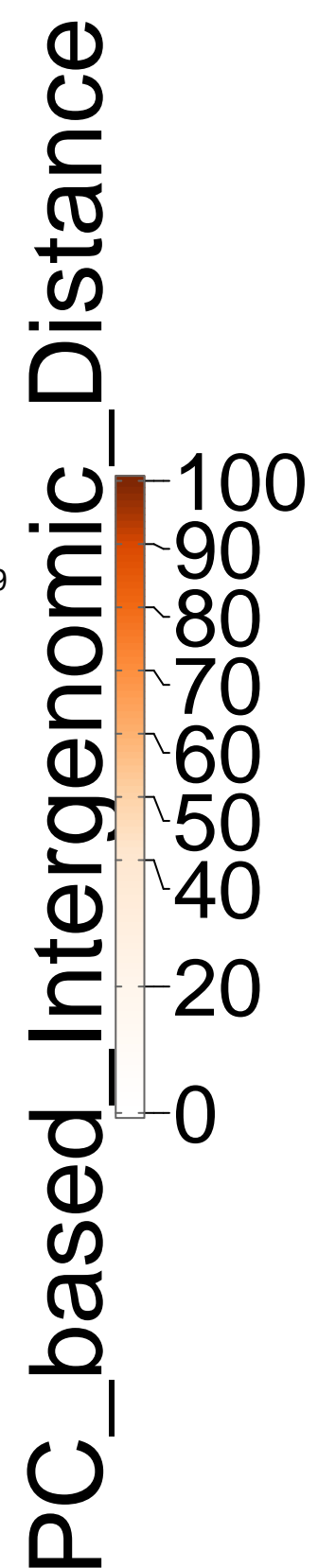


Fig. S1: Intergenomic distances for Crz_DB dataset, calculated with the following protein clustering parameters: coverage = 100%, % identity = 100%.



Family Genus

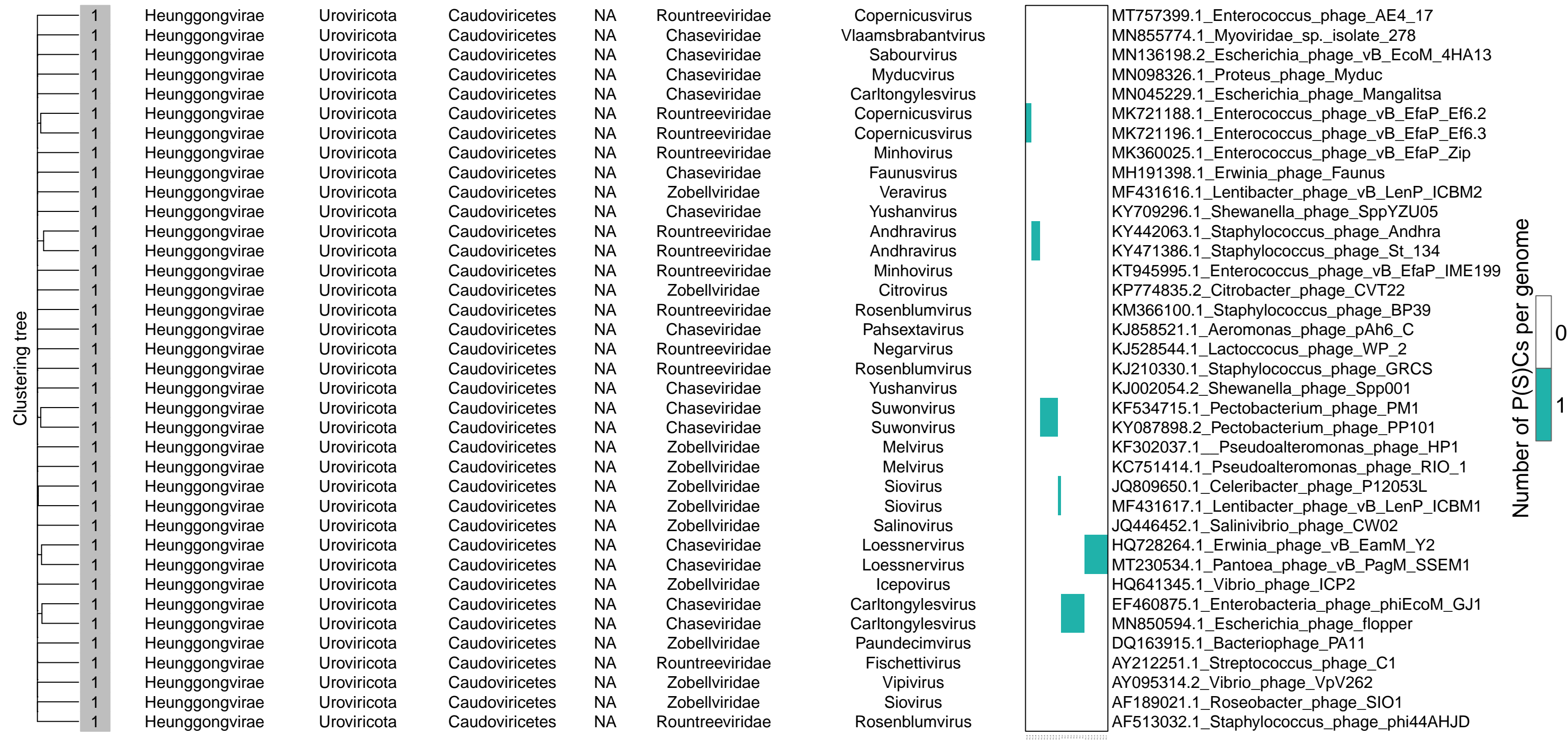


Fig. S4: Integrated visualization of the genome clustering of the Crz-DB dataset, calculated with the following protein clustering parameters: coverage = 100% and percent identity = 100%. Columns: 1) genome clustering tree; 2) Silhouette width; 3) VGC number; 4) Family; 5) Genus; 6) PC distribution; 7) genome names.

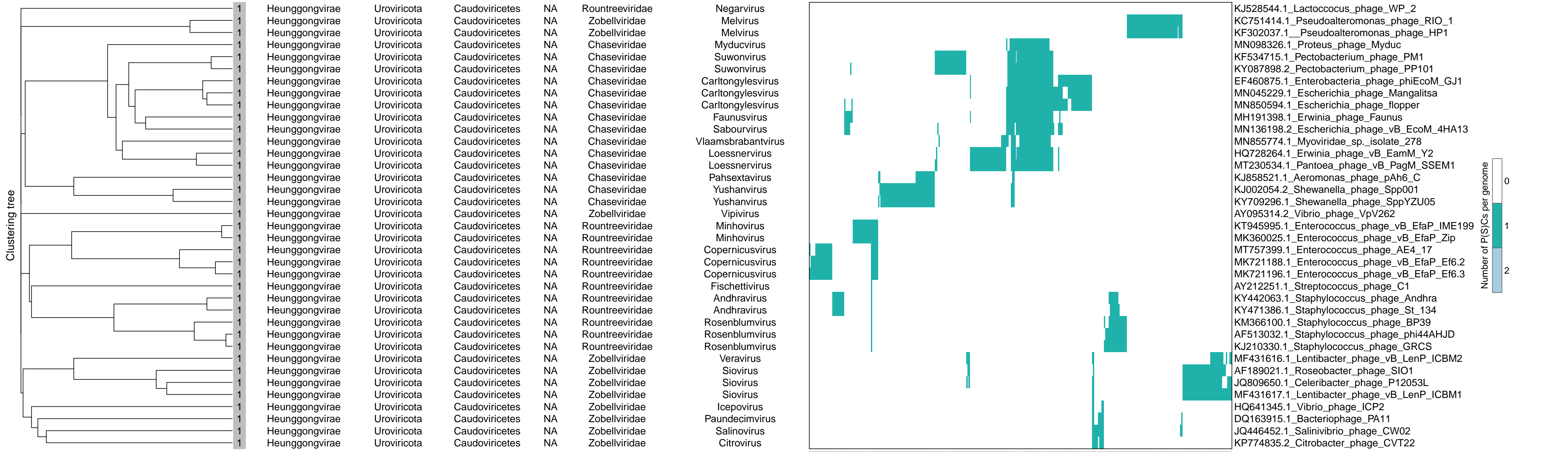


Fig. S5: Integrated visualization of the genome clustering of the Crz-DB dataset, calculated with the following protein clustering parameters: coverage = 80% and percent identity = 50%. Columns: 1) genome clustering tree; 2) Silhouette width; 3) VGC number; 4) Family; 5) Genus; 6) PC distribution; 7) genome names.

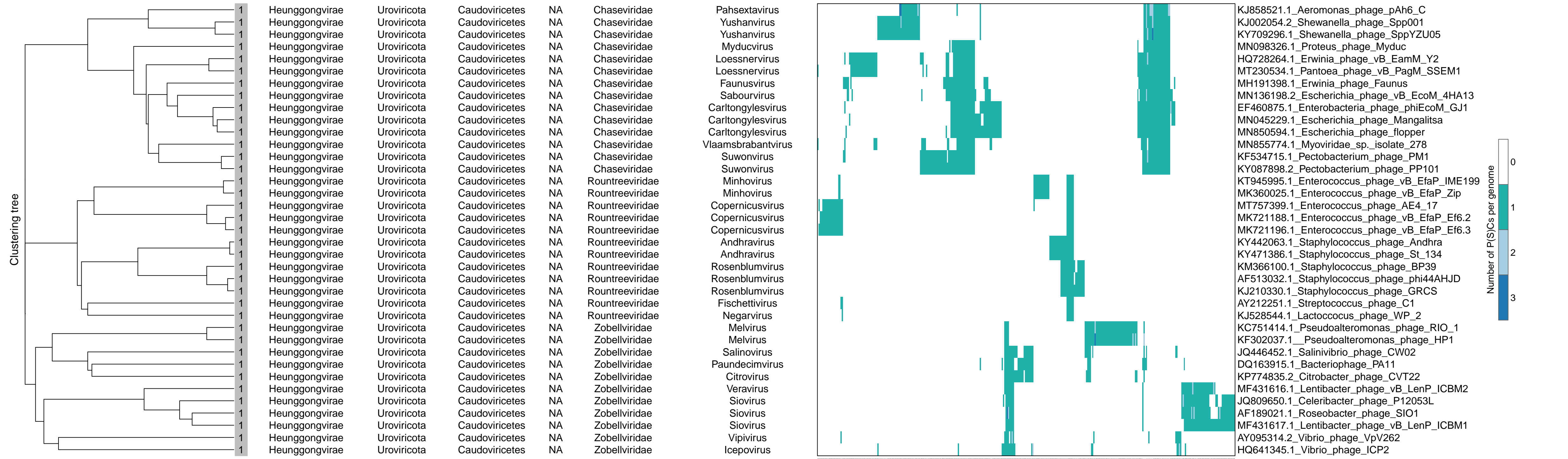


Fig. S6: Integrated visualization of the genome clustering of the Crz-DB dataset, calculated with the following protein clustering parameters: coverage = 0% and percent identity = 0%. Columns: 1) genome clustering tree; 2) Silhouette width; 3) VGC number; 4) Family; 5) Genus; 6) PC distribution; 7) genome names.