

Table S1: Impact of the data curation steps on the count of TGAs.

Species	CDS with a relic tag									CDS with minisatellites	
	not included in a TGA ^a					included in a TGA ^b				Number of CDSs before manual curation	Number of TGAs eliminated ^g
	Number of CDSs before curation	Automated curation ^c	Manual curation ^d	Number of false positive CDSs	Number of CDSs after curation ^e	Number of CDSs before curation	Automated curation ^c	Manual curation ^d	Number of CDSs after curation ^f		
<i>S. cerevisiae</i>	45	10	18	17	0	4	0	1	3	21	0
<i>C. glabrata</i>	22	8	1	12	1	3	0	3	0	39	3
<i>Z. rouxii</i>	32	2	5	17	8	2	0	2	0	15	3
<i>K. thermotolerans</i>	29	2	7	14	6	2	0	2	0	14	3
<i>S. kluyveri</i>	47	5	8	30	4	1	1	0	0	23	8
<i>K. lactis</i>	26	4	7	13	2	0	0	0	0	11	1
<i>A. gossypii</i>	29	1	3	24	1	2	1	0	1	14	2
<i>D. hansenii</i>	74	17	25	16	16	9	3	3	3	37	1
<i>Y. lipolytica</i>	56	4	11	34	7	0	0	0	0	25	12
All nine species	360	53	85	177	45	23	5	11	7	199	33

^aCDS at position n with at least one FTB score ≥ 10 but both CDSs at positions n+1 and n-1 have not the score profile defined for a TGA.

^bCDS at position n with two FTB scores ≥ 10 but just one CDS at position n+1 or n-1 has the “TGA profile”.

^cNumber of tagged CDSs automatically retrieved and included in a new or pre-existent TGA since at least one CDS at position n+2 or n-2 has the “TGA profile”.

^dNumber of tagged CDSs manually retrieved and included in a new or pre-existent TGA.

^eThis count finally gives the number of CDSs that belong to the 1CDS-relic class of TGAs.

^fNumber of CDSs that belong to a TGA containing at least two CDSs and are adjacent to a degenerated duplicated copy (nCDSs-relic class).

^gTGAs manually removed because the high value of their corresponding FTB scores was only due to the presence of minisatellites in CDS sequences.