Supplementary table

All data used in the analysis are summarized here. Megasatellite\_ID: corresponds to the gene identification. When more than one megasatellite is found in the same gene, it is followed by an index number, 1, 2, etc. Gene Name: name of the gene, when it has a known function. Start, End: refer to motif borders on the corresponding protein. Size: motif size (in amino acids). Clusters: cluster of the motif, or of the protein in which the tandemly repeated peptide was deleted. single: motif or protein without any detectable homologue. Family: Megasatellite family, based on clustering (see text). Motif\_ID: each motif in a given megasatellite was numbered, so that each motif bears a unique identifier. Amino-acid %: Amino-acid composition of megasatellites. Amino acids in proportions statistically different from the average composition of the studied proteome are indicated (see Figure 3 and text). The composition is indicated next to the first motif of the tandem repeat. Motif sequence: individual sequence of each motif.