Algorithm 1 Ensemble Voting of User Genome from GTDB to NCBI

Input: Initialize parsing genome GTDB-NCBI-r203 mapping metadata with training genome capacity $\mathcal{N}$

Input: Initialize reading user genome Q and tree T classified with GTDB-Tk

Fixed ensemble weights θ, θ' to constant 0.5

for $genome = 1, N$ **do**

  Associating genome accession ID gid with NCBI taxonomy in Dict $\mathcal{D}_{ncbi\_taxa}$

  Associating representative genome ID rid with GTDB taxonomy in Dict $\mathcal{D}_{gtdb\_taxa}$

  Clustering genome gid into representative genome rid in Dict $\mathcal{D}_{cluster}$

  Associating specific $taxon$ with NCBI taxonomy lineage in Dict $\mathcal{D}_{ncbi\_lineage}$

  Storing $\mathcal{D}_{ncbi\_taxa}$, $\mathcal{D}_{gtdb\_taxa}$, $\mathcal{D}_{cluster}$ and $\mathcal{D}_{ncbi\_lineage}$

end for

Initializing GTDB cluster with NCBI-type taxonomy in Dict $\mathcal{N}4\mathcal{G}_{cluster}$

for $rid, gids$ in $\mathcal{D}_{ncbi\_taxa}$.items() **do**

  for $rank = r_{species}, r_{kindom}$ **do**

    Initializing hit list for $\mathcal{D}_{ncbi\_taxa}$ with $\mathcal{L}_{hit}$

    for ε in $gids$ **do**

      if ε in $\mathcal{D}_{ncbi\_taxa}$ **then**

        appending $\mathcal{D}_{ncbi\_taxa}[\varepsilon][rank]$ into $\mathcal{L}_{hit}$

      end if

    end for

    if $\mathcal{L}_{hit}$ is NOT empty **then**

      Counting element occurence in $\mathcal{L}_{hit}$

      Applying heapsort to $\mathcal{L}_{hit}$ and finding top 1 element

      tax_{top}, $count_{top} = \mathcal{HEAPSORT}(\mathcal{L}_{hit})$

      if $count_{top} >= \theta$ * $\mathcal{L}_{hit}$.size() && tax_{top} is NOT unassigned **then**

        $\mathcal{N}4\mathcal{G}_{cluster}[rid] = \mathcal{D}_{ncbi\_lineage}[tax_{top}]$

        break

      end if

    end if

    if $\mathcal{N}4\mathcal{G}_{cluster}[rid]$ is NOT unassigned **then**

      Reporting representative genome cannot be converted to NCBI taxonomy

    end if

  end for

  Storing $\mathcal{N}4\mathcal{G}_{cluster}[rid]$

end for