

Supplementary Table S6. Identification of the putative regulatory elements and position at the in 5′- and 3′- regulatory regions of the DNA sequences encoding for tubulin-like proteins found in the *T. vaginalis* genome database.

Group	Name	TVAG_	Inr	M1	M2	M3	M4	M5	PS	CS	DSE
I	A	TvTUBa1	359090	TCACT (-9 to -13)		AAAGTGTC (-12 to -19)		TAAAAATC (-36 to -44)	TAAA (included in the stop codon) TAAA (14 to 17)	TAATT (7 to 11) TAATT (37 to 41)	TTTTTTT (49 to 55)
		TvTUBa2	196270	TCACT (-9 to -13)		AAAGTGTC (-12 to -19)		TAAAAATC (-36 to -44)	TAAA (included in the stop codon) TAAA (14 to 17)	TAATT (7 to 11) TAATT (37 to 41)	TTTTTTT (49 to 55) TTTT (60 to 63)
		TvTUBa3	206890	ACACT (-9 to -13)					TAAA (included in the stop codon) TAAA (14 to 17)	TAATT (7 to 11) TAATT (37 to 41)	TTTTT (49 to 54)
		TvTUBa4	312330	TCACT (-7 to -11)		GACGGTTT (-16 to -23)			TAAA (included in the stop codon)	TAATT (7 to 11)	TTTT (42 to 45)
		TvTUBa5	360870	TCACT (-9 to -11)			TAAAAATA (-42 to -49) AAAAAATT (-32 to -39)		TAAA (included in the stop codon) TAAA (14 to 17)	TAATT (7 to 11)	TTTTT (24 to 28) TTTTTTTT (51 to 58)
		TvTUBa6	448380	TCAAATCAC (-10 to -18)					TAAA (7 to 10)	AAATT (37 to 41)	TTTT (47 to 51)
		TvTUBa7	467840	TCAAATCA (-11 to -18)					TAAA (19 to 22)	TAATT (25 to 29)	TTTT (31 to 34)
	B	TvTUBβ1	034440	TCAC TTCA (-7 to -14)					TAAA (included in the stop codon) TAAA (46 to 49)	CAATT (2 to 6) AAATT (60 to 64) TAATT (70 to 74)	TTTTT (28 to 32) TTTTTTTT (37 to 43) TTTTT (66 to 70)
		TvTUBβ2	456920	TCATT (-7 to -11)					TAAA (7 to 10)	CAATT (37 to 41)	TTTTTT (55 to 60) TTTTTT (94 to 100)
		TvTUBβ3	008680	TCACT (-8 to -12)					TAAA (21 to 24)	AAATT (56 to 60)	TTTTTT (67 to 73)
		TvTUBβ4	062880	TCATA (-8 to -12)			AAAAAATA (-63 to -70)		TAAA (22 to 25)	AATT (27 to 30) TAATT (45 to 49)	TTTTTT (32 to 37) TTTTT (60 to 64)
A	A		109820	TCATT (-15 to -19)	TCATTTTT (-12 to -19) TCACTTTT (-26 to -33)		AAAAAATA (-54 to -61)		TAAA (included in the stop codon) TAAA (11 to 14) TAAA (15 to 18)	TAATT (78 to 82)	
			338530	TCATA (-8 to -12)	CAAAATTT (-38 to 45)				TAAA (24 to 27)	TAATT (32 to 36) AAATT (46 to 50)	TTTTT (59 to 63) TTTTT (69 to 73)
			207590	TCAAA (-7 to -11)			TAAAAAATT (-28 to -35) GAAAAAATT (-17 to -24)		TAAA (1 to 4)	TAATT (11 to 15) TAATT (23 to 27) AAATTT (33 to 38)	

II	B	525430	Without regulatory regions		
		523980	Without regulatory regions and ORF		
		519620	Without regulatory regions		
		073810	TCACA (-8 to -12)		
		289290	ACAGT (-20 to -24)		TAAA (24 to 27) TAATT (32 to 36) AAATT (46 to 50) TTTT (59 to 63) TTTT (67 to 70)
		200200			TAAA (18 to 21) TAAA (27 to 30) TATT (53 to 57) AATT (72 to 76) AATT (78 to 81)
		024080	Without regulatory regions and ORF		
		345420	AAAAAACG (-31 to -38)		TAAA (included in the stop codon) AAATT (9 to 12)
		148390	TCACA (-8 to -12)		
		148400	TCACA (-25 to -29)		TAAA (included in the stop codon) TAATT (44 to 48) TTTT (83 to 86)
		043330	TCACT (-17 to -21)		TAAA (79 to 82) TTTTTT (94 to 100)
		448410	TCACT (-17 to -21)		TAAA (51 to 54) TAAA (61 to 64) TTTT (66 to 69) TTTTTT (71 to 77)
	C	065740	TCACT (-2 to -6)	TCATTTTT (-14 to -21)	TAAA (38 to 41)
		184510	TAATT (Inr-like) (-15 to -19)		TAAA (included in the stop codon) AAATT (44 to 47) TTTTTT (36 to 43)
		369500	TCATT (-15 to -19)	TCACTTTT (-26 to -33) TCATTTTT (-12 to -19)	AAAAAATA (-54 to -61) TAAA (29 to 32) AAATT (79 to 83)
		073800	TCACA (-25 to -29)		TAAA (51 to 54) AAATT (60 to 64)
		257730	TCACT (-17 to -21)		

The 5'- and 3'-regulatory regions of all DNA sequence encoding for tubulin-like proteins were downloaded from each contig including 100 bp up and downstream of each ORF found in the *Tv* genome database, analyzed to find putative promoter elements or motifs Inr, Inr-like, motif 1 (M1), motif 2 (M2), motif 3 (M3), motif 4 (M4), motif 5 (M5), or/and polyadenylation signal (PS), cleavage site (CS) and downstream element (DSE), respectively. We did a manual search of these putative elements to find them in all sequences. The search was based on reported sequences by different groups (49-53; *extra ref a-l). Several sequences contain the necessary elements to be expressed, since they possess a start codon (ATG), stop codon (TAA/TAG), putative Inr/Inr-like/motifs in the 5'-upstream regulatory region and putative PS, CS, and DSE elements in the 3'-downstream regulatory region. Others only possess regulatory elements in the 5'- upstream regulatory regions or only regulatory elements in the 3'-downstream regulatory regions. We did not find regulatory sequences in four DNA sequences and two of them did not have an open reading frame (ORF), and lack of the ATG star codon.

***Extra references with 5'- and 3'-regulatory sequences in different reported genes.**

- a) Figueroa-Angulo E. E., Estrella-Hernández P., Salgado-Lugo H., Ochoa-Leyva A., Gómez Puyou A., Campos S. S., Montero-Moran G., Ortega-López J., Saab-Rincón G. and Arroyo R. 2012. Cellular and biochemical characterization of two closely related triosephosphate isomerases from *Trichomonas vaginalis*. *Parasitology*. 139(13):1729 – 1738.
- b) Hernández-Romano P., Hernández R., Arroyo R., Alderete J. and López-Villaseñor I. 2010. Identification and characterization of a surface-associated, subtilisin-like serine protease in *Trichomonas vaginalis*. *Parasitology*, 137(11):1621-1635.
- c) León-Félix J., Ortega-López J., Orozco-Solís R. and Arroyo R. 2004. Two novel asparaginyl endopeptidase-like cysteine proteinases from the protist *Trichomonas vaginalis*: their evolutionary relationship within the clan CD cysteine proteinases. *Gene*. 335:25-35.
- d) León-Sicaire C. R., León-Félix J. and Arroyo R. 2004. tvcp12: a novel *Trichomonas vaginalis* cathepsin L-like cysteine proteinase-encoding gene. *Microbiology*. 150(5):1131-1138.
- e) Liston D. R. and Johnson P. J. 1999. Analysis of a ubiquitous promoter element in a primitive Eukaryote: early evolution of the initiator element. *Molecular and Cellular Biology*. 19, 2380-2388.
- f) Mancilla-Olea M. I., Ortega-López J., Figueroa-Angulo E. E., Ávila-González L., Cárdenas-Guerra R. E., Miranda-Ozuna J. F. T., González-Robles A., Hernández-García M. S., Sánchez-Ayala L. and Arroyo R. 2018. *Trichomonas vaginalis* cathepsin D-like aspartic proteinase (Tv-CatD) is positively regulated by glucose and degrades human hemoglobin. *The International Journal of Biochemistry & Cell Biology*. Volume 97:1-15.
- g) Puente-Rivera J., Ramón-Luing L. A., Figueroa-Angulo E. E., Ortega-López J. and Arroyo R. 2014. Trichocystatin-2 (TC-2): An endogenous inhibitor of cysteine proteinases in *Trichomonas vaginalis* is associated with TvCP39. *The International Journal of Biochemistry & Cell Biology*. 54:255-265.
- h) Puente-Rivera J., Villalpando J. L., Villalobos-Osnaya A., Vázquez-Carrillo L. I., León-Ávila G., Ponce-Regalado M. D., López-Camarillo C., Elizalde-Contreras J. M., Ruiz-May E., Arroyo R., Alvarez-Sánchez M. E. 2017. The 50kDa metalloproteinase TvMP50 is a zinc-mediated *Trichomonas vaginalis* virulence factor. *Molecular and Biochemical Parasitology*. 217:32-41.
- i) Ramón-Luing L. A., Rendón-Gandarilla F. J., Puente-Rivera J., Ávila-González L. and Arroyo R. 2011. Identification and characterization of the immunogenic cytotoxic TvCP39 proteinase gene of *Trichomonas vaginalis*. *The International Journal of Biochemistry & Cell Biology*. 43(10):1500-1511..j) Smith A. J., Chudnovsky L., Simoes-Barbosa A., Delgadillo-Correa M. G., Jonsson Z. O., Wohlschlegel J. A., and Johnson P. J. 2011. Novel core promoter elements and a cognate transcription factor in the divergent unicellular eukaryote *Trichomonas vaginalis*. *Molecular and cellular biology*, 31(7), 1444–1458.
- k) Solano-González E., Burrola-Barraza E., León-Sicaire C., Ávila-González L., Gutiérrez-Escolano L., Ortega-López J. and Arroyo, R. 2007. The trichomonad cysteine proteinase TVCP4 transcript contains an iron-responsive element. *FEBS Letters*. 581.
- l) Amos B., Aurrecochea C., Barba M, Barreto A, Basenko E. Y., Bazant W., Belnap R., Blevins A. S., Böhme U., Brestelli J., Brunk B. P., Caddick M., Callan D., Campbell L., Christensen M. B., Christophides G. K., Crouch K., Davis K., DeBarry J., Doherty R., Duan Y., Dunn M., Falke D., Fisher S., Flicek P., Fox B., Gajria B., Giraldo-Calderón G. I., Harb O. S., Harper E., Hertz-Fowler C., Hickman M. J., Howington C., Hu S., Humphrey S., Iodice J., Jones A., Judkins J., Kelly S. A., Kissinger J. C., Kun Kwon D., Lamoureux K, Lawson D., Li W., Lies K., Lodha D., Long J., MacCallum R. M., Maslen G., McDowell M. A., Nabrzyski J., Roos D. S., Rund S. S. C., Wever Schulman S., Shanmugasundram A., Sitnik V., Spruill D., Starns D., Stoeckert C. J., Tomko S. S., Wang H, Warrenfeltz S., Wieck R., Wilkinson P. A., Xu L. and Zheng J. 2022. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research*. 50(D1):D898–D911.