

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).

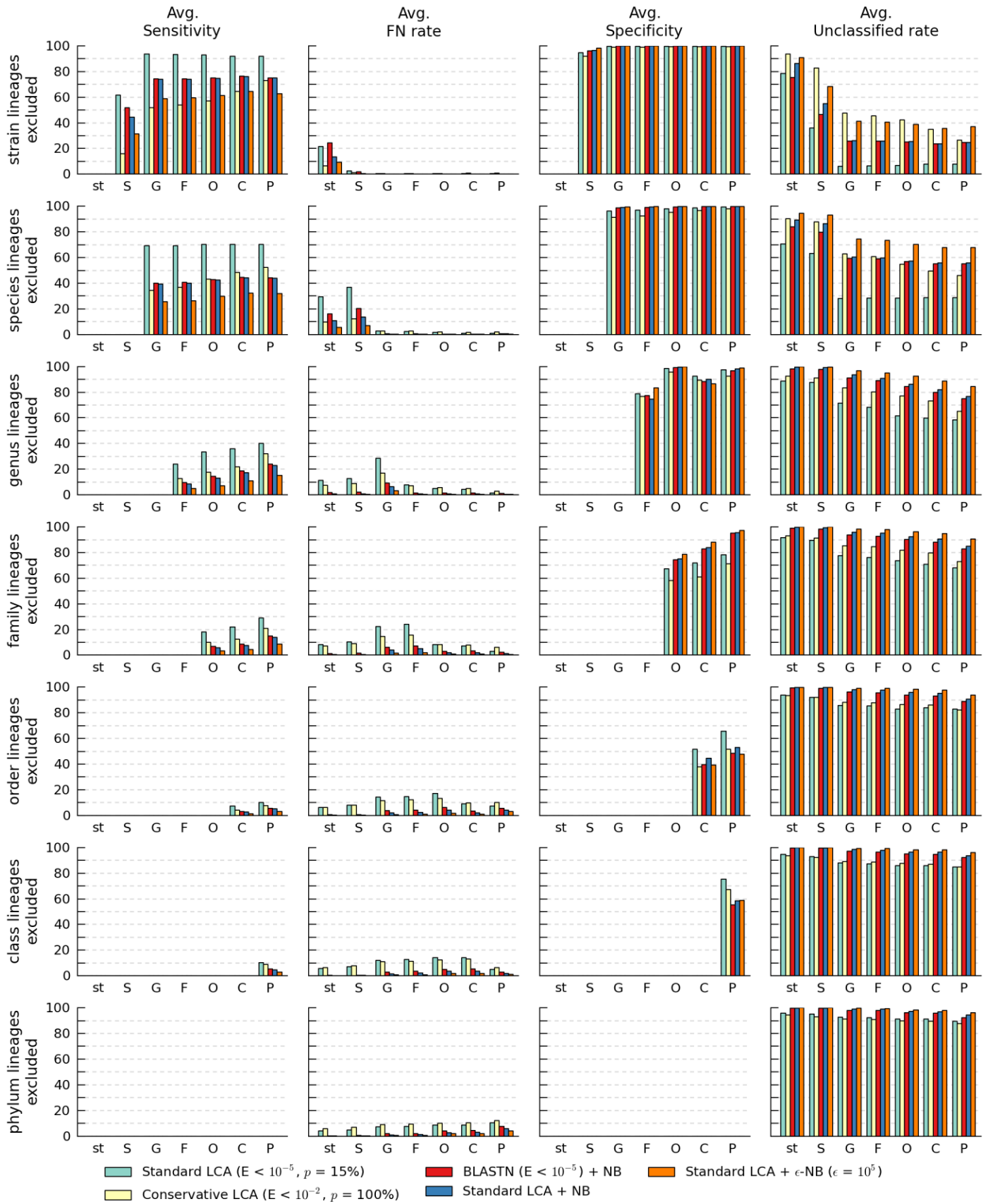


Figure S7. Average classification performance of rank-flexible classifiers on 200 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the strain (st), species (S), genus (G), family (F), order (O), class (C), or phylum (P) of their source genome.

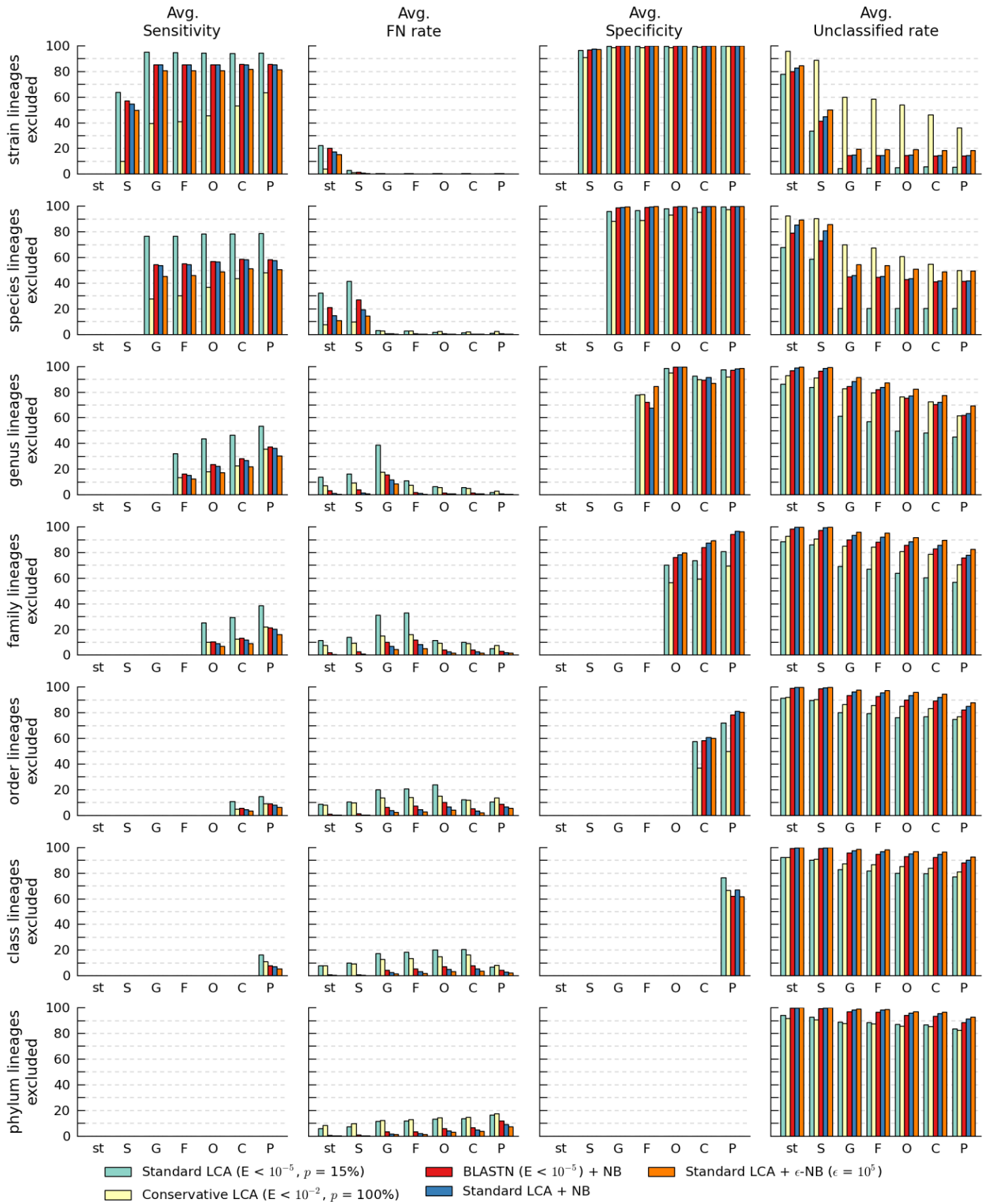


Figure S8. Average classification performance of rank-flexible classifiers on 400 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the strain (st), species (S), genus (G), family (F), order (O), class (C), or phylum (P) of their source genome.

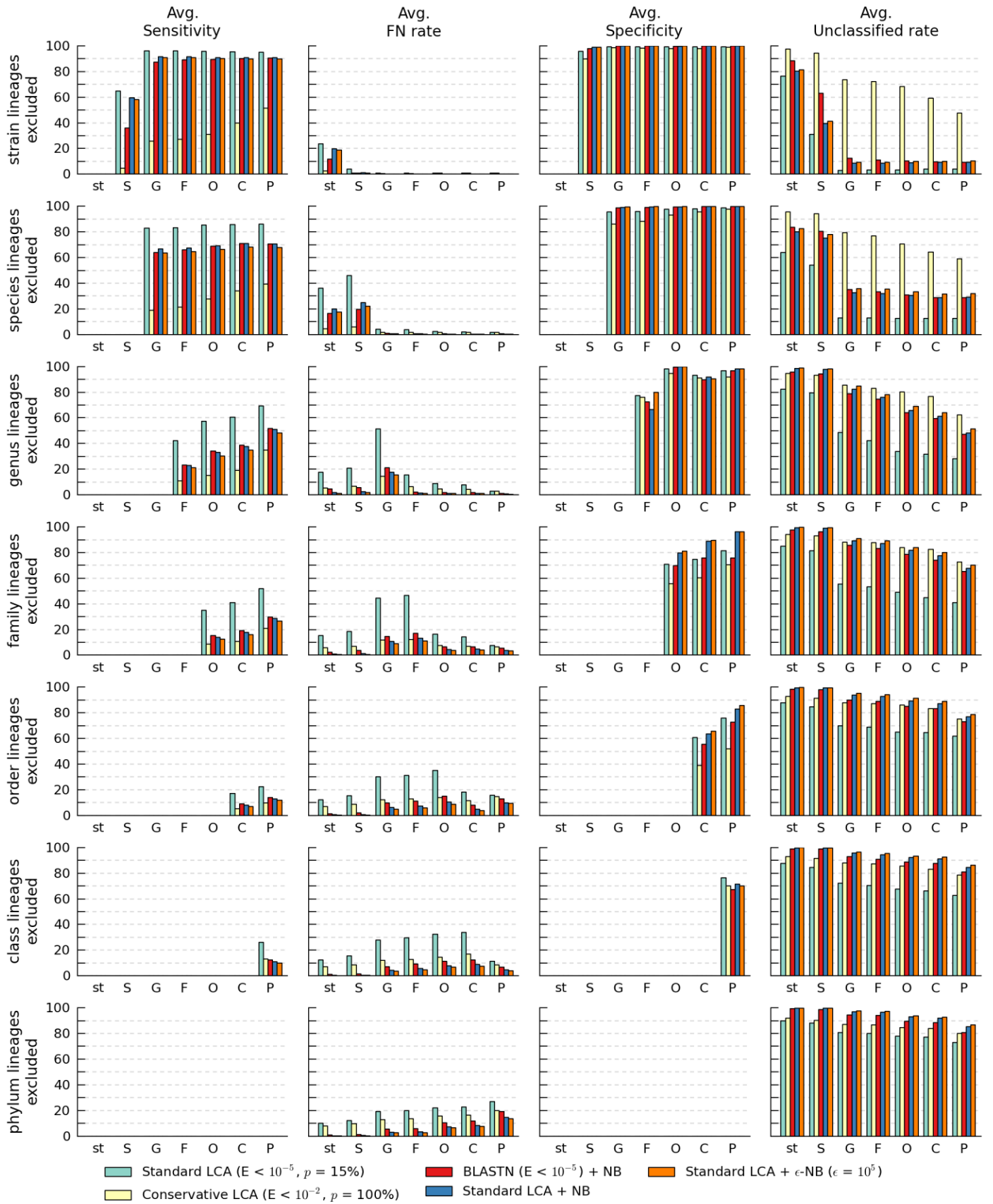


Figure S9. Average classification performance of rank-flexible classifiers on 1000 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the strain (st), species (S), genus (G), family (F), order (O), class (C), or phylum (P) of their source genome.

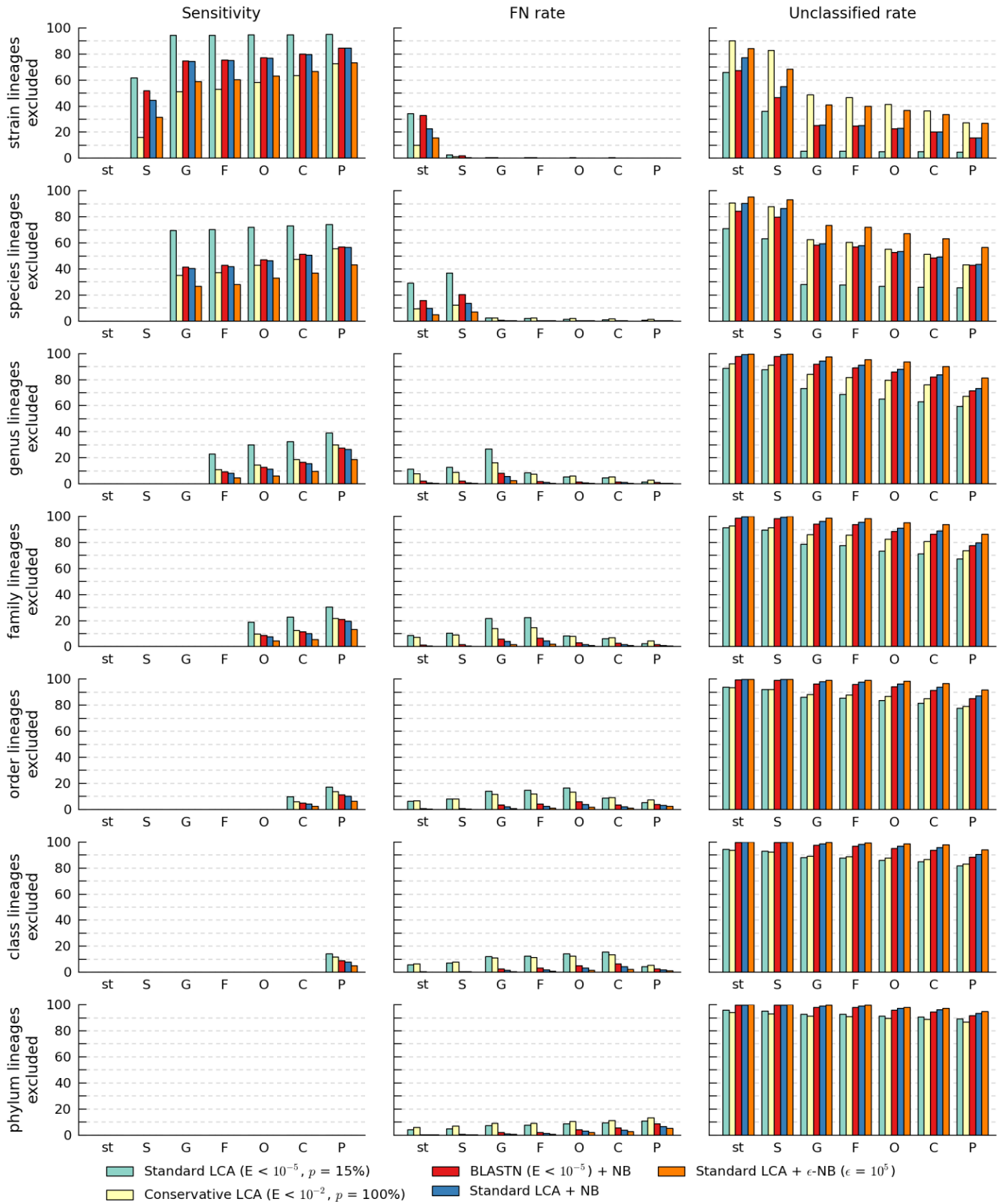


Figure S10. Absolute classification performance of rank-flexible classifiers over all 200 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the strain (st), species (S), genus (G), family (F), order (O), class (C), or phylum (P) of their source genome.

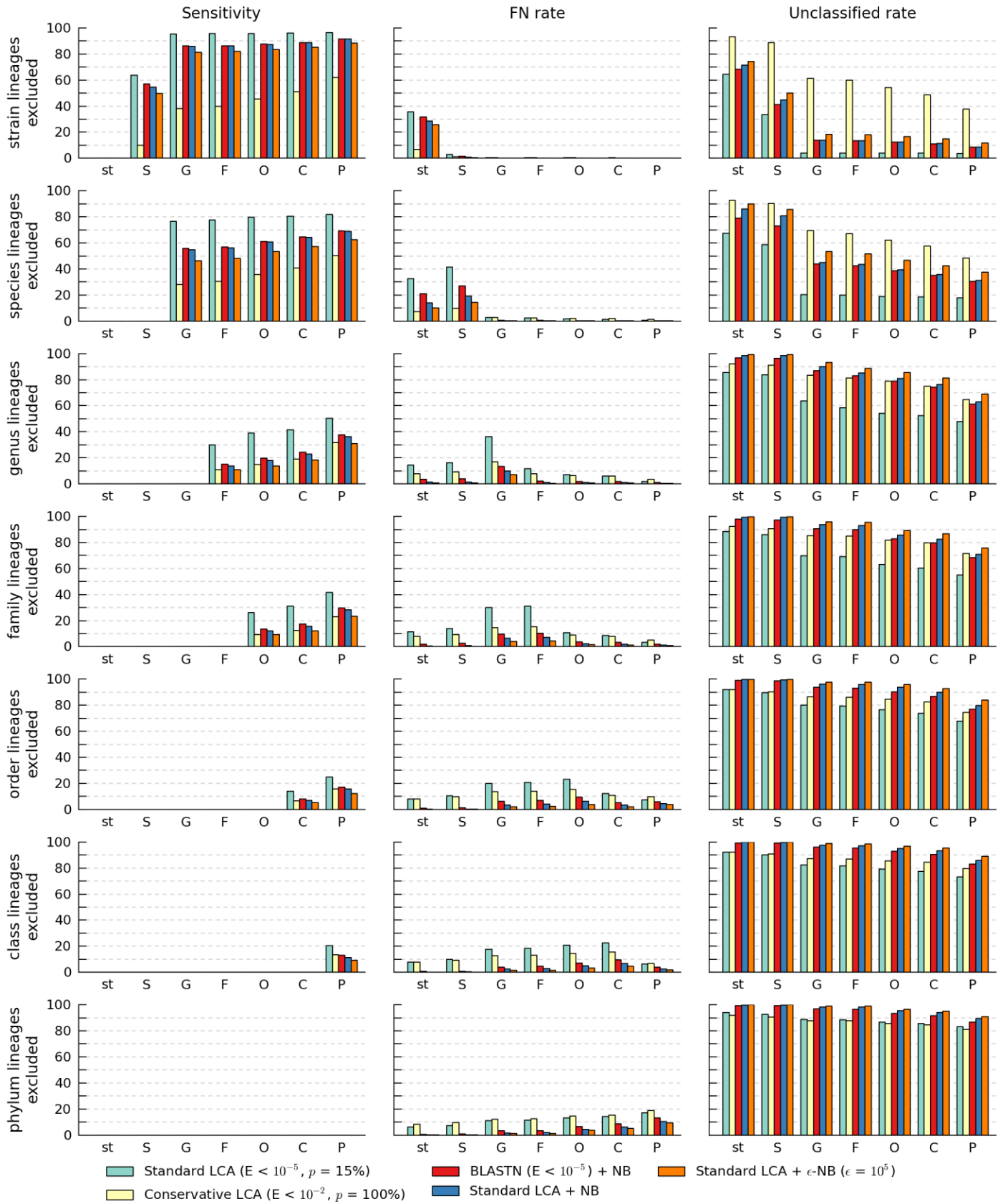


Figure S11. Absolute classification performance of rank-flexible classifiers over all 400 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the strain (st), species (S), genus (G), family (F), order (O), class (C), or phylum (P) of their source genome.

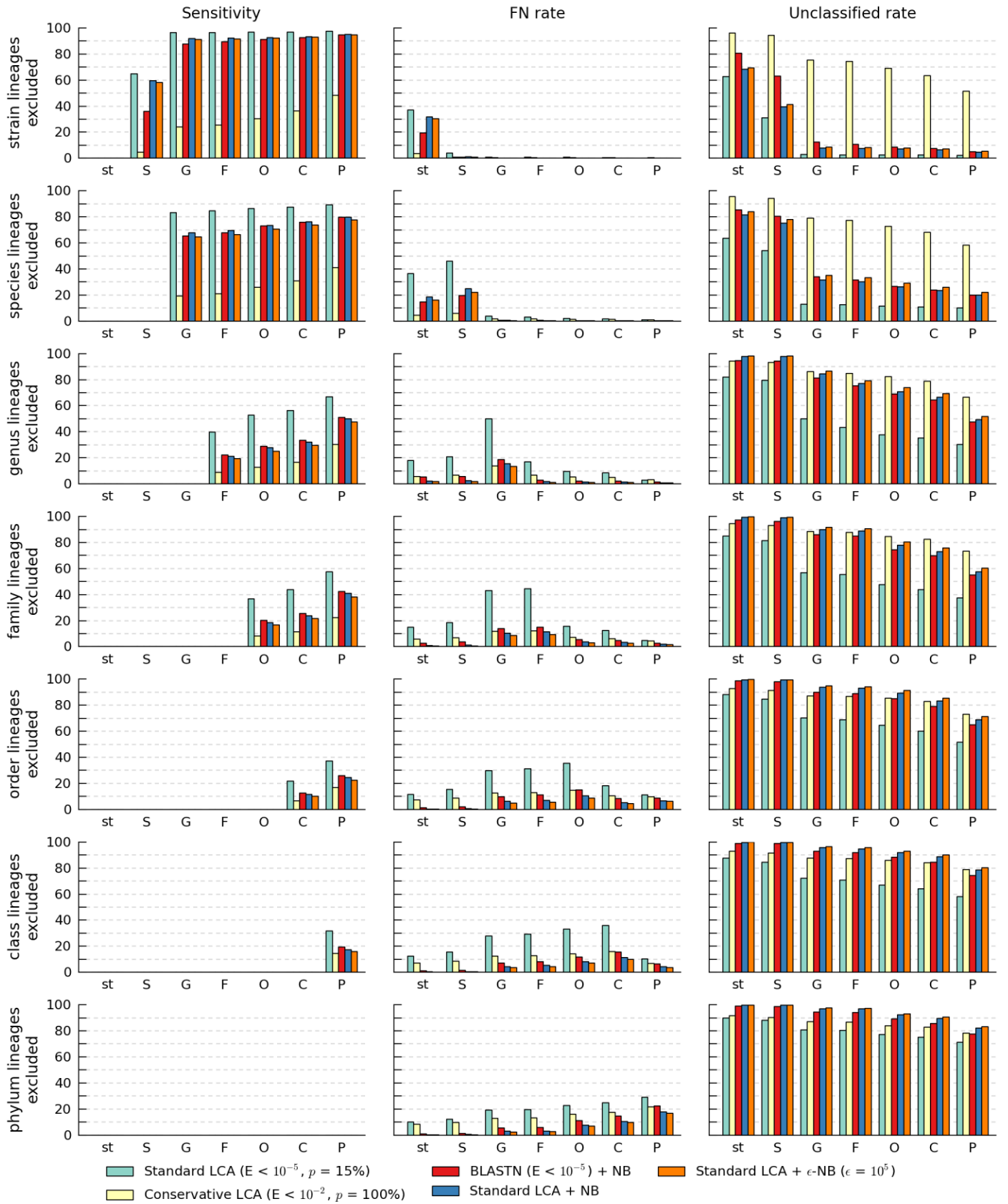


Figure S12. Absolute classification performance of rank-flexible classifiers over all 1000 bp query fragments using a leave-one-out evaluation framework. Fragments were considered correctly classified if they were assigned to the strain (st), species (S), genus (G), family (F), order (O), class (C), or phylum (P) of their source genome.