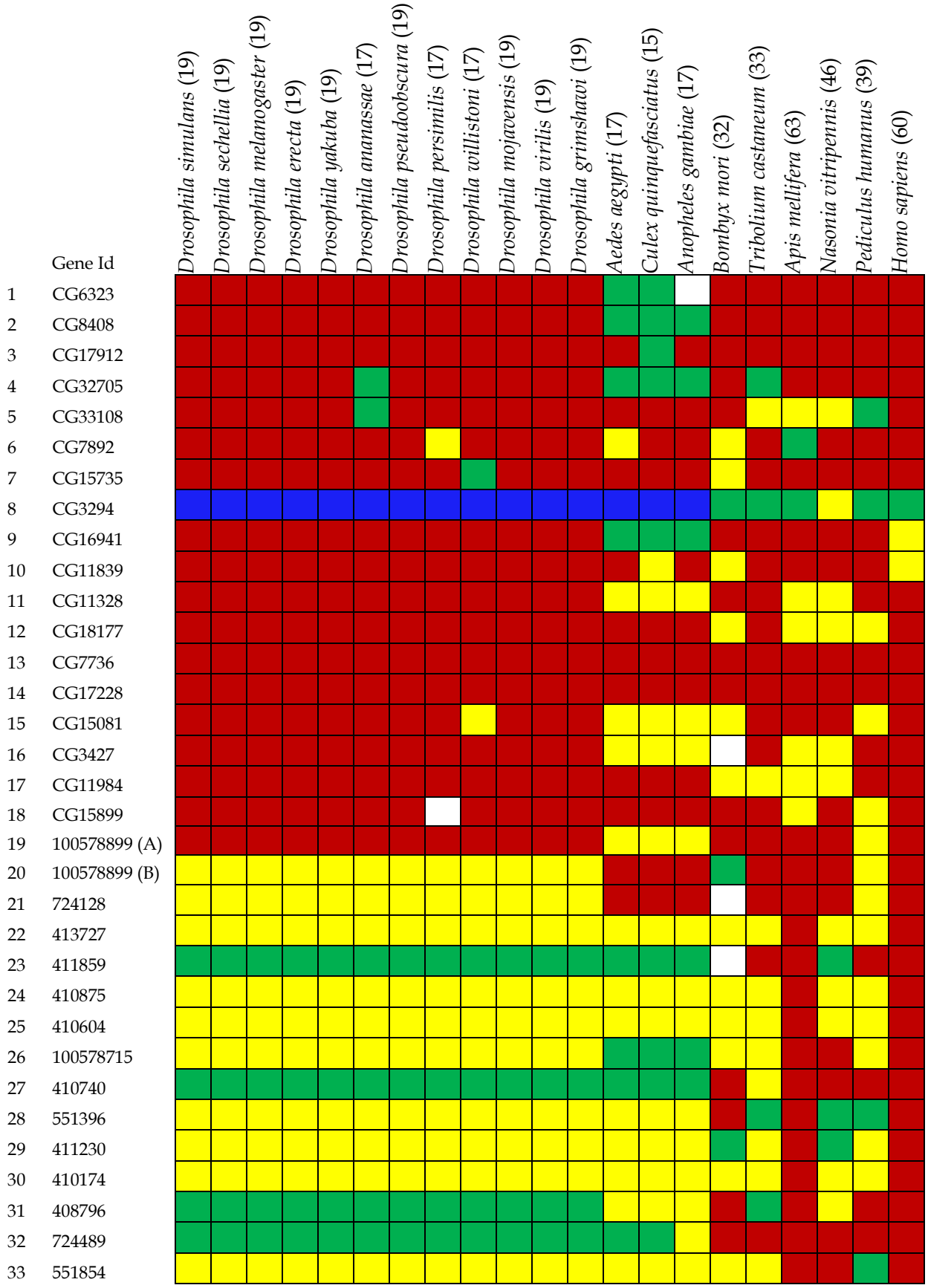


Figure S1. Scatter plot of correlation between genome size and number of U12-type introns.


Gene Id
Drosophila simulans (19)
Drosophila sechellia (19)
Drosophila melanogaster (19)
Drosophila erecta (19)
Drosophila yakuba (19)
Drosophila ananassae (17)
Drosophila pseudoobscura (19)
Drosophila persimilis (17)
Drosophila willistoni (17)
Drosophila mojavensis (19)
Drosophila virilis (19)
Drosophila grimshawi (19)
Aedes aegypti (17)
Culex quinquefasciatus (15)
Anopheles gambiae (17)
Bombyx mori (32)
Tribolium castaneum (33)
Apis mellifera (63)
Nasonia vitripennis (46)
Pediculus humanus (39)
Homo sapiens (60)
1 CG6323
2 CG8408
3 CG17912
4 CG32705
5 CG33108
6 CG7892
7 CG15735
8 CG3294
9 CG16941
10 CG11839
11 CG11328
12 CG18177
13 CG7736
14 CG17228
15 CG15081
16 CG3427
17 CG11984
18 CG15899
19 100578899 (A)
20 100578899 (B)
21 724128
22 413727
23 411859
24 410875
25 410604
26 100578715
27 410740
28 551396
29 411230
30 410174
31 408796
32 724489
33 551854

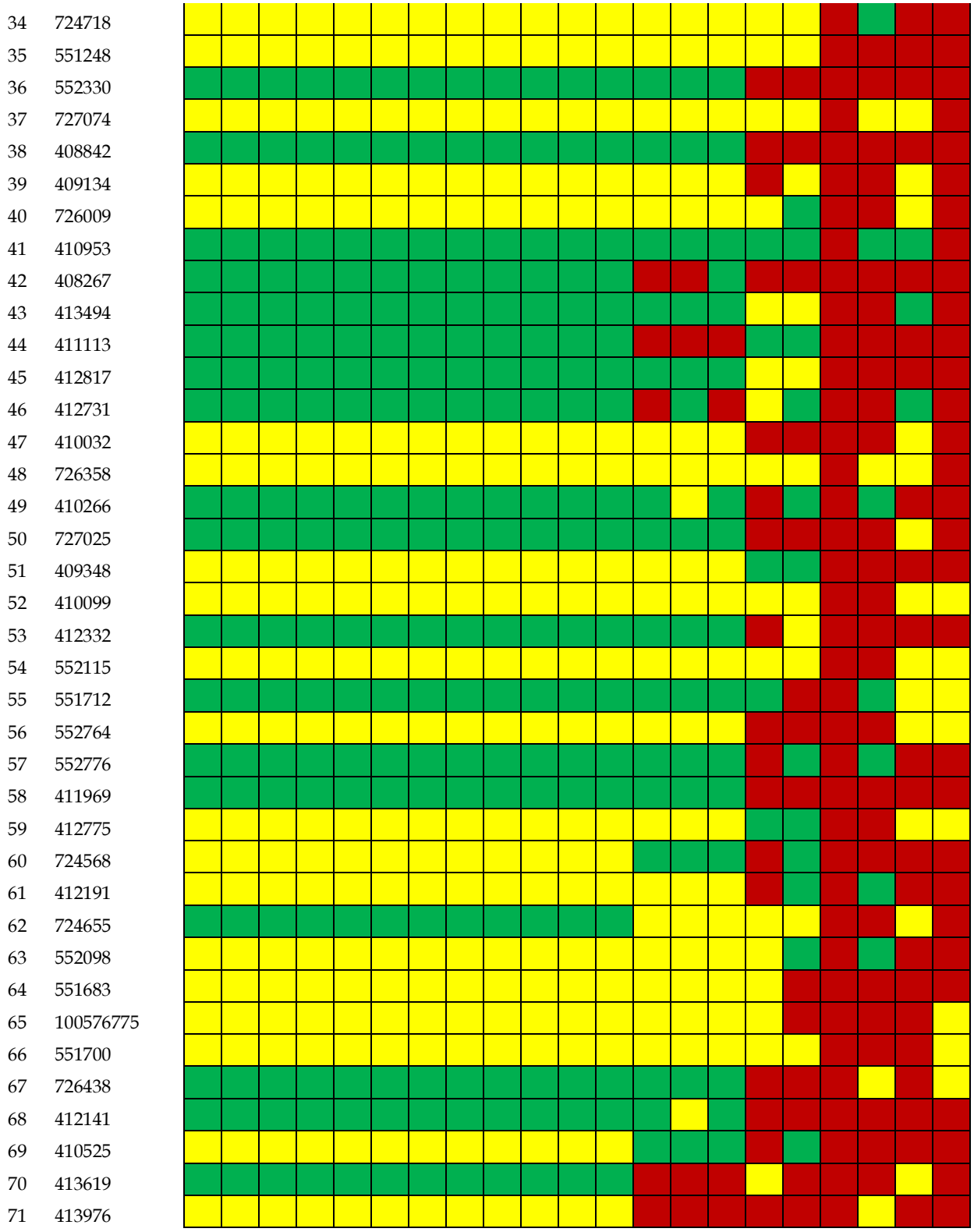


Figure S2. Status of seventy-one U12-type introns in the human and twenty insect genomes. Red – U12-type intron, green – U2-type intron, yellow –intron deletion (absence), blue – U12-type intron gain, and white – data unavailable.

Table S1. Insect genes harboring U12-type introns used in the study.

Seed Genome	Gene number	Annotation Symbol / Gene ID	Termini	Name and Function	Source
Drosophila melanogaster	1	CG6323	GT-AG	*Tsp97E* (Tetraspanin 97E); *TSPAN13* (tetraspanin 13)	U12DB
	2	CG8408	GT-AG	*TMEM41B* (Transmembrane protein 41B)	U12DB
	3	CG17912	GT-AG	*ZNF207* (Zinc finger protein 207)	U12DB
	4	CG32705	GT-AG	*ZDHHC8* (zinc finger, DHHC-type containing 8)	U12DB
	5	CG33108	GT-AG	*C19orf54* (chromosome 19 open reading frame 54)	Lin et al.
	6	CG7892	GT-AG	*nmo* (nemo); *NLK* (nemo-like kinase)	U12DB
	7	CG15735	GT-AG	LSM12, RNA-processing, Lsm domain	Lin et al.
	8	CG3294	GT-AG	*ZRSR2* (zinc finger (CCCH type), RNA-binding motif and serine/arginine rich 2), *URP*; U2AF small subunit related protein ;	U12DB
	9	CG16941	GC-AG	*SF3A1* (splicing factor 3a, subunit 1, 120kDa)	U12DB
	10	CG11839	AT-AC	*CCDC16* (coiled-coil domain containing 16); ZNF830 (zinc finger protein 830)	U12DB
	11	CG11328	AT-AC	*Nhe3*; *SLC9A7* (solute carrier family 9 (sodium/hydrogen exchanger), member 7)	U12DB
	12	CG18177	AT-AC	FLJ14154 hypothetical protein	U12DB
	13	CG7736	AT-AC	*Syx6* (Syntaxin 6); *STX6* (syntaxin 6)	U12DB
	14	CG17228	AT-AC	*pros* (prospero); *PROX1* (prospero homeobox 1)	U12DB
	15	CG15081	AT-AC	*l(2)03709* (lethal (2) 03709); *PHB2* (prohibitin 2)	U12DB
	16	CG3427	AT-AC	*Epac*; *RAPGEF3* (Rap guanine nucleotide exchange factor (GEF)3)	U12DB
	17	CG11984	AT-AC	*KCMF1* (potassium channel modulatory factor 1)	U12DB
	18	CG15899	AT-AC	*Ca-a1T*, calcium channel activity	U12DB

Apis mellifera	19	100578899 (A)	GT-AG	*Ca-a1D* (Ca2+-channel protein α1 subunit D); *CACNA1D* (calcium channel, voltage-dependent, L type, alpha 1D subunit)	U12DB
	20	100578899 (B)	GT-AG	*Ca-a1D* (Ca2+-channel protein α1 subunit D); *CACNA1D* (calcium channel, voltage-dependent, L type, alpha 1D subunit)	U12DB
	21	724128	AT-AC	*p16-ARC* (ARP2/3 complex 16 kDa subunit) (p16-ARC)	Mount et al.
	22	413727	GT-AG	*pit/pitchoune* (ATP-dependent RNA helicase activity)	Mount et al.
	23	411859	GT-AG	*Pten* (PDZ domain binding)	U12DB
	24	410875	GT-AG	CG6712 (RNA binding)	U12DB
	25	410604	GT-AG	*tws/ twins* (protein serine/threonine phosphatase activity)	U12DB
	26	100578715	GT-AG	*wls/ wntless* (Wnt receptor signaling pathway)	U12DB
	27	410740	GT-AG	CG15814 (protein/zinc ion binding)	U12DB
	28	551396	GT-AG	*lethal (1)* G0095 (binding)	U12DB
	29	411230	GT-AG	*Zpr1* (protein/zinc ion binding)	U12DB
	30	410174	GT-AG	*Cyclin T* (cyclin-dependent protein kinase regulator activity)	U12DB
	31	408796	GT-AG	CG5625 (biological process vesicle-mediated transport)	U12DB
	32	724489	GT-AG	CG15735 (ubiquitin like)	U12DB
	33	551854	GT-AG	*PpD3*- Protein phosphatase D3(protein serine/threonine phosphatase activity; metal ion binding)	U12DB
	34	724718	GT-AG	*Ry- Rosy* (xanthine dehydrogenase activity; iron-sulfur cluster binding; electron carrier activity)	Mount et al.
	35	551248	GT-AG	*CG14222*, acyl-CoA N-acyltransferase activity; N-acetyltransferase activity	U12DB

36	552330	GT-AG	*Nhe3*- Na+/H+ hydrogen exchanger 3 (sodium:hydrogen antiporter activity)	U12DB
37	727074	AT-AC	*Spase 22/23* Spase 22/23-subunit (Pepetidases activity)	U12DB
38	408842	GT-AG	*Trn*- Transportin (protein transmembrane transporter activity)	Mount et al.
39	409134	GT-AG	*Tcp1-like* (hydrogen-exporting ATPase activity, phosphorylative mechanism; ATP binding)	U12DB
40	726009	GT-AG	CG5941 (protein binding; zinc ion binding)	U12DB
41	410953	GT-AG	*Mio*-Mlx interactor(transcription factor activity)	U12DB
42	408267	GT-AG	*Tomosyn* (syntaxin-1 binding)	U12DB
43	413494	GT-AG	CG5608 (binding)	U12DB
44	411113	GT-AG	CG5594 (potassium:chloride symporter activity; amino acid transmembrane transporter activity)	U12DB
45	412817	GT-AG	CG32164 (protein transporter activity; binding)	U12DB
46	412731	GT-AG	CG3573 (inositol-polyphosphate 5-phosphatase activity; inositol trisphosphate phosphatase activity.)	U12DB
47	410032	GT-AG	*Trxr-1*- Thioredoxin reductase-1(thioredoxin-disulfide reductase activity; antioxidant activity; NADPH / FADbinding)	Mount et al.
48	726358	GT-AG	*htt*- huntingtin(microtubule binding; protein binding; binding)	Mount et al.
49	410266	GT-AG	*RfC3*- Replication factor C subunit 3 (DNA binding; ATP binding; DNA clamp loader activity)	Mount et al.
50	727025	GT-AG	*DISC1*- Disrupted in schizophrenia 1 protein	U12DB
51	409348	GT-AG	*l(2)gl*- lethal (2) giant larvae(myosin II binding; myosin binding; protein	Mount et al.

			binding)	
52	410099	AT-AC	unknown	Mount et al.
53	412332	GT-AG	*MED23*- Mediator complex subunit 23(RNA polymerase II transcription mediator activity)	Mount et al.
54	552115	AT-AC	*FK506-bp2*- FK506-binding protein 2(FK506 binding; peptidyl-prolyl cis-trans isomerase activity)	Mount et al.
55	551712	GT-AG	CG8446 (lipoyltransferase activity)	Mount et al.
56	552764	GT-AG	CG7414 (regulation of translation; ribosome assembly; proteolysis)	Mount et al.
57	552776	GT-AG	*Srp68* (7S RNA binding; mRNA binding)	Mount et al.
58	411969	GT-AG	*rump*- rumpelstiltskin (mRNA 3'-UTR binding; mRNA binding; nucleotide binding)	Mount et al.
59	412775	GT-AG	*kl-2*- male fertility factor kl2	Mount et al.
60	724568	GT-AG	*CalpC*- Calpain C (Calcium-dependent cysteine-type endopeptidase activity; calcium ion binding)	Mount et al.
61	412191	GT-AG	*dia*- Diaphanous (Actin binding; Rho GTPase binding)	Mount et al.
62	724655	GT-AG	*Set* (cyclin binding)	Mount et al.
63	552098	GT-AG	*Syx8*- Syntaxin 8 (SNAP receptor activity)	Mount et al.
64	551683	GT-AG	(CG2021) - nuclear mRNA splicing, via spliceosome	Mount et al.
65	100576775	GT-AG	CG9004 (protein binding)	Mount et al.
66	551700	GT-AG	CG4159 (pseudouridylate synthase activity; pseudouridine synthase activity; RNA binding)	Mount et al.
67	726438	GT-AG	*Cep97* (protein phosphatase type 1 regulator activity; protein binding)	Mount et al.

	68	412141	GT-AG	unknown	Mount et al.
	69	410525	GT-AG	*PrBP*- Prenyl-binding protein (3',5'-cyclic-GMP phosphodiesterase activity; protein binding)	Mount et al.
Anopheles gambiae	70	1278356	GT-AG	cg14216 (phosphoprotein phosphatase activity)	U12DB
	71	1277345	GT-AG	*Derlin-2*, unknown	U12DB

Table S2. Functional categories over-represented in genes containing the U12-type introns in *Insecta*.

GO category	**Genes in GO category**
Ion Transport	FBGN0040297, FBGN0028703, FBGN0026787, FBGN0029846, FBGN0001991, FBGN0039465, FBGN0036141, FBGN0039172, FBGN0030850, FBGN0024921, FBGN0042177
WD40 repeat-like	FBGN0030412, FBGN0037135, FBGN0002121, FBGN0004889
Nucleic acid binding	FBGN0029833, FBGN0037135, FBGN0038464, FBGN0032408, FBGN0025140, FBGN0038811, FBGN0032244, FBGN0032600, FBGN0032940, FBGN0260010, FBGN0004595, FBGN0003308, FBGN0020653, FBGN0001313, FBGN0011817, FBGN0031628, FBGN0260450, FBGN0032059, FBGN0005777, FBGN0023508, FBGN0039172, FBGN0004889, FBGN0031054, FBGN0026379, FBGN0036039, FBGN0031043, FBGN0025455, FBGN0029846, FBGN0085421, FBGN0013954, FBGN0034089
Transcription Regulation	FBGN0020653, FBGN0025455, FBGN0036141, FBGN0001991, FBGN0002121, FBGN0004889, FBGN0004595, FBGN0037135, FBGN0014879, FBGN0032940, FBGN0034795
Transmembrane	FBGN0029846, FBGN0039465, FBGN0001991, FBGN0039172, FBGN0040297, FBGN0028703, FBGN0026787, FBGN0030850, FBGN0036141, FBGN0039172, FBGN0002121
Ion binding	FBGN0003308, FBGN0031628, FBGN0260450, FBGN0005777, FBGN0032600, FBGN0030873, FBGN0001991, FBGN0085478, FBGN0030096, FBGN0037655, FBGN0026379

Transferase activity	FBGN0025455, FBGN0036039, FBGN0085421, FBGN0034089, FBGN0031043, FBGN0011817, FBGN0020653, FBGN0037135, FBGN0001313, FBGN0039172, FBGN0025140, FBGN0031054, FBGN0038811, FBGN0003308, FBGN0032059, FBGN0005777, FBGN0260450, FBGN0032244, FBGN0029846, FBGN0023508, FBGN0013954, FBGN0004889, FBGN0026379
Phosphatase/ Hydrolase activity	FBGN0032059, FBGN0005777, FBGN0023508, FBGN0004889, FBGN0031054, FBGN0026379, FBGN0260450, FBGN0032244, FBGN0037135, FBGN0001313, FBGN0039172, FBGN0025140, FBGN0020653, FBGN0031043, FBGN0036039, FBGN0011817, FBGN0038811, FBGN0003308, FBGN0025455, FBGN0029846, FBGN0085421, FBGN0013954, FBGN0034089

Table S3. Functional categories over-represented in genes containing the U12-type introns only in *Diptera*.

GO term	**Genes in GO category**
Ion Transport	FBGN0029846, FBGN0039465, FBGN0001991, FBGN0028703, FBGN0030850
Transferase	FBGN0036039, FBGN0029846, FBGN0023508, FBGN0085421, FBGN0011817, BGN0031054
Ion binding	FBGN0030412, FBGN0001991, FBGN0038438, FBGN0031437, FBGN0037655, FBGN0011817, FBGN0004595, FBGN0031628, FBGN0032600, FBGN0023508, FBGN0030364, FBGN0037084, FBGN0038464
Nucleic acid binding	FBGN0031628, FBGN0030412, FBGN0032600, FBGN0023508, FBGN0030364, FBGN0038438, FBGN0031437,FBGN0037084, FBGN0038464, FBGN0001991, FBGN0037655, FBGN0011817, FBGN0004595

Table S4. Total and U12-type intron number in selected genomes.

Organism	Total number of introns	U12-type introns
Apis mellifera	54,818	63
Drosophila melanogaster	38,539	19
Anopheles gambiae	36,186	17
A. aegypti	41,678	17
N. vetripennis	86,395	46
Tetraodon nigroviridis	187,875	392
H. sapiens	207,344	695
A. thaliana	70,379	309