

Whitefly genomes contain ribotoxin coding genes acquired from plants

Walter J. Lapadula*, María Laura Mascotti, Maximiliano Juri Ayub*

IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, Ejército de los Andes 950, D5700HHW, San Luis, Argentina.

*Correspondence and requests for materials should be addressed to W.J.L (e-mail: wlapadula@gmail.com) or M.J.A. (e-mail: mjuriayub@hotmail.com)

Supplementary information

Supplementary Table 1. Contig region containing BRIP genes. Annotated proteins GenPept cdes of *B. robocri* are shown along with their location in different genome databases. The highest identity hit in BLASTP searches for each protein is indicated, along with the identity percent and coverage percent.

Annotation	Arrow code Figure 1 A	GenPept	Closest BLASTP	Organism	Id (%)/QC (%)	Orientation according to NW_017547285	NW_017547285			PGTPO1000858			ML134445		
							<i>Bemisia tabaci</i> MEAM1			<i>Bemisia tabaci</i> SSA1			<i>Bemisia tabaci</i> MCO-Q		
							Initiation	End	Gen length (bp)	Initiation	End	Gen length (bp)	Initiation	End	Gen length (bp)
elongation of very long chain fatty acids protein 7-1	A	XP_018902197	VVC28064	<i>Cinara cedri</i>	48/93	-	985766	992173	6407	199829	206692	6863	165303	172259	6956
elongation of very long chain fatty acids protein 7-1	B	XP_018902200	XP_015364617	<i>Diuraphis noxia</i>	51/97.7	-	993246	1009513	16267	186383	196288	11905	179925	187284	7359
ribosome-inactivating protein	C	XP_018902206	XP_002619608	<i>Helicoverpa zea</i>	37/58.9	-	1027989	1027989	3284	NA	NA	NA	NA	152684	2590
rabenosyn-5	D	XP_018902267	XP_014271881	<i>Halymorpha halys</i>	50/94.4	+	1034969	1044873	9904	152706	162699	9993	123566	133452	9886
protein Y1P6	E	XP_018902269	XP_023722963	<i>Cryptotermes secundus</i>	67/98.3	-	1046825	1051805	4980	146162	151093	4931	116371	121278	4908
TFIID subunit 1-like	F	XP_018902266	PNF40462	<i>Cryptotermes secundus</i>	66/99.7	+	1052478	1080660	27882	118682	145465	26783	89677	115706	26029
IMiR fusion partner-like 3	G	XP_018902290	XP_023715865	<i>Cryptotermes secundus</i>	71/100	-	1087652	1089807	2135	105245	111375	2130	79739	81871	2132
ubiquitin carboxyl-terminal hydrolase 16	H	XP_018902226	XP_030036242	<i>Manduca sexta</i>	34/100	+	1093723	1108974	15251	88677	105921	17244	60327	76392	16065
coiled-coil domain-containing protein 93 isoform X1	I	XP_018902228	XP_022184911	<i>Nilaparvata lugens</i>	47/98.6	-	1113316	1128262	14946	69713	83924	14211	40005	54622	14617
hoxy-6-polyphenyl-1,4-benzoquinol methylase, mitochondrion	J	XP_018902233	RZF39936	<i>Laelophax striatellus</i>	70/92.1	+	1128922	1129800	878	68198	69074	878	18484	19359	878
2-phosphorylase phosphatase 1	K	XP_018902210	XP_023717146	<i>Cryptotermes secundus</i>	45/98.6	-	1131522	1135731	3609	63248	64880	3632	31851	36346	4455
TBC1 domain family member 13	L	XP_018902284	XP_023712160	<i>Cryptotermes secundus</i>	72/96.9	-	1143805	1152910	9105	46521	53455	6934	20863	NA	NA
ribosome-inactivating protein luffaculin 1-like	M	XP_018902288	AAR15568	<i>Gynostemma pentaphyllum</i>	34/64.4	+	1158736	1161290	2554	37269	40465	3196	10307	12611	2304
HM/homeobox protein Awb isoform X1	N	XP_018902192	RZF48072	<i>Laelophax striatellus</i>	78/93.9	+	1202084	1240031	109647	NA	NA	NA	NA	NA	NA
ubiquitin carboxyl-terminal hydrolase isozyme L3	O	XP_018902286	XP_023724885	<i>Cryptotermes secundus</i>	90.7	-	1248855	1359235	8340	NA	NA	NA	NA	NA	NA
uncharacterised protein CD1161 isoform X1	P	XP_018902286	XP_024214830	<i>Halymorpha halys</i>	60/97.5	+	1358825	1363863	5038	NA	NA	NA	NA	NA	NA

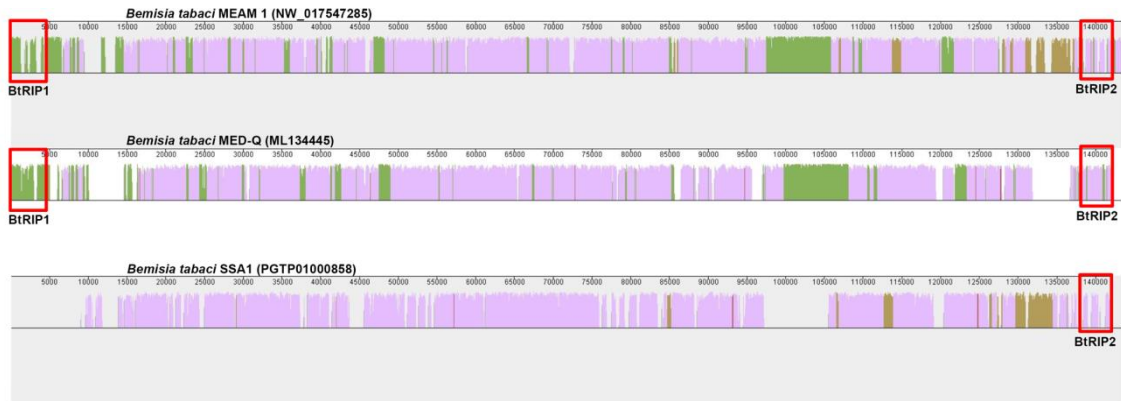


Figure S1. Scheme depicting the *B. tabaci* genomic region harboring *BtRIP1* and *BtRIP2* genes. The contig region containing *RIP* genes from *B. tabaci* MEAM 1 (NW017547285), *B. tabaci* MED/Q (ML134445) and *B. tabaci* SSA-ECA (PGTP01000858) were compared using MAUVE software [1]. Sections of the similarity plot colored in violet are conserved among all three genomes; portions in green are conserved among the MEAM 1 and MED-Q sequences, segments conserved among MEAM 1 and SSA 1 are colored in brown. Empty spaces represent regions with no sequence available. Regions corresponding to *BtRIP1* and *BtRIP2* are indicated with red boxes.

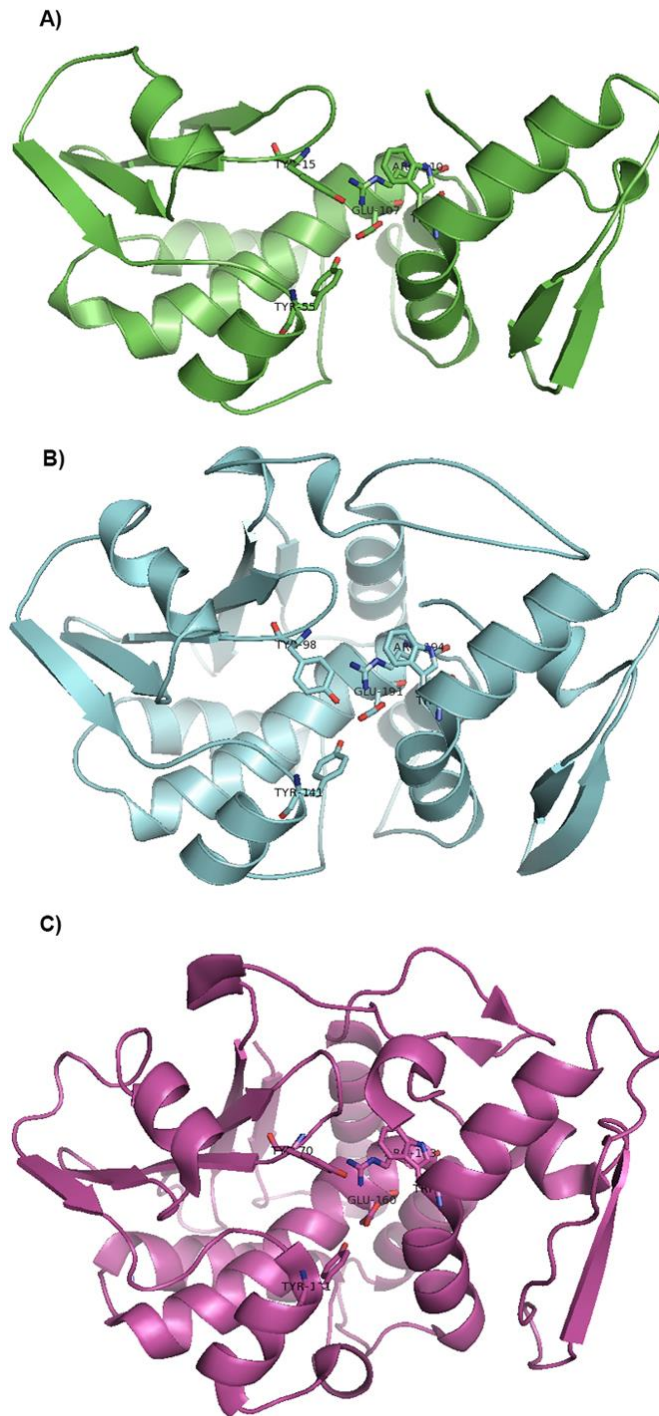


Figure S2. Homology models of BtRIP1 and BtRIP2. Models were constructed in Swiss-Model server [2]. Structures of momordin (PDB 3MY6 or 4YP2) displaying 30-32% sequence identity were used as templates for both BtRIP1 and BtRIP2. **A.** Homology model of BtRIP1 (green). **B.** Homology model of BtRIP2 (cyan) **C.** Structure of momordin PDB 3MY6 (magenta). Active site residues (Y, Y, E, R, W) are numbered and shown as sticks.

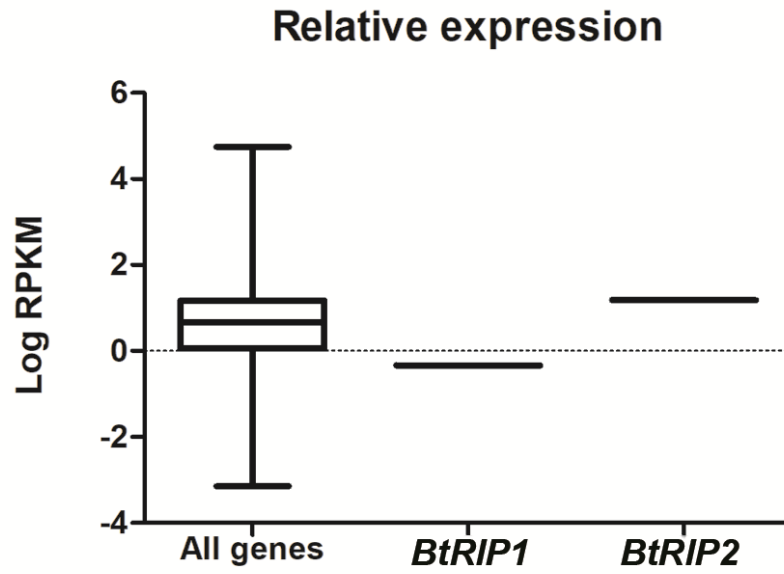


Figure S3. Expression level of *BtRIPs*. Data provided by Dr. Dr. Zhangjun Fei were plotted as log of RPKM. Values of *BtRIP1* and *BtRIP2* were compared with the whole set of *B. tabaci* genes according to Chen *et al.* [3]. The box-and-whisker plot shows expression levels as quartiles.

