Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity.

Francisco Callejas-Hernández[1], Alberto Rastrojo[1], Cristina Poveda[1], Núria Gironès*[1,2] and Manuel Fresno*[1,2]

[1] Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas, Universidad Autónoma de Madrid, Cantoblanco, Madrid, Spain.

[2] Instituto Sanitario de Investigación Princesa, Madrid, Spain.

* Address correspondence to: mfresno@cbm.csic.es or ngirones@cbm.csic.es Tel: +34 91 1964565; Fax: +34 91 1964420;

Supporting Information: Figures
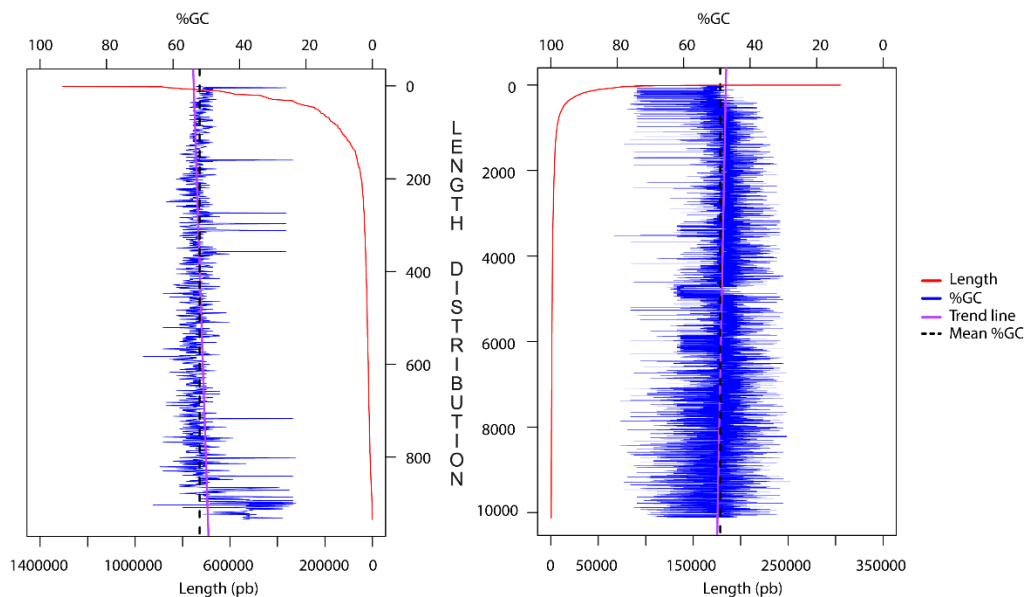
S1 Fig. %GC distribution on the contigs of the new assembled genomes. Bug2148 contigs (left panel). Y contigs (right panel). Contig length distribution (red trendline) %GC distribution (blue line) and %GC trend line (pink line).
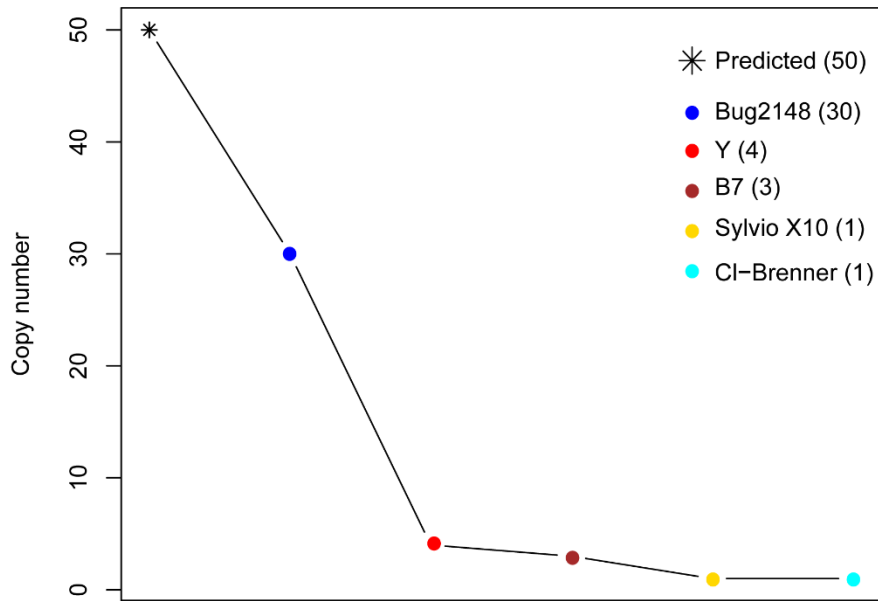
S2 Fig. MisSassembly control (Monoglyceride lipase). Copy number of Monoglyceride lipase in our new assembled and previously available genomes.
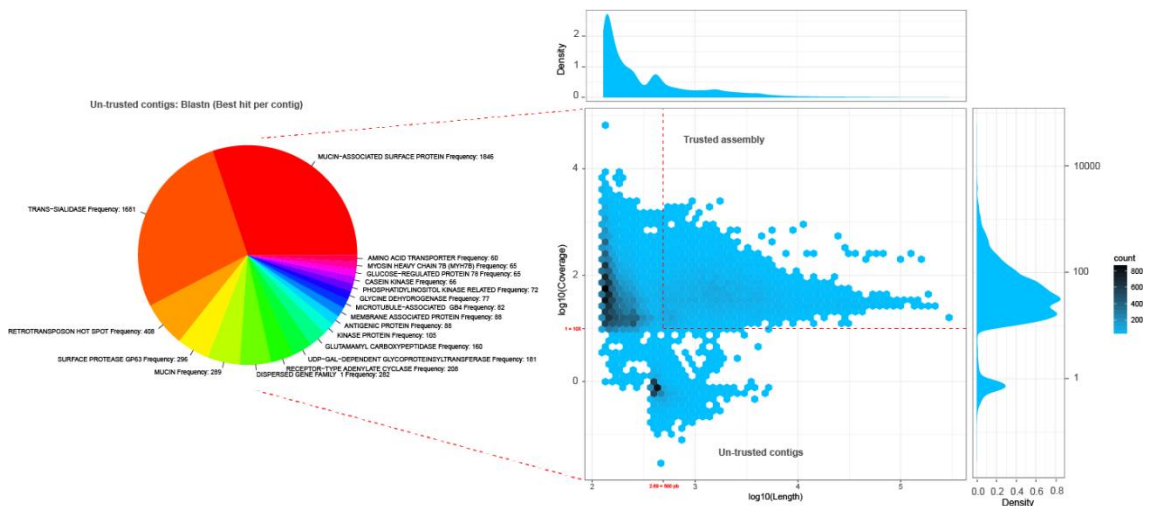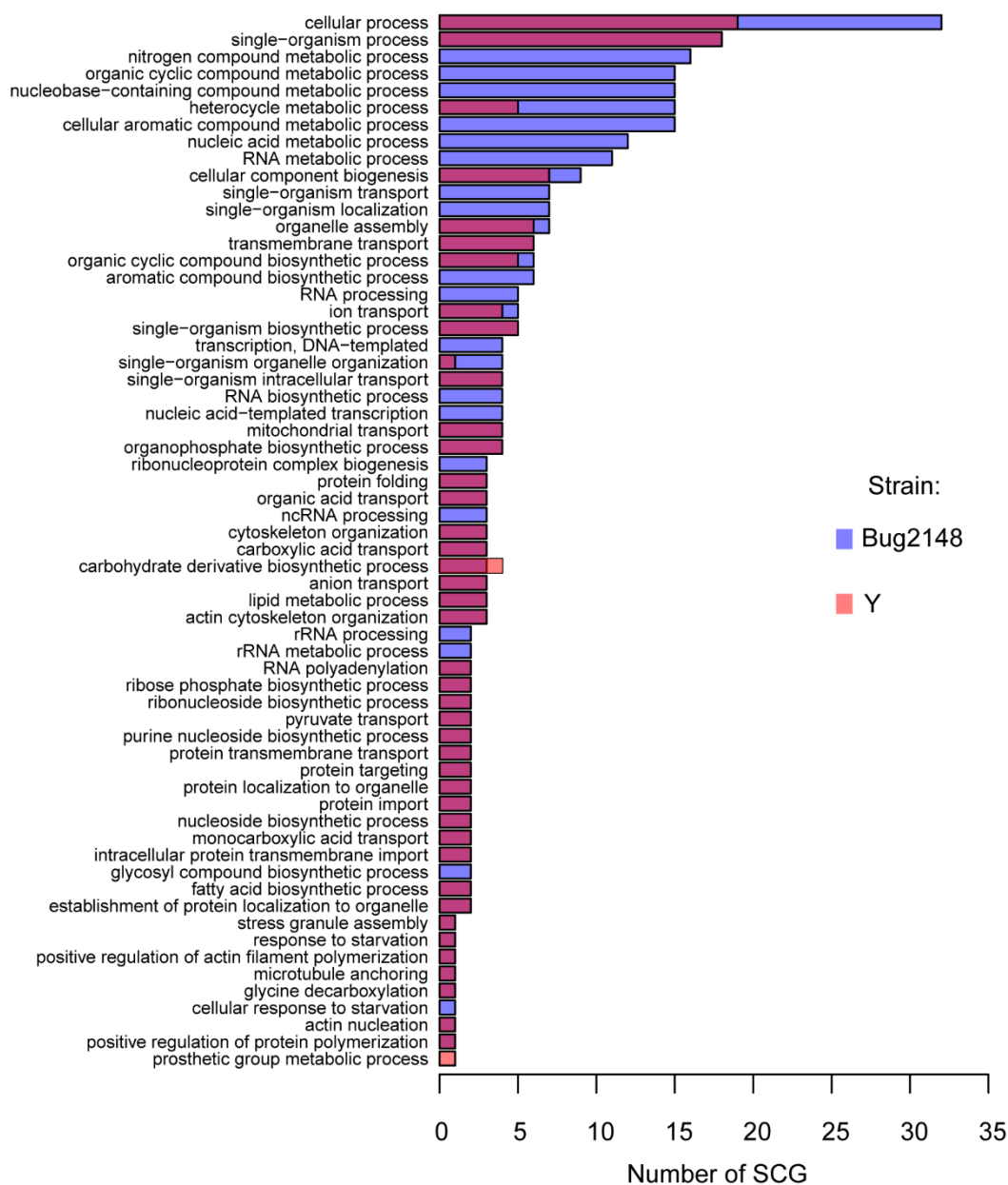

S3 Fig. Profile of contigs assembled by Illumina reads and spades assembler (Y strain), length versus coverage (log10 based). Untrusted contigs were defined as contigs with coverage below of 10X and length shorter than 500 nucleotides and therefore, trusted contigs constitute the assembled genome (Y strain). In this case, coverage above 10,000X is indicative of high levels of genomic plasticity and aneuploidies, in agreement to previous speculations but also it allow us to identify the specific genetic families causing these collapses.

T. cruzi DGF proteins; phylogram

Brener Esmeraldo-like

Brener non-Esmeraldo-like

New: Y

New: Bug2148

B7

Dm28c

SylvioX10
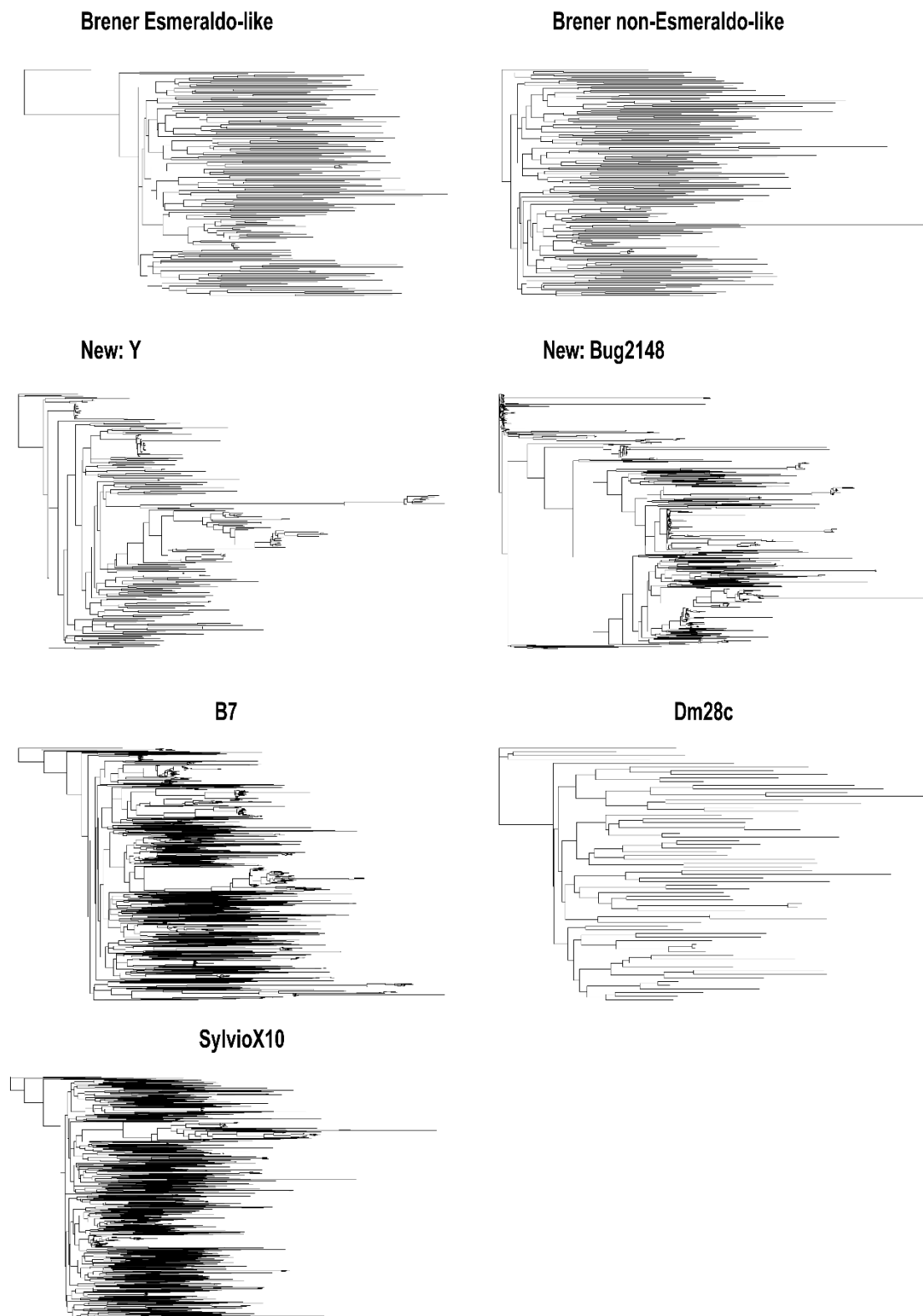
S4 Fig. Hypothetical and annotated DGF proteins in T. cruzi genomes.

S5 Fig. SCG associated to specific biological processes. SCG association to Biological process was obtained by Gene ontology (GO) analysis in Tritryp DB (http://tritrypdb.org/tritrypdb/), redundancy was manually filtered when possible.

Supporting Information: Files

S1 File. G+C contig mean per assembled contig.

S2 File. Annotated functions in *Trypanosoma cruzi* available draft genomes.

S3 File. Contigs containing the lowest and highest linguistic complexity.