

Classifying short genomic fragments from novel lineages using composition and homology

Donovan H. Parks^{1,§}, Norman J. MacDonald^{1,§}, and Robert G. Beiko^{1,*}

¹Faculty of Computer Science, Dalhousie University, 6050 University Avenue, Halifax, Nova Scotia, Canada B3H 1W5

§ These authors contributed equally to this work.

* To whom correspondence should be addressed (beiko@cs.dal.ca).

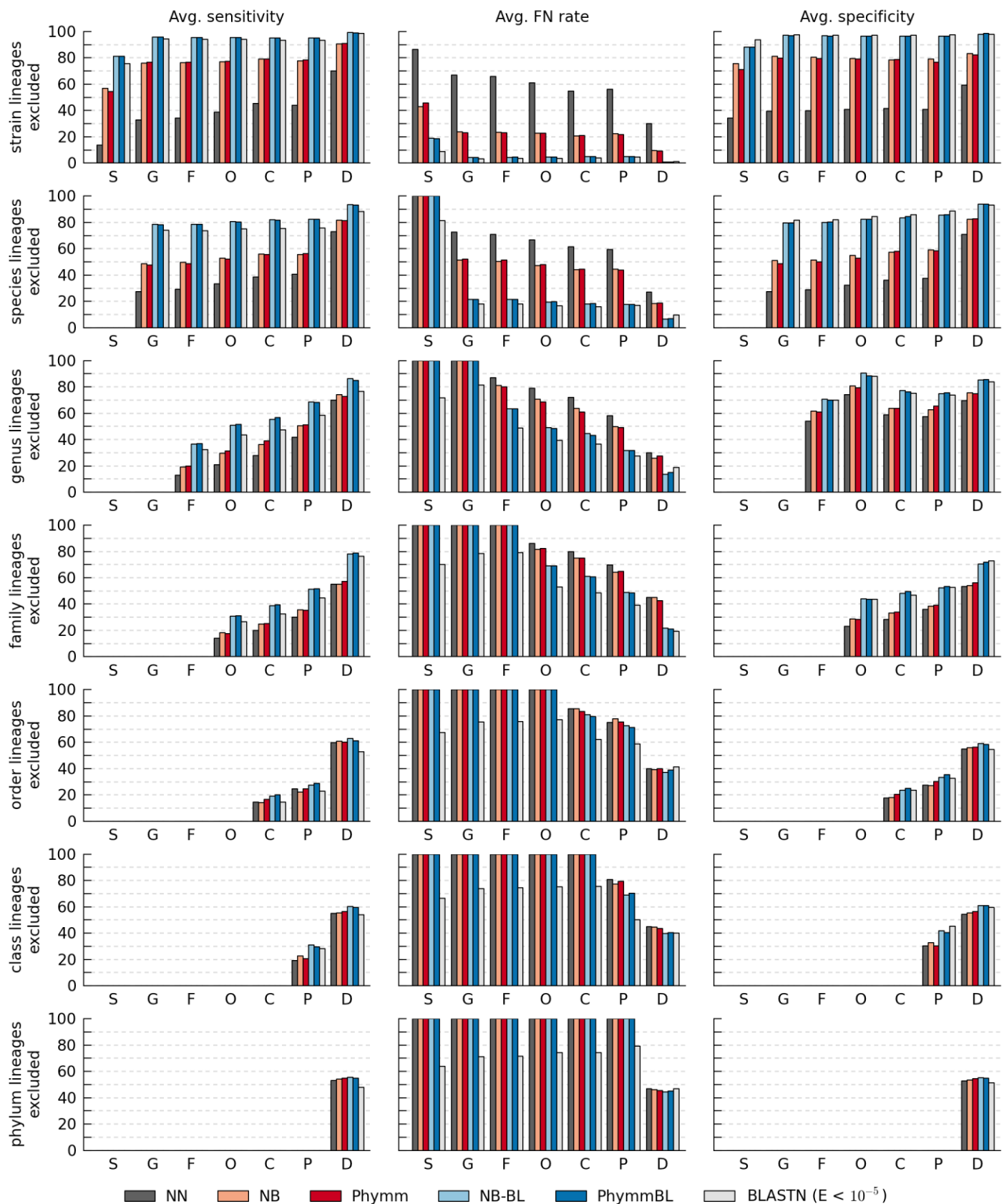


Figure S1. Average classification performance of rank-specific classifiers on 200 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the species (S), genus (G), family (F), order (O), class (C), phylum (P), or domain (D) of their source genome.

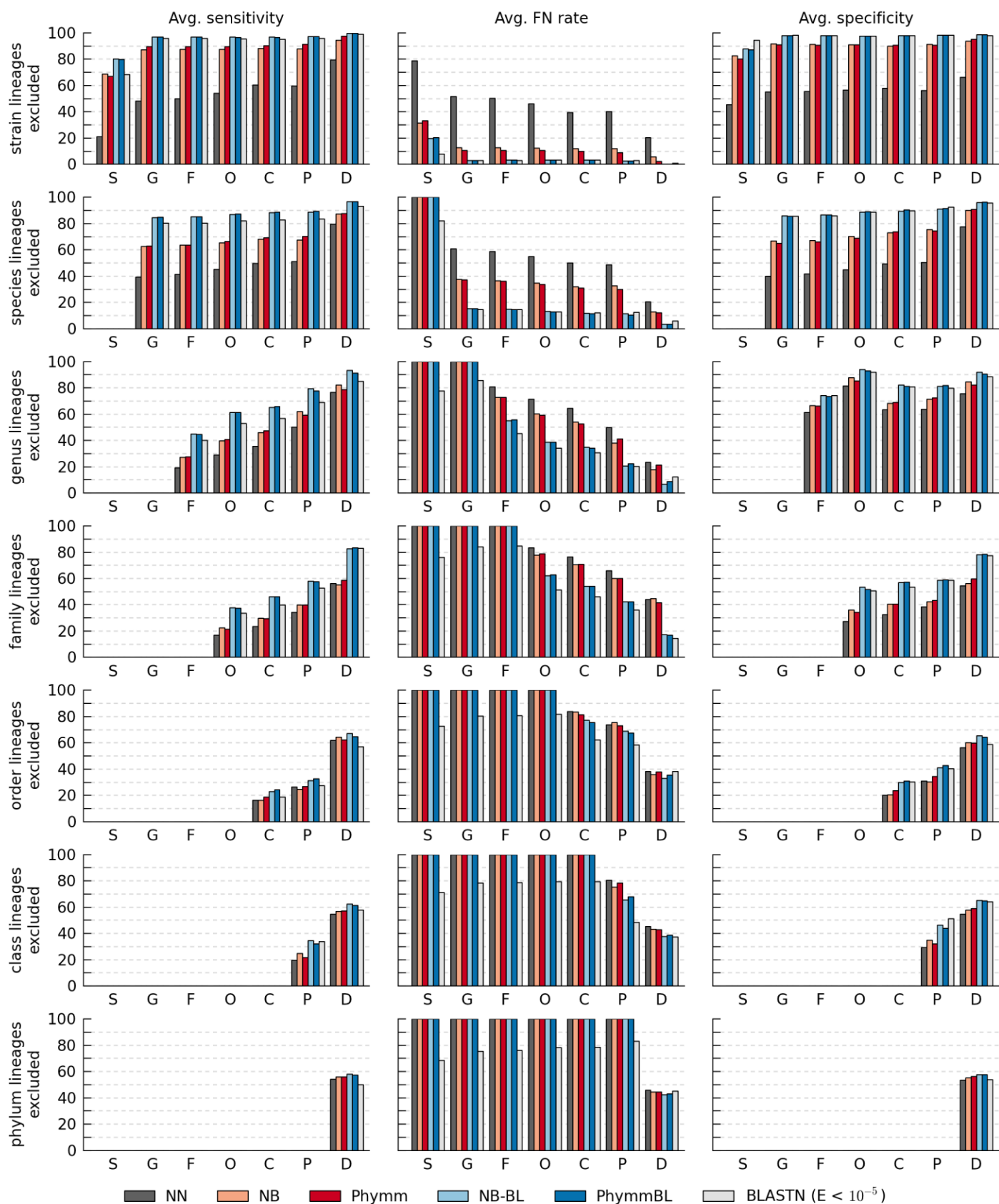


Figure S2. Average classification performance of rank-specific classifiers on 400 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the species (S), genus (G), family (F), order (O), class (C), phylum (P), or domain (D) of their source genome.

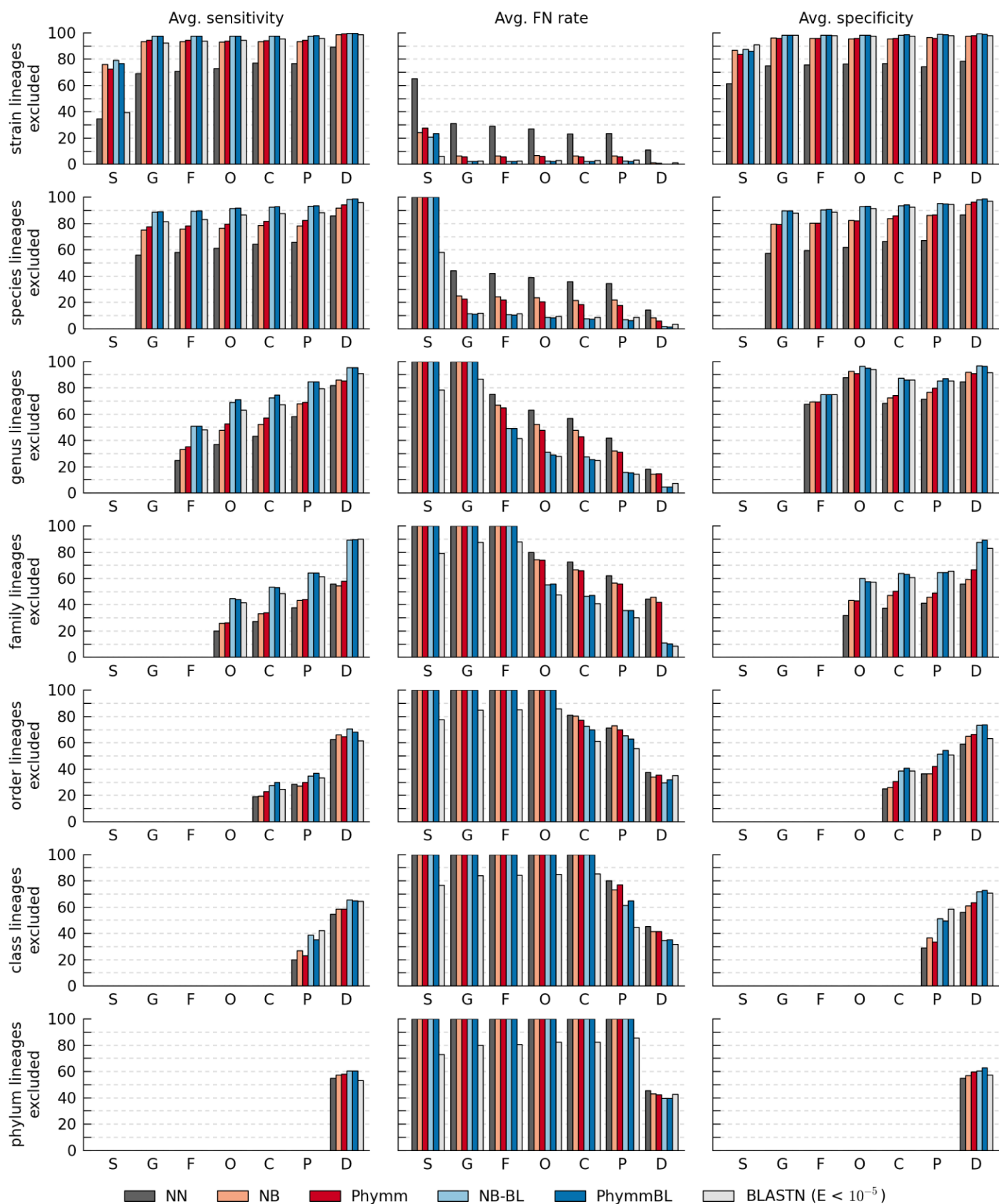


Figure S3. Average classification performance of rank-specific classifiers on 1000 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the species (S), genus (G), family (F), order (O), class (C), phylum (P), or domain (D) of their source genome.

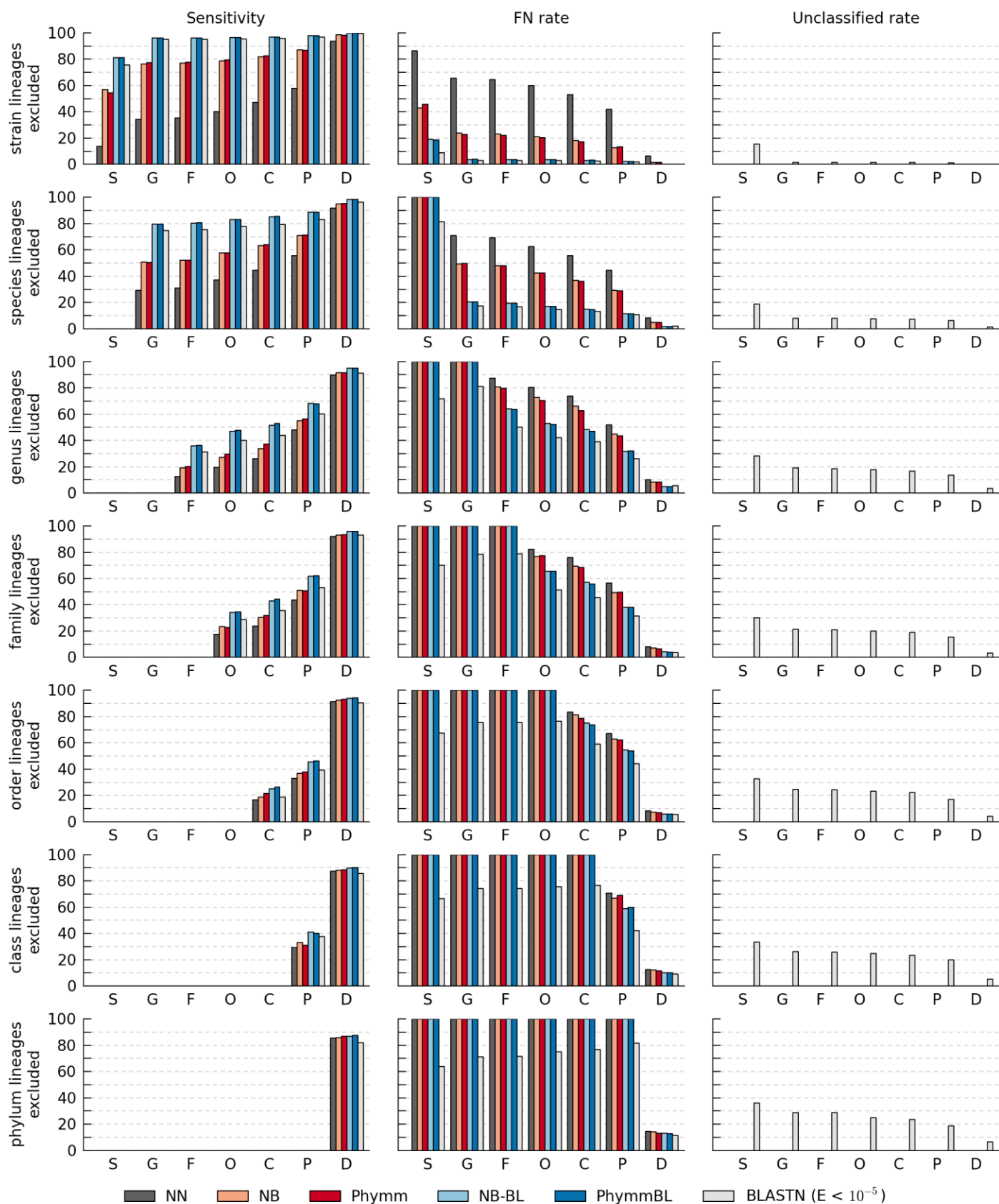


Figure S4. Absolute classification performance of rank-specific classifiers over all 200 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the species (S), genus (G), family (F), order (O), class (C), phylum (P), or domain (D) of their source genome.

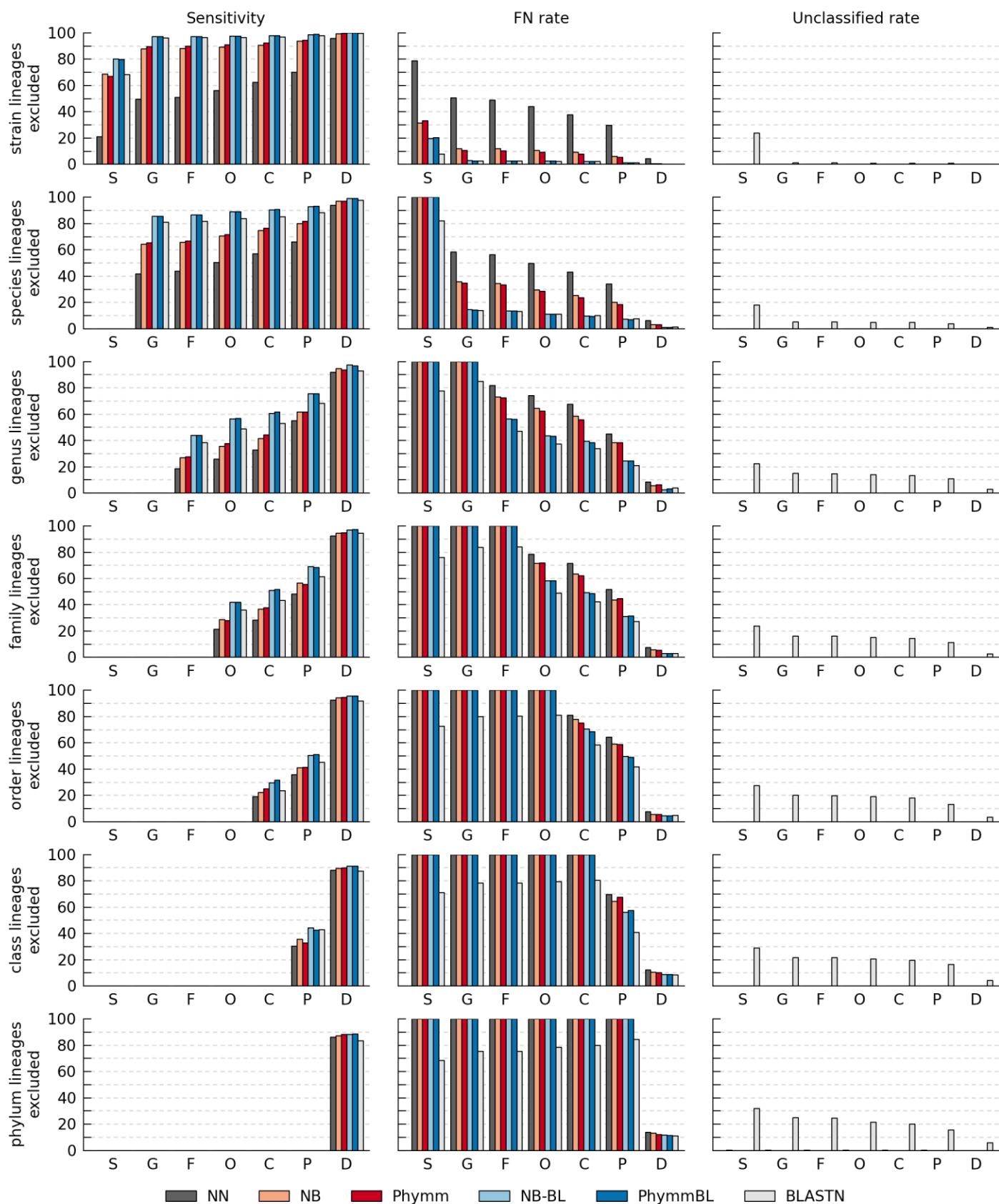


Figure S5. Absolute classification performance of rank-specific classifiers over all 400 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the species (S), genus (G), family (F), order (O), class (C), phylum (P), or domain (D) of their source genome.

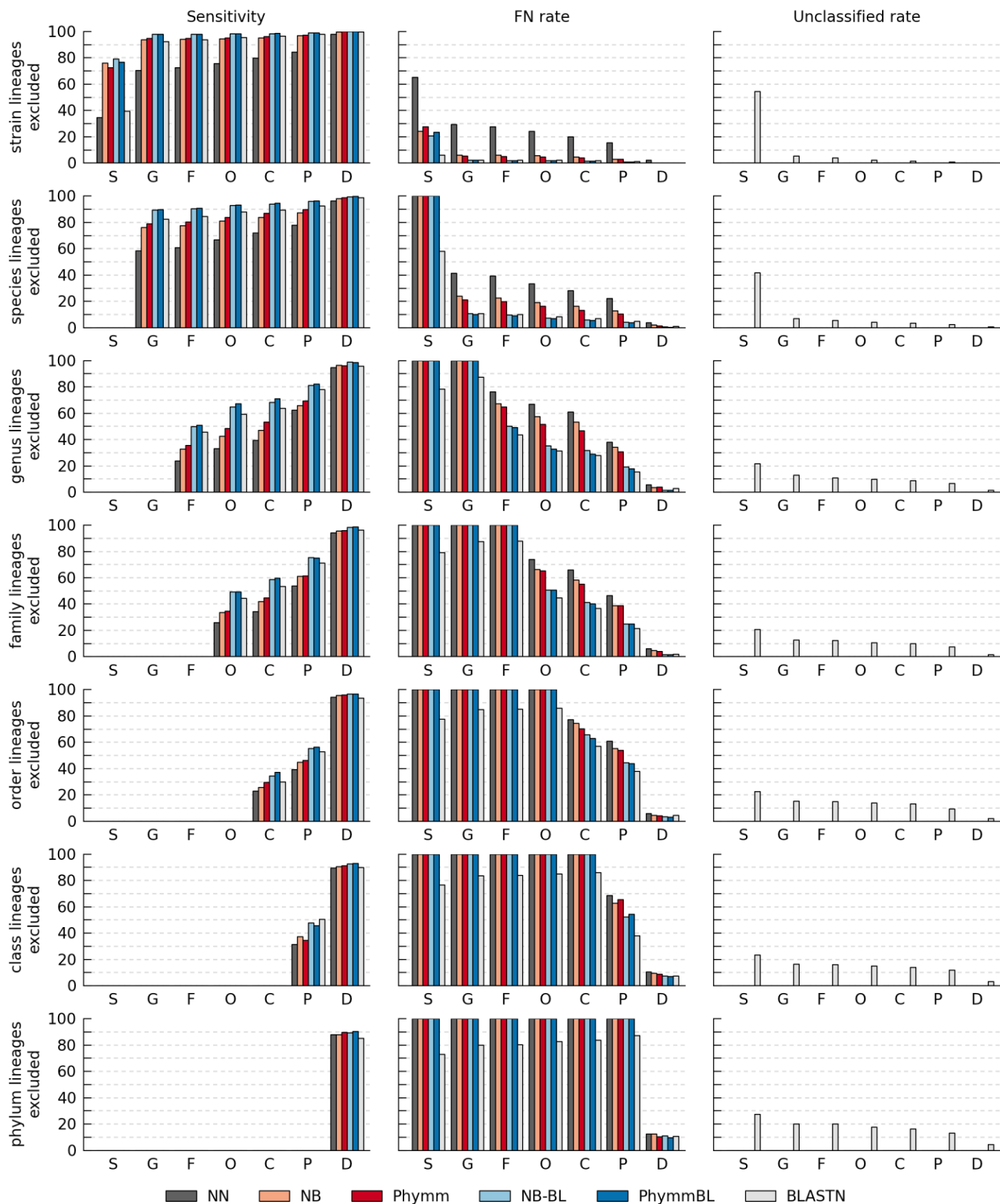


Figure S6. Absolute classification performance of rank-specific classifiers over all 1000 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the species (S), genus (G), family (F), order (O), class (C), phylum (P), or domain (D) of their source genome.