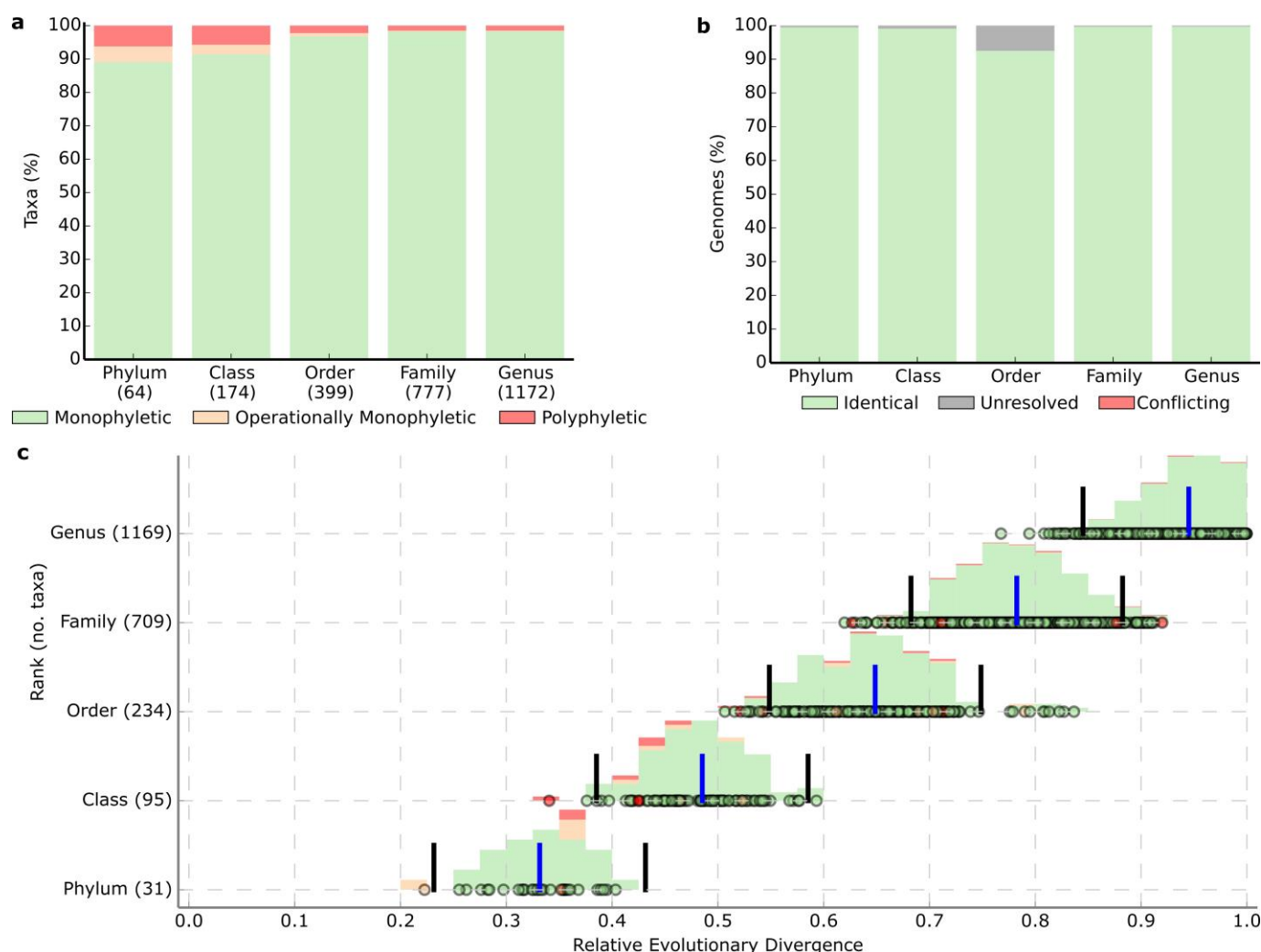
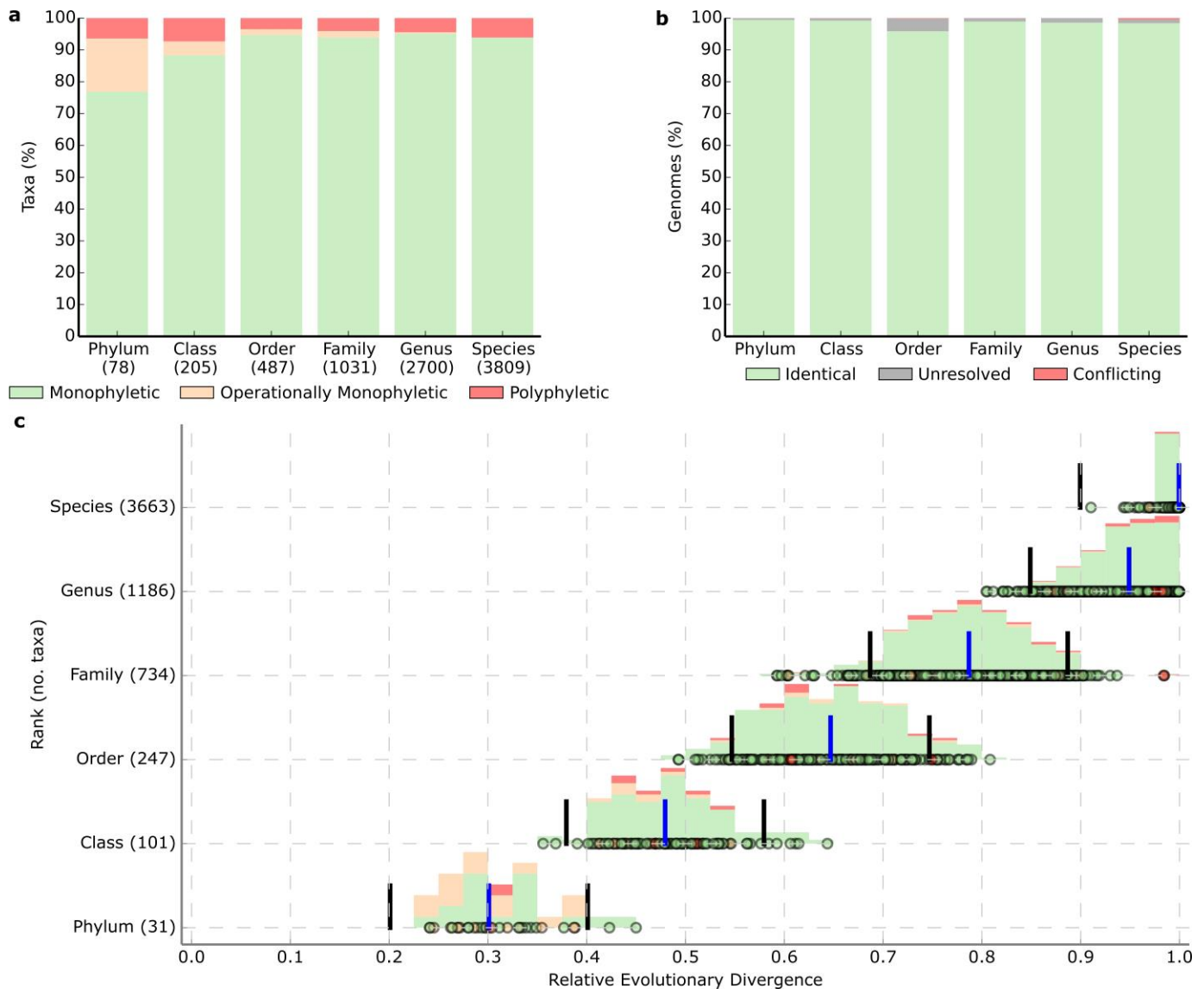




Supplementary Figure 1

Congruence of the GTDB taxonomy on a tree inferred with ExaML from the concatenation of 120 proteins (bac120) and species-dereplicated genome set.

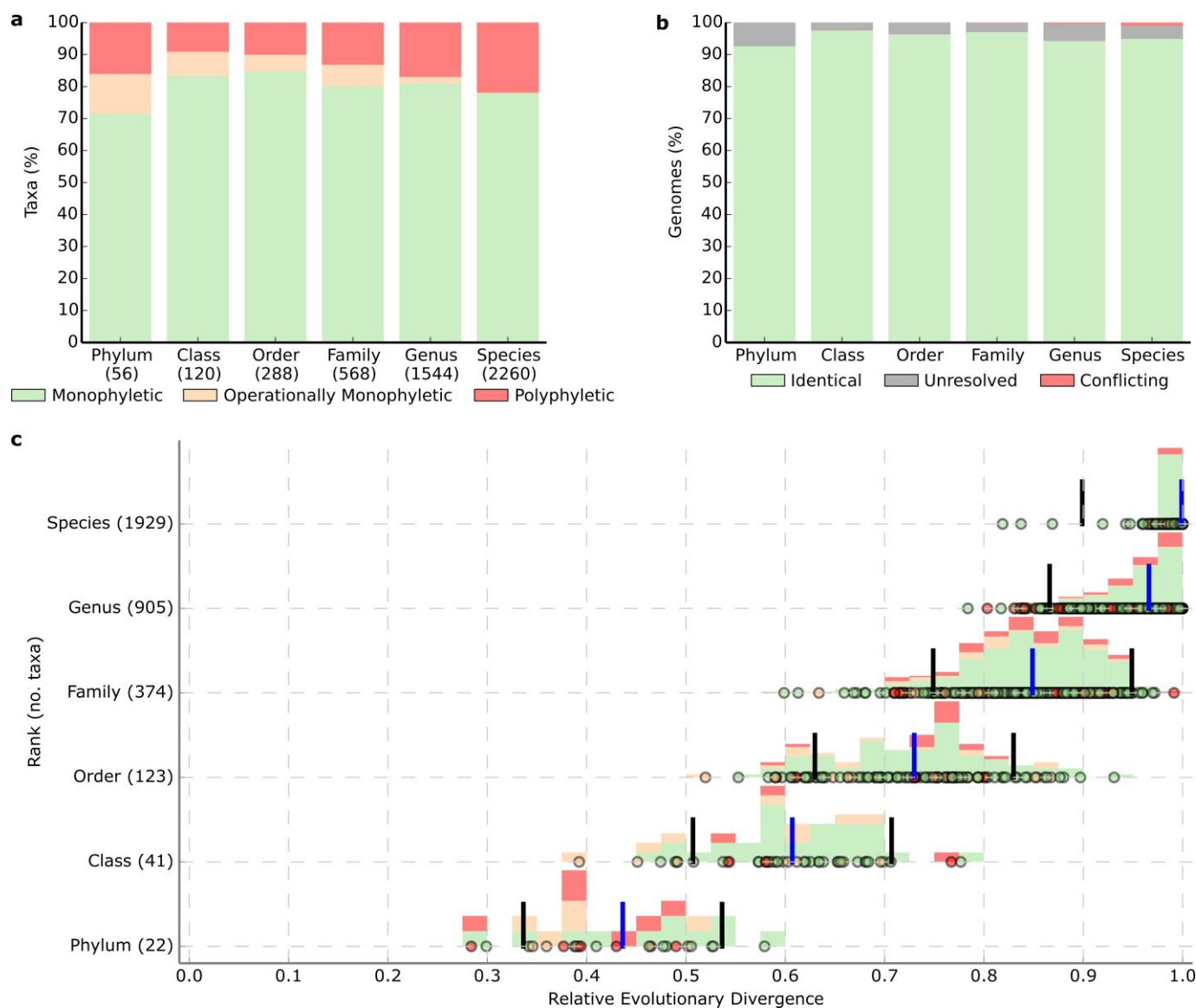
Percentage of GTDB taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the ExaML tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. **(b)** Percentage of the 10,462 genomes within the ExaML tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the GTDB taxonomy when taxonomy is assigned based on their placement in the inferred tree. **(c)** RED of taxa with ≥2 immediate subordinate taxa in the ExaML tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses.



Supplementary Figure 2

Congruence of the GTDB taxonomy on a tree inferred with FastTree from the concatenation of 16 ribosomal proteins (rp1).

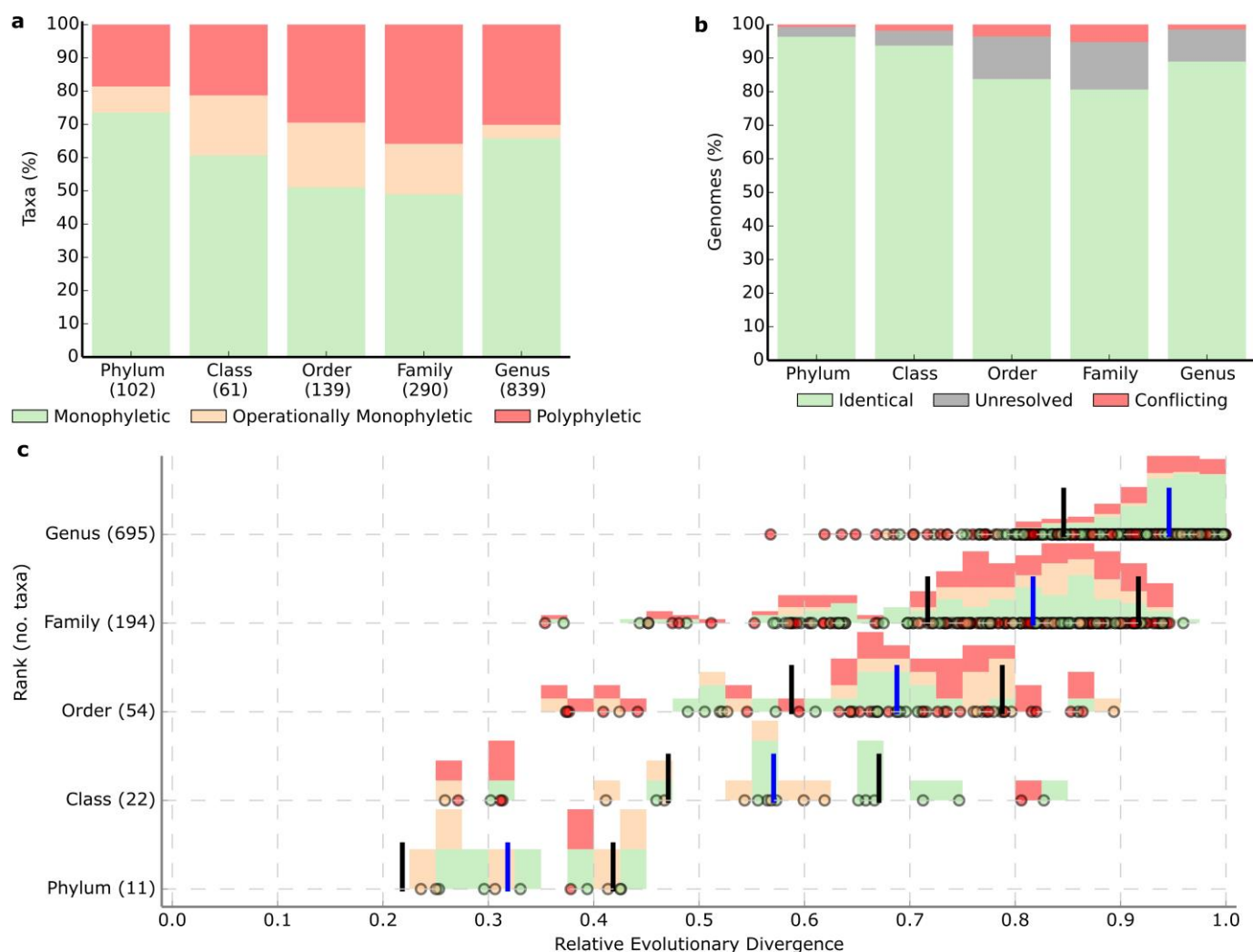
(a) Percentage of GTDB taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the rp1 tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. (b) Percentage of the 20,699 genomes within the rp1 tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the GTDB taxonomy when taxonomy is assigned based on their placement in the inferred tree. (c) RED of taxa with ≥2 immediate subordinate taxa in the rp1 tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses.



Supplementary Figure 3

Congruence of the GTDB taxonomy on a tree inferred with FastTree by using the 16S rRNA gene.

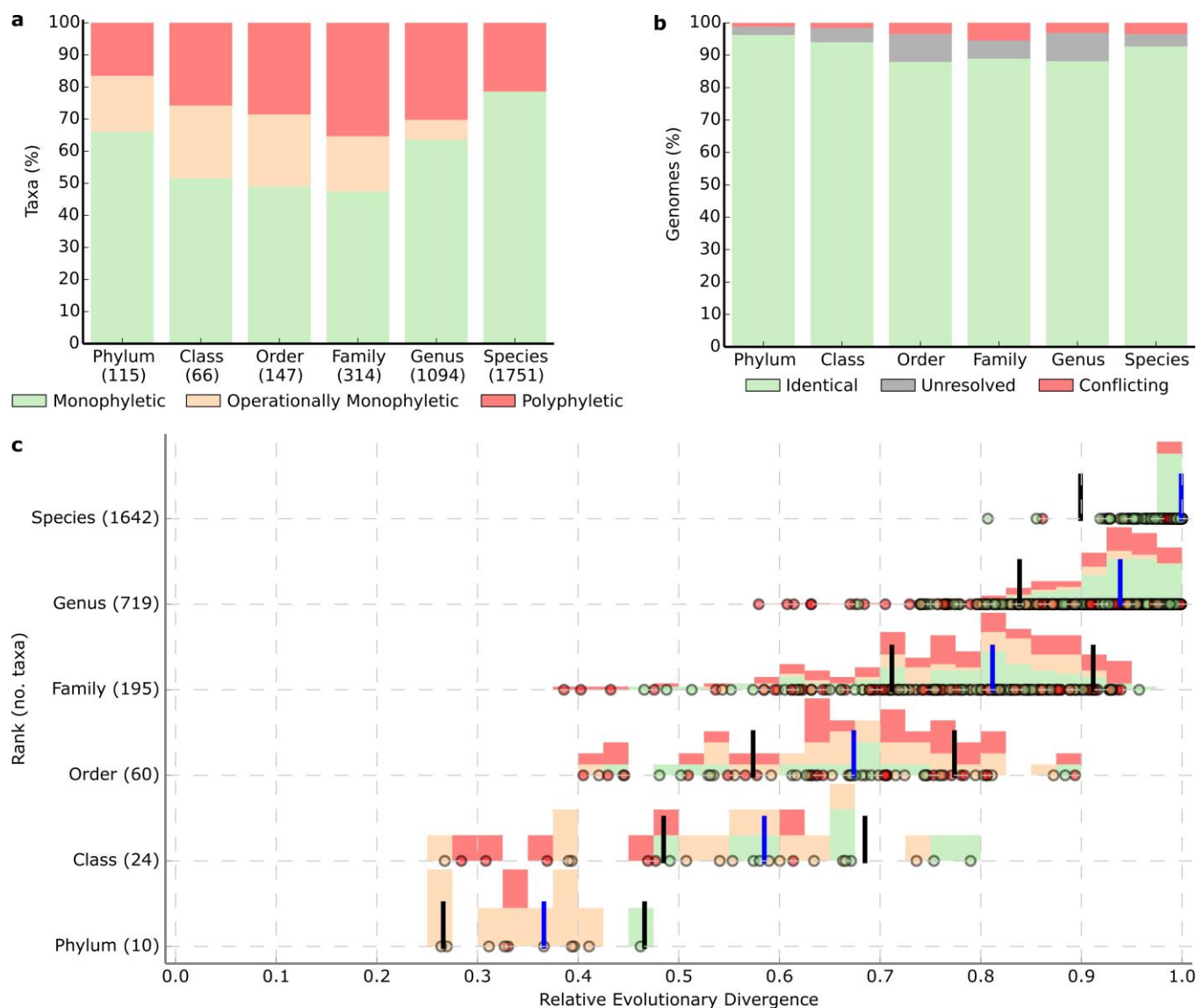
(a) Percentage of GTDB taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the 16S rRNA gene tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. (b) Percentage of the 12,435 genomes within the 16S rRNA gene tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the GTDB taxonomy when taxonomy is assigned based on their placement in the inferred gene tree. (c) RED of taxa with ≥2 immediate subordinate taxa in the 16S rRNA gene tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses.



Supplementary Figure 4

Congruence of the NCBI taxonomy on a tree inferred with ExaML from the concatenation of 120 proteins (bac120) and species-dereplicated genome set.

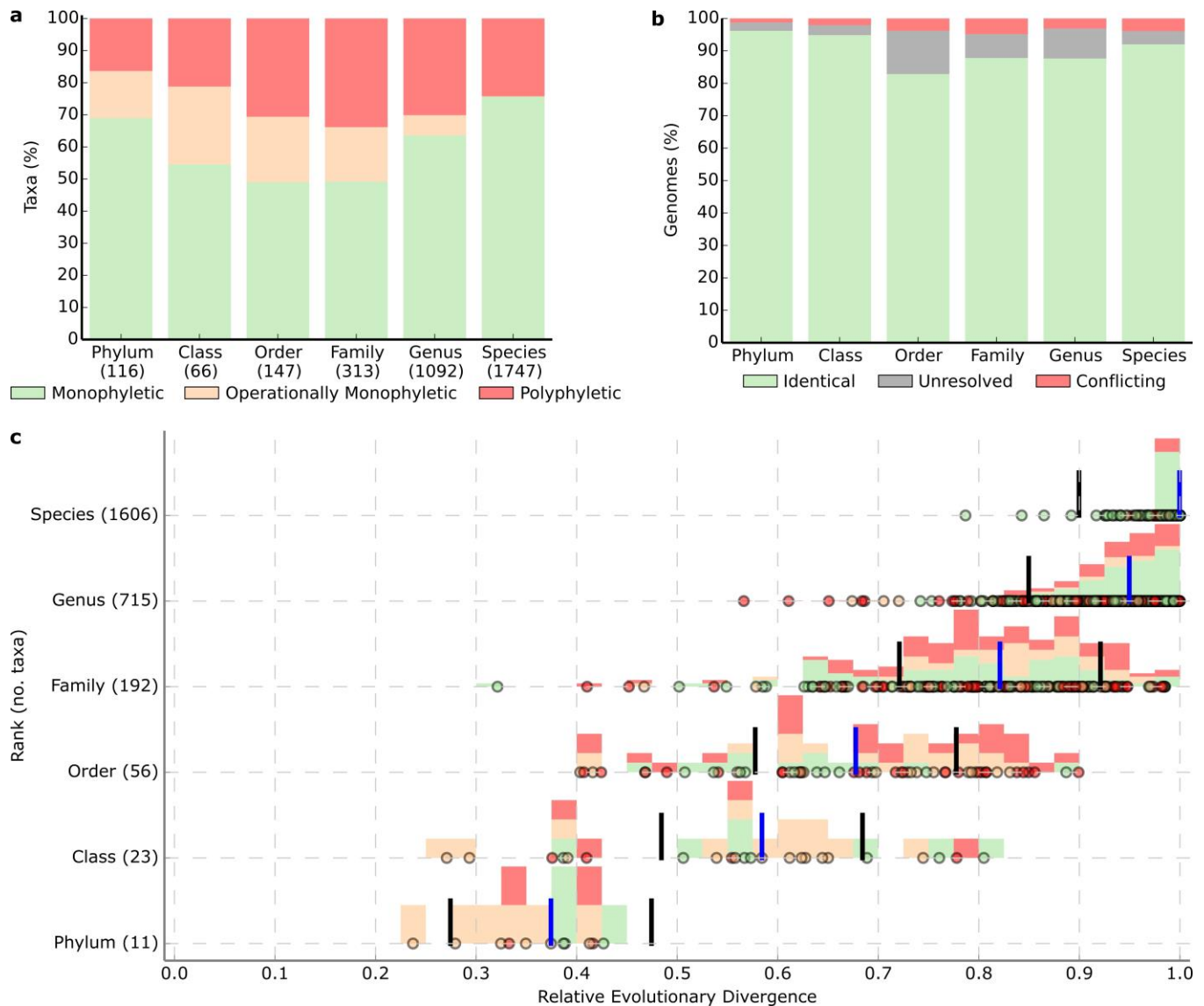
(a) Percentage of NCBI taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the ExaML tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. (b) Percentage of the 8,905 RefSeq/GenBank genomes within the ExaML tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the NCBI taxonomy when taxonomy is assigned based on their placement in the inferred tree. (c) RED of taxa with ≥ 2 immediate subordinate taxa in the ExaML tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses.



Supplementary Figure 5

Congruence of the NCBI taxonomy on a tree inferred with FastTree from the concatenation of 120 proteins (bac120).

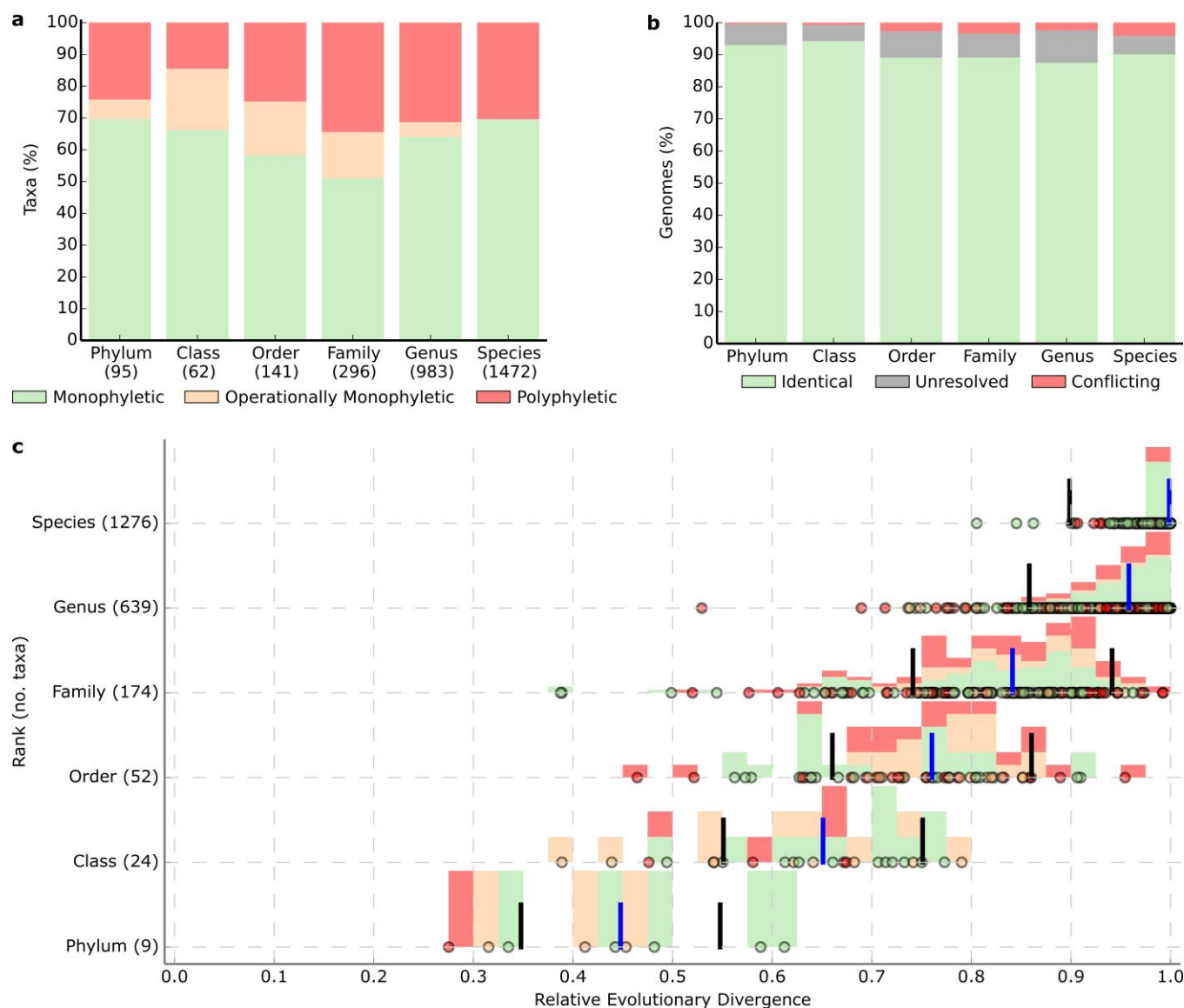
(a) Percentage of NCBI taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the bac120 tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. (b) Percentage of the 16,248 RefSeq/GenBank genomes within the bac120 tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the NCBI taxonomy when taxonomy is assigned based on their placement in the inferred tree. (c) RED of taxa with ≥2 immediate subordinate taxa in the bac120 tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses (note that this is a reproduction of Fig. 2a).



Supplementary Figure 6

Congruence of the NCBI taxonomy on a tree inferred with FastTree from the concatenation of 16 ribosomal proteins (rp1).

(a) Percentage of NCBI taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the rp1 tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. (b) Percentage of the 16,306 RefSeq/GenBank genomes within the rp1 tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the NCBI taxonomy when taxonomy is assigned based on their placement in the inferred tree. (c) RED of taxa with ≥ 2 immediate subordinate taxa in the rp1 tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses.



Supplementary Figure 7

Congruence of the NCBI taxonomy on a tree inferred with FastTree by using the 16S rRNA gene.

(a) Percentage of NCBI taxa at each rank which are monophyletic, operationally monophyletic, or polyphyletic within the 16S rRNA gene tree. Results were calculated over all taxa comprised of >1 genomes and the number of taxa considered at each rank is shown in parentheses. (b) Percentage of the 12,174 RefSeq/GenBank genomes within the 16S rRNA gene tree with identical, unresolved, or conflicting taxonomic assignments at each rank relative to the NCBI taxonomy when taxonomy is assigned based on their placement in the inferred gene tree. (c) RED of taxa with ≥ 2 immediate subordinate taxa in the 16S rRNA gene tree, with the same coloring as used in panel a. The number of taxa plotted at each rank is given in parentheses.



Supplementary Figure 8

Comparison of GTDB and SILVA.

Comparison of GTDB and SILVA taxonomic assignments across 10,779 bacterial genomes from RefSeq/GenBank release 80. These genomes are part of the 21,943 dereplicated genomes for which a 16S rRNA gene could be reliably matched by sequence similarity to a 16S rRNA gene in SILVA release 128. For each rank, a taxon was classified as being unchanged if its name was identical in both taxonomies, passively changed if the GTDB taxonomy provided name information absent in the SILVA taxonomy, or actively changed if the name was different between the two taxonomies after adjusting the SILVA taxonomy for colloquial designations indicating of missing taxonomic information (see *Methods*). Changes between the GTDB and SILVA taxonomies are given in **Supp. Table 9**.