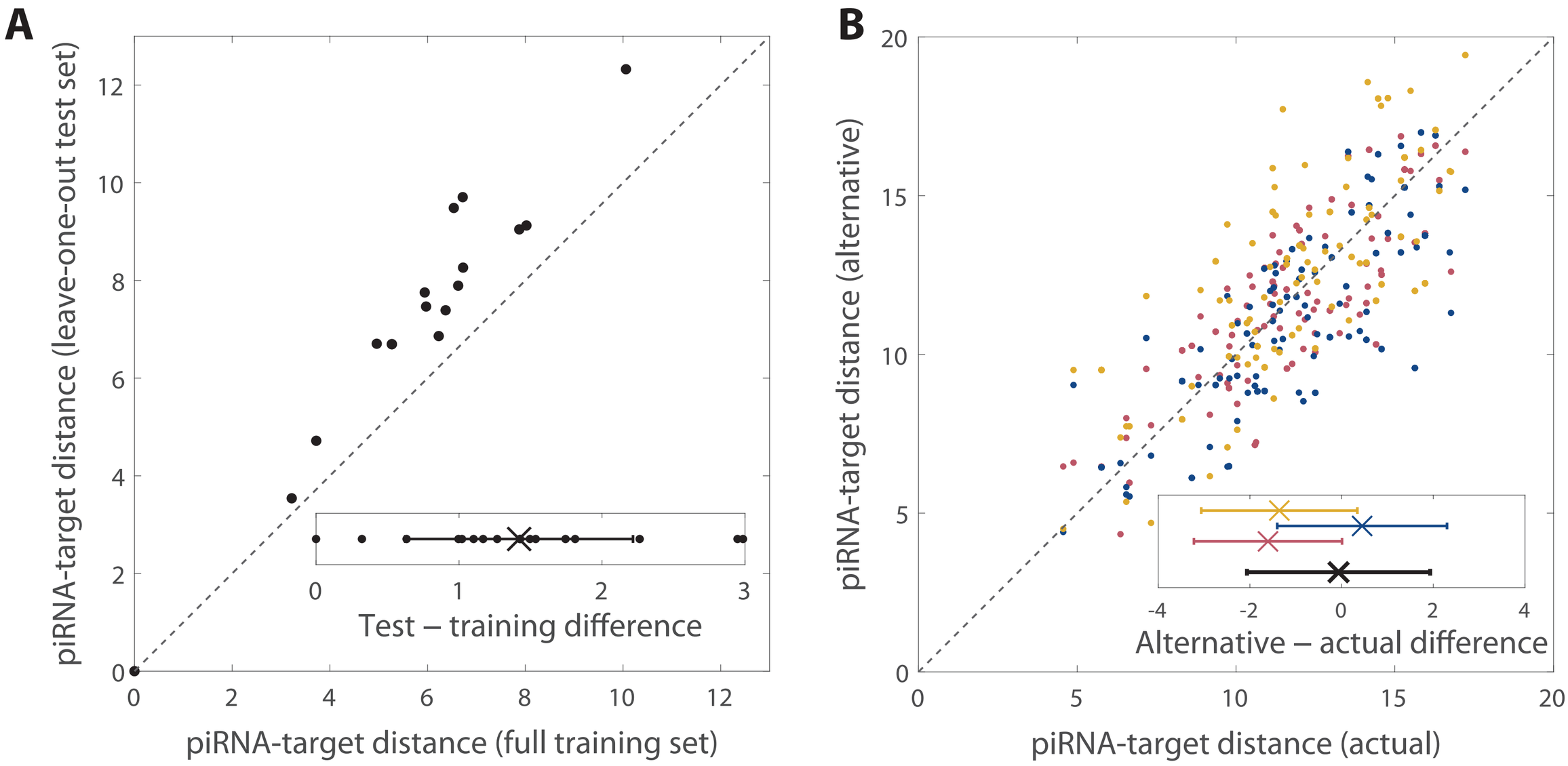
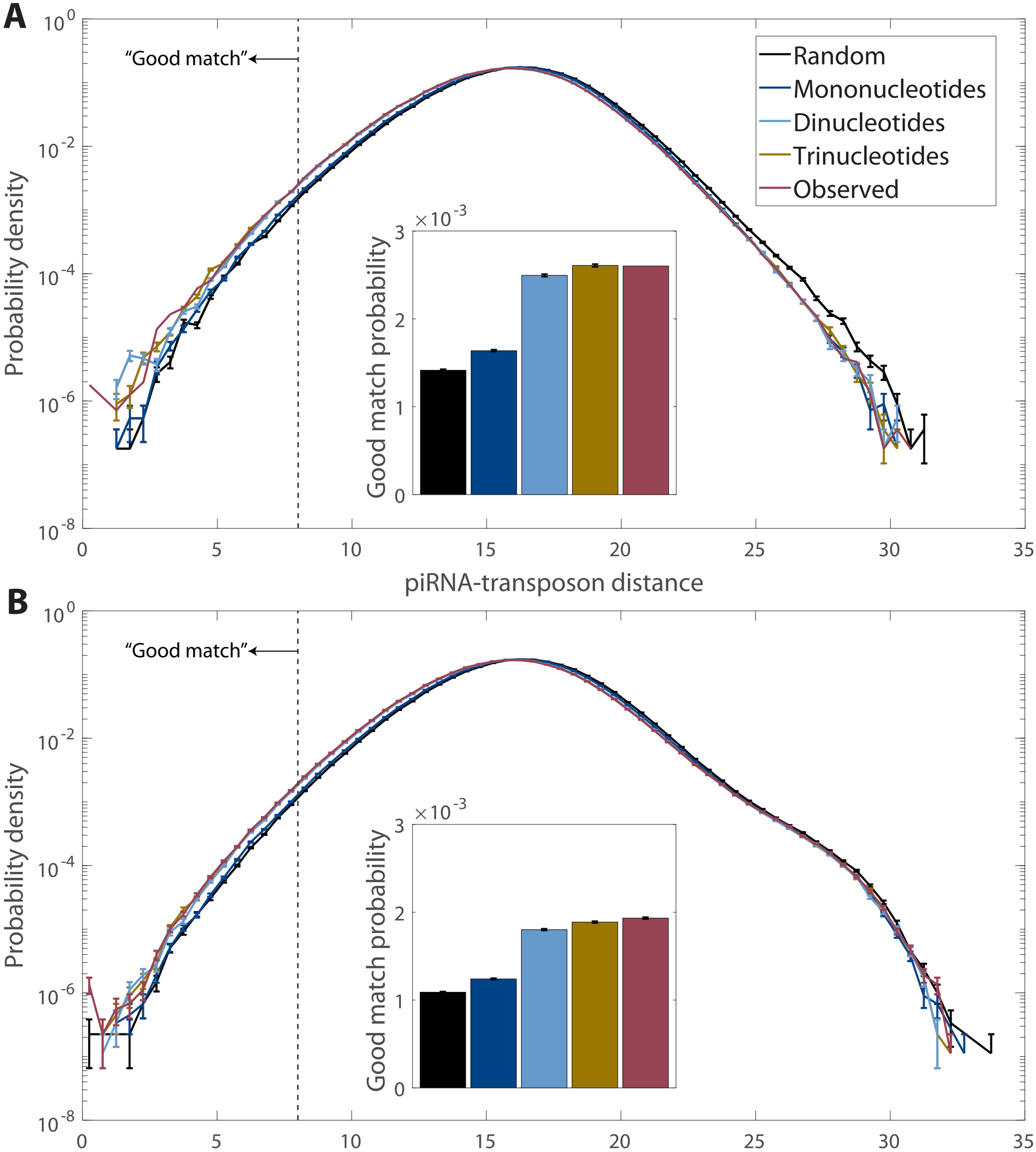
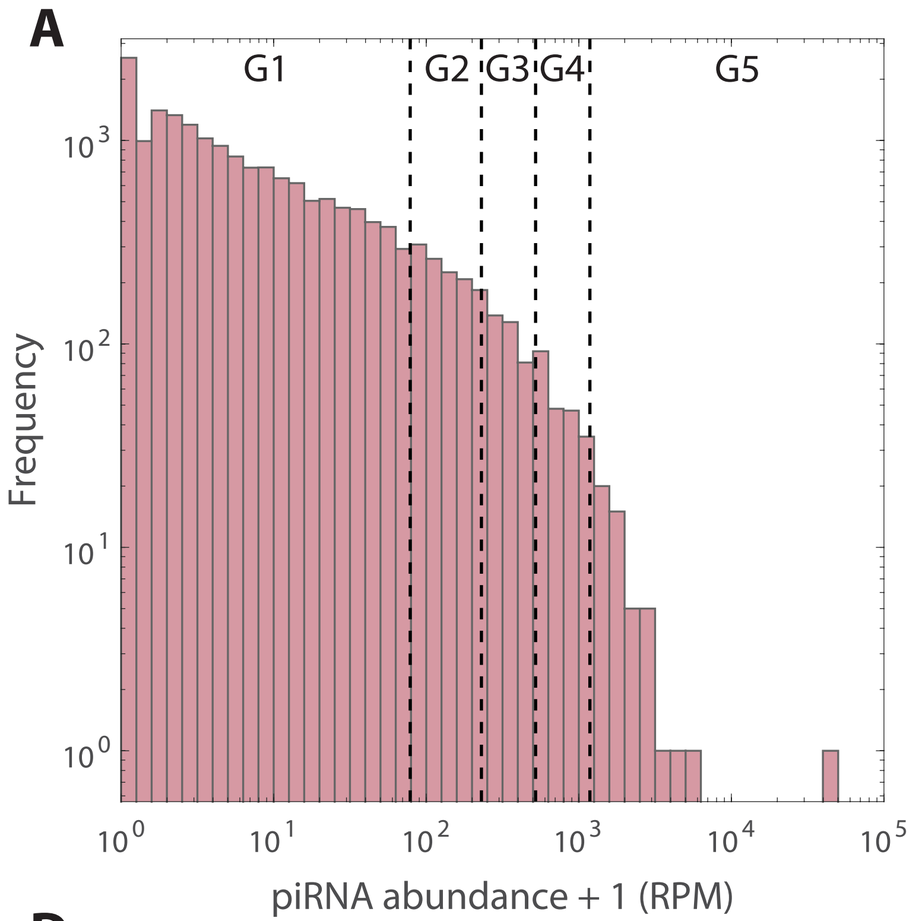
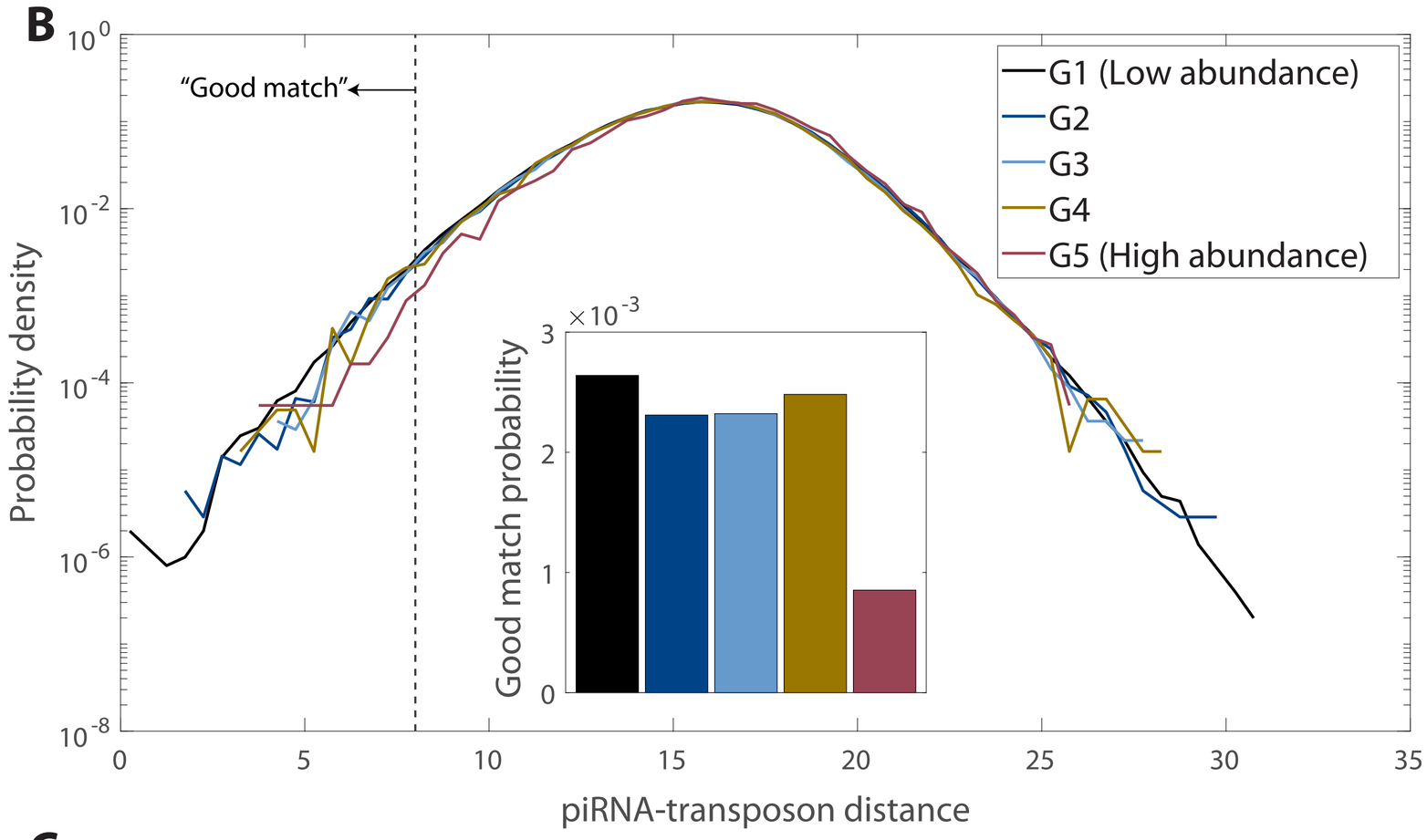


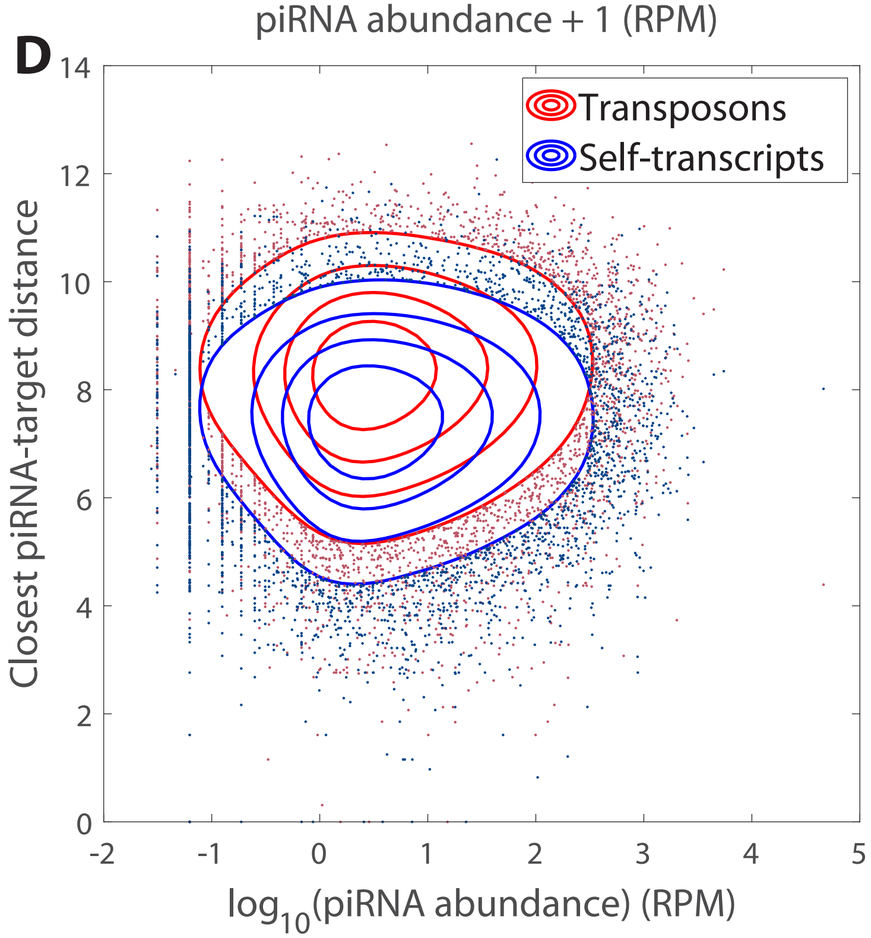
Supplementary Figure 1. Cross-validation of distance metric. **(A)** Leave-one-out cross-validation. Comparison between piRNA-target distances of experimentally validated piRNA-target pairs using the full training set (the piRNA distance utilized in the rest of the paper, incorporating all known pairs) or leave-one-out test sets (where the pair being measured is omitted when calculating the log-odds probabilities). The dashed line indicates perfect equivalence. Inset: the difference between the distance calculated between each piRNA-target pair in the leave-one-out test set and in the full training set, for each of the 17 experimentally confirmed piRNA-target pairs. The cross and error bar indicate mean and standard deviation. **(B)** Cross-validation using putative piRNA-target pairs in piRTarBase, identified by CLASH and differential expression after PRG-1 knockout. Alternative piRNA-target distances were constructed using 17 randomly selected putative piRNA-target pairs, out of a total $n = 354$. The distances between the remaining 337 putative piRNA-target pairs were measured using both an alternative distance metric and the actual distance used in the paper, and the results of the two distance metrics were compared. Different colors correspond to three different random choices of 17 putative pairs; the dashed line shows perfect equivalence. Inset: the mean and standard deviation of the difference between the alternative and actual piRNA-target distance. Colored bars show the three random choices of 17 putative pairs used in the scatterplot, while the black bar shows the mean and standard deviation across 100 random choices of sets of 17 putative pairs.

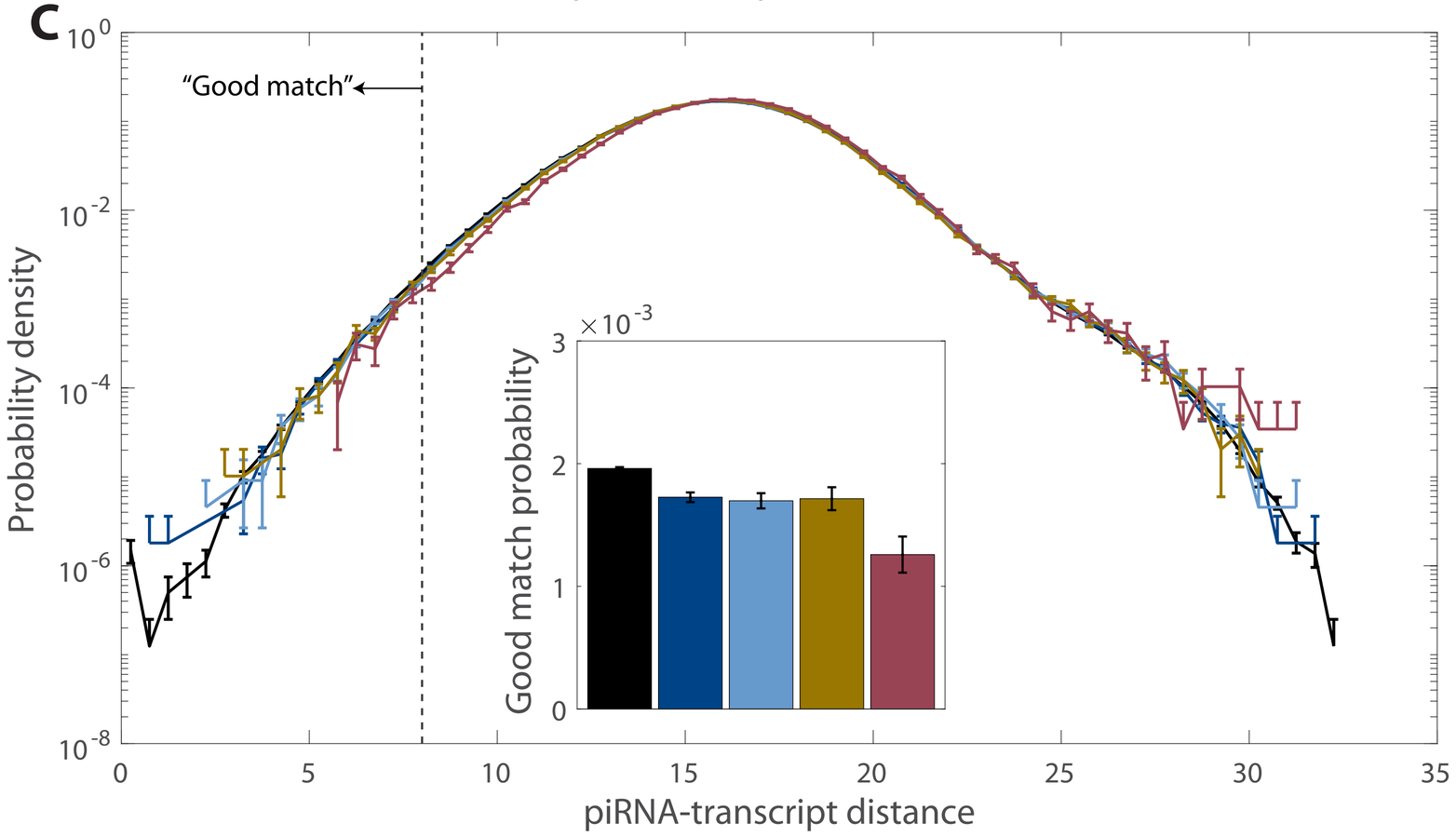


Supplementary Figure 2. Comparison between real and random C. elegans piRNA-transposon and piRNA-transcript pairing, with random piRNAs generated preserving mono-, di-, and trinucleotide probabilities across each piRNA. **(A)** Probability distribution of the closest-match piRNA-transposon distance for every pair of C. elegans piRNAs (*n* = 17,849) and transposons (*n* = 627), similar to Fig. 4A. In addition to real piRNA sequences (red) and fully random piRNA sequences (black), results are shown for random piRNA sequences which maintain the position-specific mononucleotide (dark blue), dinucleotide (light blue), and trinucleotide (yellow) probabilities of real piRNAs. Inset: proportion of piRNA-transposon pairs where the closest-matching site has a piRNA-target distance of less than 8 (left of the dashed vertical line in the main panel), for each set of piRNAs. Error bars indicate counting error. **(B)** Same as (A), but for a random sample of all C. elegans transcripts (*n* = 1,000) rather than transposons.









Supplementary Figure 3. Analysis of C. elegans piRNA abundance data in the context of piRNA-transposon and piRNA-target pairing. **(A)** Histogram of piRNA abundance (in reads per million) on a log-log scale for the 17,833 piRNAs with abundance data. Dashed lines show the division of piRNAs into five groups sorted by abundance, such that each group contains the same total number of reads per million. **(B)** Probability distribution of the closest-match piRNA-transposon distance for every pair of these C. elegans piRNAs (n = 17,833) and transposons (n = 627), with piRNAs divided into each of the five abundance-ordered groups in (A). Inset: proportion of piRNA-transposon pairs where the closest-matching site has a piRNA-target distance of less than 8 (left of the dashed vertical line in the main panel), for each set of piRNAs. **(C)** Same as (B), but for a random sample of all C. elegans transcripts (n = 1,000) rather than transposons. Error bars indicate counting error. **(D)** Combined contour- and scatter-plot showing the abundance of each piRNA (n = 17,833), along with the piRNA-target distance between it and its closest match on a transposon (red) or self-transcript (blue). Each contour curve is a region with equal density of observations in the scatterplot, with higher density in the inner contours and lower density in the outer contours. Outside the outermost contour, individual observations are plotted.

Supplementary Table 1: References to RepBase Reports transposon consensus sequences

Organism	Author	Name	Year	Volume	Issue	Page
C. brenneri	Kojima,K.K. and Jurka,J.	NeSL-1_CBre	2013	13	1	760
C. brenneri	Kojima,K.K. and Jurka,J.	NeSL-2_CBre	2013	13	1	764
C. brenneri	Kojima,K.K. and Jurka,J.	NeSL-3_CBre	2013	13	1	766
C. brenneri	Kojima,K.K. and Jurka,J.	TE-1_CBre	2013	13	1	769
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-1_CBre	2018	18	3	354
C. brenneri	Kojima,K.K. and Jurka,J.	RTE-1_CBre	2018	18	3	357
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-2_CBre	2018	18	3	366
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-3_CBre	2018	18	3	368
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-4_CBre	2018	18	3	370
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-5_CBre	2018	18	3	372
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-6_CBre	2018	18	3	374
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-7_CBre	2018	18	3	376
C. brenneri	Kojima,K.K. and Jurka,J.	CR1-8_CBre	2018	18	3	378
C. brenneri	Kojima,K.K. and Jurka,J.	RTE-2_CBre	2018	18	3	383
C. brenneri	Kojima,K.K. and Jurka,J.	RTE-3_CBre	2018	18	3	384
C. briggsae	Feschotte,C. and Wessler,S.R.	Merlin1_CB	2003	3	5	92
C. briggsae	Pavlicek,A. and Jurka,J.	MERLIN5_CB	2004	4	6	185
C. briggsae	Kapitonov,V.V. and Jurka,J.	Mirage1_CB	2004	4	6	186
C. briggsae	Kapitonov,V.V. and Jurka,J.	MirageN1a_CB	2004	4	6	187
C. briggsae	Kapitonov,V.V. and Jurka,J.	MirageN1b_CB	2004	4	6	188
C. briggsae	Pavlicek,A. and Jurka,J.	HAT10_CB	2004	4	8	204
C. briggsae	Pavlicek,A. and Jurka,J.	HAT6_CB	2004	4	8	205
C. briggsae	Pavlicek,A. and Jurka,J.	HAT7_CB	2004	4	8	206
C. briggsae	Pavlicek,A. and Jurka,J.	HAT8_CB	2004	4	8	207
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER38A_CB	2004	4	8	208
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER38B_CB	2004	4	8	209
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER38C_CB	2004	4	8	210
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER40_CB	2004	4	8	211
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER41_CB	2004	4	8	212
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER42_CB	2004	4	8	213
C. briggsae	Pavlicek,A. and Jurka,J.	MUDR4_CB	2004	4	8	214
C. briggsae	Pavlicek,A. and Jurka,J.	MUDR5A_CB	2004	4	8	215
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY1-I_CB	2004	4	9	225
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY1-LTR_CB	2004	4	9	226
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY2-I_CB	2004	4	9	227
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY2-LTR_CB	2004	4	9	228
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY3-I_CB	2004	4	9	229
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY3-LTR_CB	2004	4	9	230
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY4-I_CB	2004	4	9	231
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY4-LTR_CB	2004	4	9	232
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY5-I_CB	2004	4	9	233
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY5-LTR_CB	2004	4	9	234
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY6-I_CB	2004	4	9	235
C. briggsae	Pavlicek,A. and Jurka,J.	GYPSY6-LTR_CB	2004	4	9	236
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER43_CB	2004	4	9	238
C. briggsae	Pavlicek,A. and Jurka,J.	MARINER44B_CB	2004	4	9	239

C. briggsae	Pavlicek,A. and Jurka,J.	MARINER44_CB	2004	4	9	240
C. briggsae	Pavlicek,A. and Jurka,J.	MERLIN6_CB	2004	4	9	241
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON10_CB	2004	4	10	263
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON11_CB	2004	4	10	264
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON12_CB	2004	4	10	265
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON13_CB	2004	4	10	266
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON14_CB	2004	4	10	267
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON1_CB	2004	4	10	268
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON2_CB	2004	4	10	269
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON3_CB	2004	4	10	270
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON4_CB	2004	4	10	271
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON5_CB	2004	4	10	272
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON6_CB	2004	4	10	273
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON7_CB	2004	4	10	274
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON8_CB	2004	4	10	275
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON9A_CB	2004	4	10	276
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON9B_CB	2004	4	10	277
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON9C_CB	2004	4	10	278
C. briggsae	Pavlicek,A. and Jurka,J.	HELITRON9D_CB	2004	4	10	279
C. briggsae	Pavlicek,A. and Jurka,J.	DIRS-1_CB	2005	5	4	82
C. briggsae	Pavlicek,A. and Jurka,J.	MuDR-10_CB	2005	5	4	85
C. briggsae	Pavlicek,A. and Jurka,J.	MuDR-12A_CB	2005	5	4	86
C. briggsae	Pavlicek,A. and Jurka,J.	MuDR-13_CB	2005	5	4	87
C. briggsae	Pavlicek,A. and Jurka,J.	MuDR-14_CB	2005	5	4	88
C. briggsae	Pavlicek,A. and Jurka,J.	MuDR-15_CB	2005	5	4	89
C. briggsae	Pavlicek,A. and Jurka,J.	MuDR-8_CB	2005	5	4	90
C. briggsae	Pavlicek,A. and Jurka,J.	MUDR-9_CB	2005	5	4	91
C. briggsae	Kapitonov,V.V. and Jurka,J.	Mariner-N1_CB	2005	5	5	123
C. briggsae	Pavlicek,A. and Jurka,J.	HAT14_CB	2005	5	7	177
C. briggsae	Pavlicek,A. and Jurka,J.	HAT15_CB	2005	5	7	178
C. briggsae	Pavlicek,A. and Jurka,J.	HAT16_CB	2005	5	7	179
C. briggsae	Pavlicek,A. and Jurka,J.	HAT17_CB	2005	5	7	180
C. briggsae	Pavlicek,A. and Jurka,J.	DNA2-3_CB	2006	6	3	115
C. briggsae	Pavlicek,A. and Jurka,J.	DNA8-1_CB	2006	6	3	116
C. briggsae	Pavlicek,A. and Jurka,J.	DNA8-2_CB	2006	6	3	117
C. briggsae	Pavlicek,A. and Jurka,J.	DNA8-3_CB	2006	6	3	118
C. briggsae	Kojima,K.K. and Jurka,J.	NeSL-1_CBri	2013	13	1	761
P. pacificus	Jurka,J.	DNA-1_PPac	2010	10	7	952
P. pacificus	Jurka,J.	DNA-2B_PPac	2010	10	7	953
P. pacificus	Jurka,J.	DNA-2_PPac	2010	10	7	954
P. pacificus	Jurka,J.	DNA-3_PPac	2010	10	7	955
P. pacificus	Jurka,J.	DNA-4_PPac	2010	10	7	956
P. pacificus	Jurka,J.	DNA-5_PPac	2010	10	7	957
P. pacificus	Jurka,J.	Mariner-1_PPc	2010	10	7	958
P. pacificus	Jurka,J.	Mariner-2_PPc	2010	10	7	959
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-1_PPc-LTR	2010	10	7	995
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-2_PPc-I	2010	10	7	996

P. pacificus	Jurka,J. and Kohany,O.	Gypsy-2_PPc-LTR	2010	10	7	997
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-3_PPc-I	2010	10	7	998
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-3_PPc-LTR	2010	10	7	999
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-4_PPc-I	2010	10	7	1000
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-4_PPc-LTR	2010	10	7	1001
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-5_PPc-I	2010	10	7	1002
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-5_PPc-LTR	2010	10	7	1003
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-6_PPc-I	2010	10	7	1004
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-6_PPc-LTR	2010	10	7	1005
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-7_PPc-I	2010	10	7	1006
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-7_PPc-LTR	2010	10	7	1007
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-8_PPc-I	2010	10	7	1008
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-8_PPc-LTR	2010	10	7	1009
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-9_PPc-I	2010	10	7	1010
P. pacificus	Jurka,J. and Kohany,O.	Gypsy-9_PPc-LTR	2010	10	7	1011
P. pacificus	Kojima,K.K. and Jurka,J.	RTE-1_PPac	2010	10	7	1060
P. pacificus	Kojima,K.K. and Jurka,J.	RTE-2_PPac	2010	10	7	1061
P. pacificus	Jurka,J.	RTE-3_PPac	2010	10	7	1062
P. pacificus	Jurka,J.	RTE-4_PPac	2010	10	7	1063
P. pacificus	Jurka,J.	RTE-5_PPac	2010	10	7	1064
P. pacificus	Jurka,J.	RTE-6_PPac	2010	10	7	1065
P. pacificus	Jurka,J.	RTE-7_PPac	2010	10	7	1066
P. pacificus	Jurka,J.	RTE-10_PPac	2018	18	3	358
P. pacificus	Jurka,J.	RTE-11_PPac	2018	18	3	359
P. pacificus	Jurka,J.	RTE-12_PPac	2018	18	3	360
P. pacificus	Jurka,J.	RTE-13_PPac	2018	18	3	361
P. pacificus	Jurka,J.	RTE-14_PPac	2018	18	3	362
P. pacificus	Jurka,J.	CR1-2_PPac	2018	18	3	363
P. pacificus	Jurka,J.	CR1-1_PPac	2018	18	3	365
P. pacificus	Jurka,J.	RTE-8_PPac	2018	18	3	385
P. pacificus	Jurka,J.	RTE-9_PPac	2018	18	3	386
P. pacificus	Jurka,J.	CR1-3_PPac	2018	18	3	387
P. pacificus	Jurka,J.	CR1-4_PPac	2018	18	3	388
P. pacificus	Jurka,J.	CR1-5_PPac	2018	18	3	389
P. pacificus	Jurka,J.	CR1-6_PPac	2018	18	3	390
P. pacificus	Jurka,J.	CR1-7_PPac	2018	18	3	391
P. pacificus	Jurka,J.	CR1-8_PPac	2018	18	3	392
P. pacificus	Jurka,J.	CR1-2B_PPac	2018	18	3	393