

---

**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

- 1: Fixed ensemble weights  $\theta, \theta'$  to constant 0.5
- 2:
- 3: **for**  $genome = 1, N$  **do**
- 4:   Associating genome accession ID  $gid$  with NCBI taxonomy in Dict  $\mathcal{D}_{ncbi\_taxa}$
- 5:   Associating representative genome ID  $rid$  with GTDB taxonomy in Dict  $\mathcal{D}_{gtdb\_taxa}$
- 6:   Clustering genome  $gid$  into representative genome  $rid$  in Dict  $\mathcal{D}_{cluster}$
- 7:   Associating specific  $taxon$  with NCBI taxonomy lineage in Dict  $\mathcal{D}_{ncbi\_lineage}$
- 8:   Storing  $\mathcal{D}_{ncbi\_taxa}$ ,  $\mathcal{D}_{gtdb\_taxa}$ ,  $\mathcal{D}_{cluster}$  and  $\mathcal{D}_{ncbi\_lineage}$
- 9: **end for**
- 10:
- 11: Initializing GTDB cluster with NCBI-type taxonomy in Dict  $\mathcal{N}4\mathcal{G}_{cluster}$
- 12: **for**  $rid, gids$  in  $\mathcal{D}_{ncbi\_taxa}.items()$  **do**
- 13:   **for**  $rank = r_{species}, r_{kindom}$  **do**
- 14:     Initializing hit list for  $\mathcal{D}_{ncbi\_taxa}$  with  $\mathcal{L}_{hit}$
- 15:     **for**  $\varepsilon$  in  $gids$  **do**
- 16:       **if**  $\varepsilon$  in  $\mathcal{D}_{ncbi\_taxa}$  **then**
- 17:         appending  $\mathcal{D}_{ncbi\_taxa}[\varepsilon][rank]$  into  $\mathcal{L}_{hit}$
- 18:       **end if**
- 19:     **end for**
- 20:     **if**  $\mathcal{L}_{hit}$  is *NOT* empty **then**
- 21:       Counting element occurrence in  $\mathcal{L}_{hit}$
- 22:       Applying heapsort to  $\mathcal{L}_{hit}$  and finding top 1 element
- 23:        $tax_{top}, count_{top} = \mathcal{HEAPSORT}(\mathcal{L}_{hit})$
- 24:       **if**  $count_{top} \geq \theta * \mathcal{L}_{hit}.size()$  &&  $tax_{top}$  is *NOT* unassigned **then**
- 25:          $\mathcal{N}4\mathcal{G}_{cluster}[rid] = \mathcal{D}_{ncbi\_lineage}[tax_{top}]$
- 26:       **break**
- 27:     **end if**
- 28:     **end if**
- 29:     **if**  $\mathcal{N}4\mathcal{G}_{cluster}[rid]$  is *NOT* unassigned **then**
- 30:       Reporting representative genome cannot be converted to NCBI taxonomy
- 31:     **end if**
- 32:   **end for**
- 33:   Storing  $\mathcal{N}4\mathcal{G}_{cluster}[rid]$
- 34: **end for**
- 35:

---