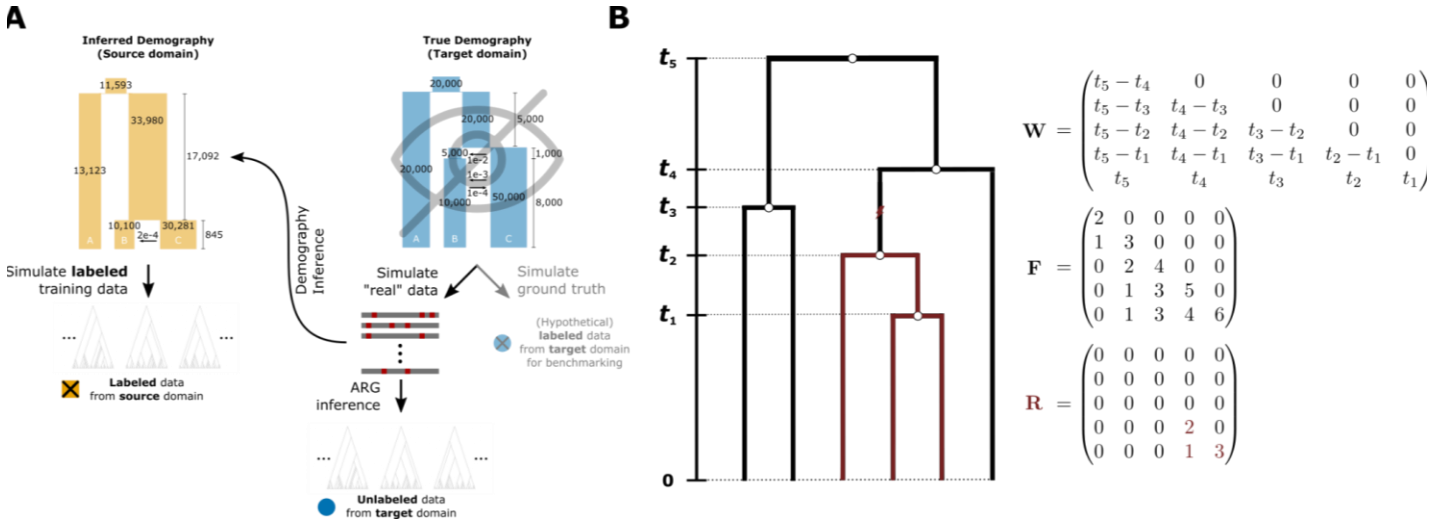
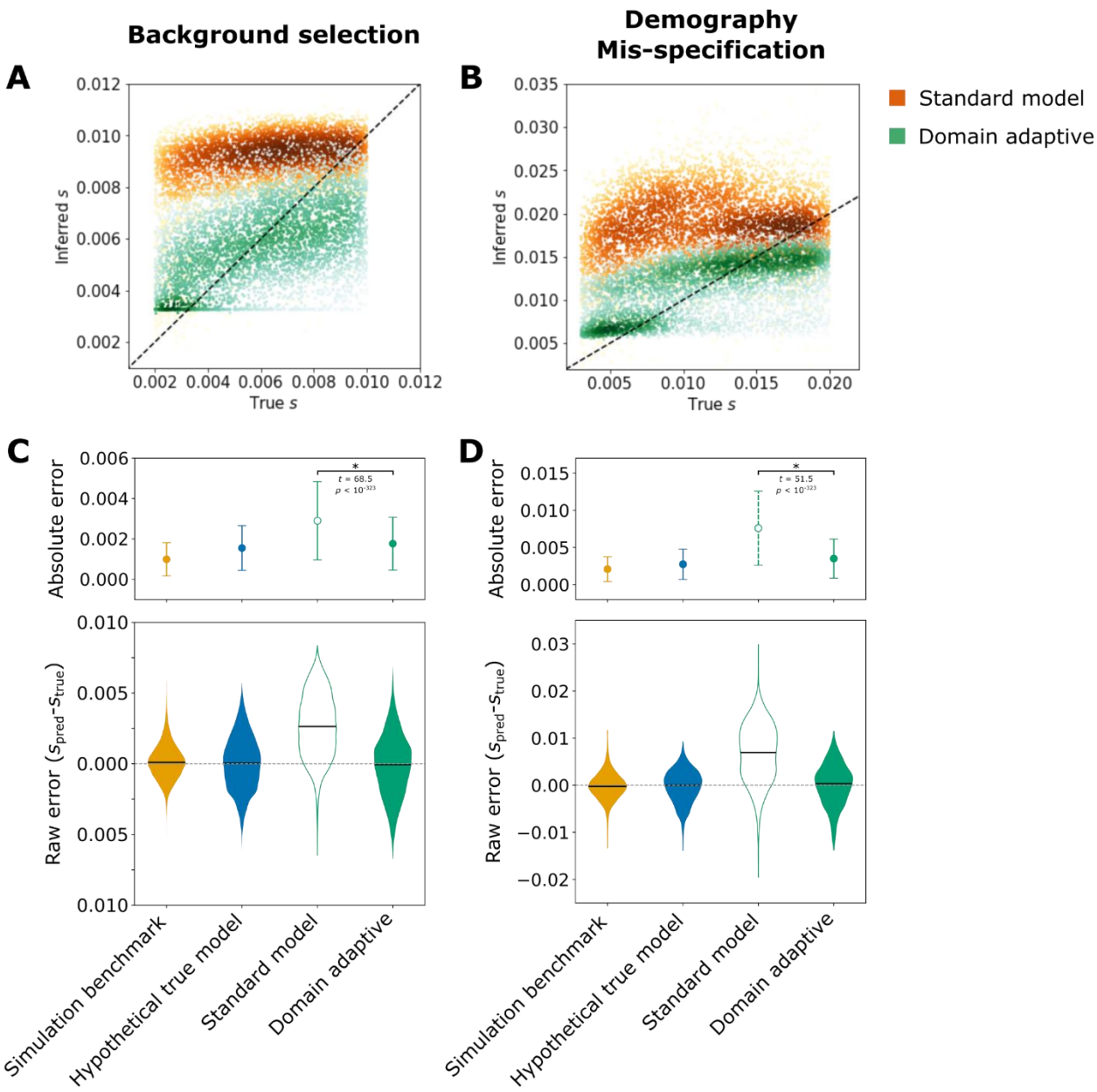
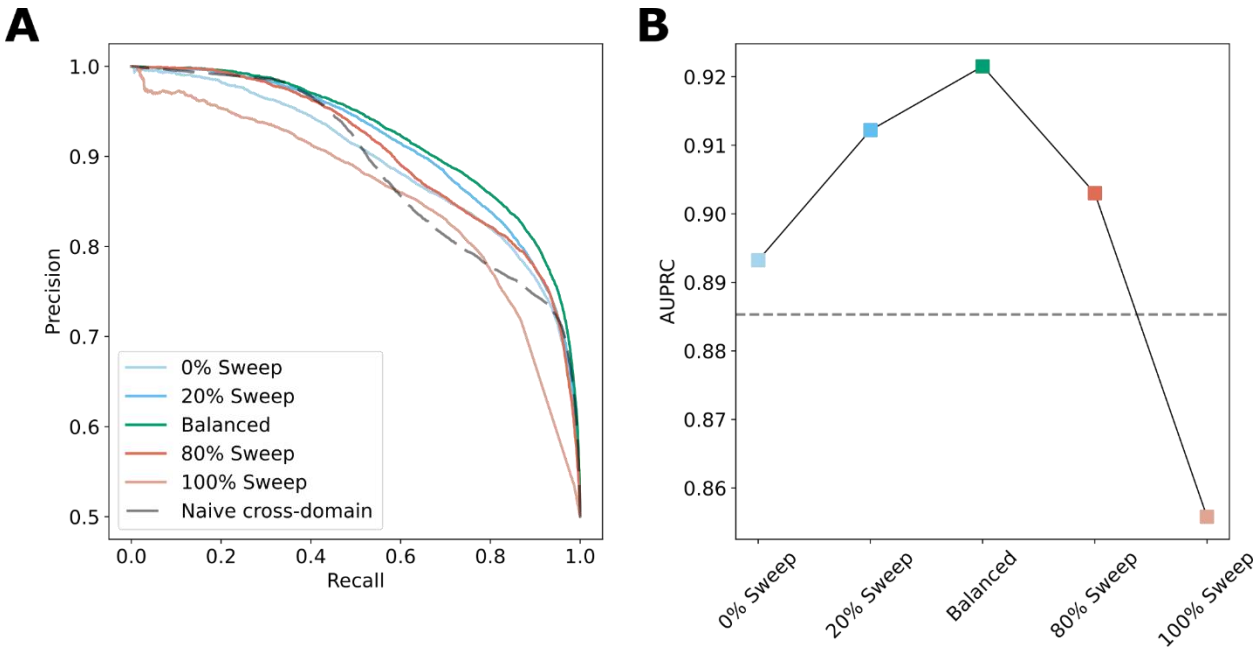
Supplementary Figures



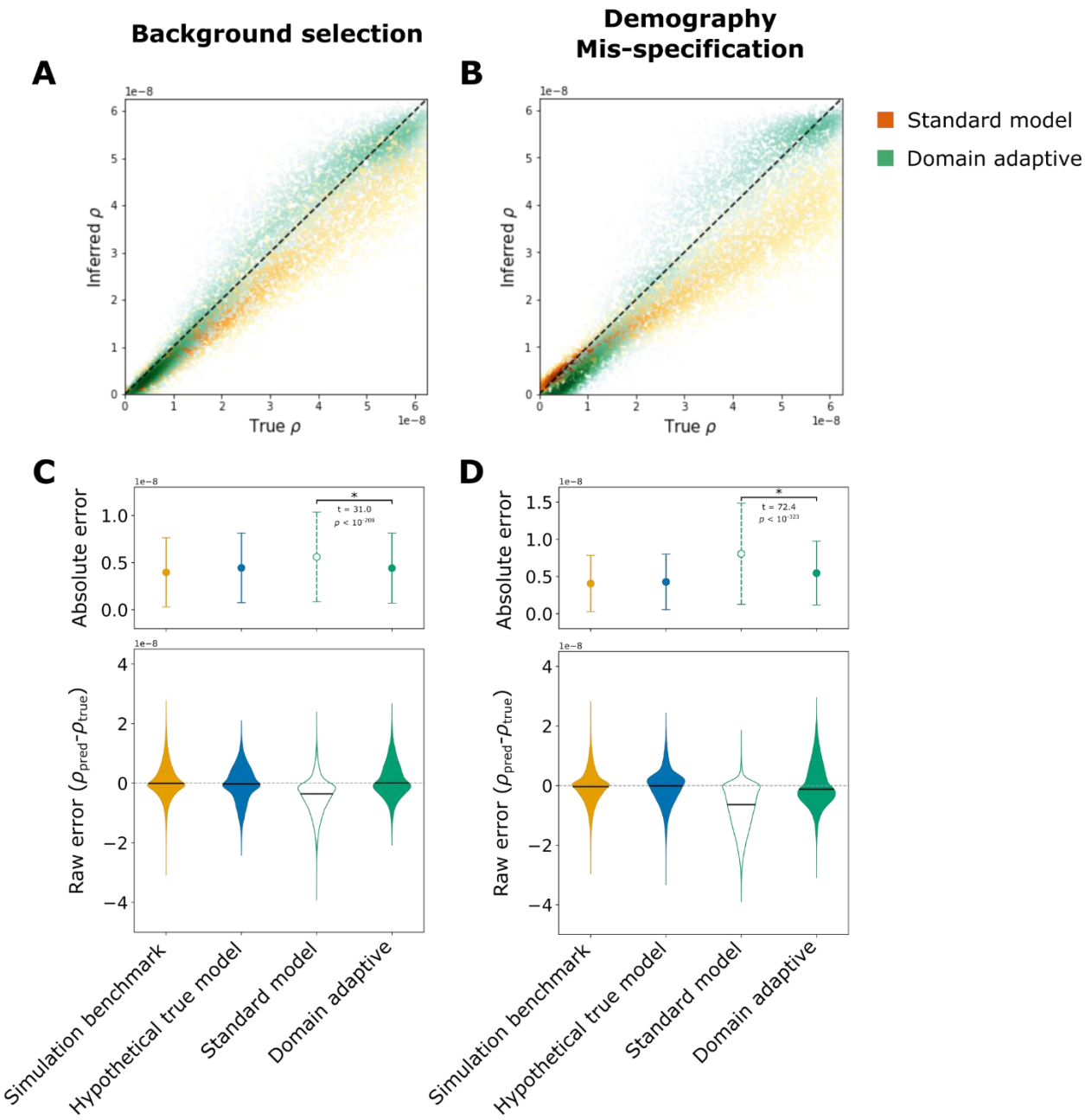
Supplementary Figure 1. Domain-adaptive SIA. A) The workflow of a simulation study that aims to benchmark the performance of the domain-adaptive SIA model in a realistic setting of demographic mis-specification. **B)** An improved version of SIA input features that encodes the full genealogy (adapted from (Kim et al. 2020)). A genealogy with *n* taxa at a polymorphic site is uniquely encoded by three (*n*-1) x (*n*-1) lower triangular matrices. The weight matrix **W** encodes the coalescent intervals where $w_{ij} = t_{n-j} - t_{n-1-i}, \forall\, i \geq j$, and the topology matrix **F** encodes the number of lineages persistent in the coalescent intervals *corresponding* to **W** (i.e. $f_{ij} = \#\ of\ lineages\ between\ t_{n-j} and\ t_{n-1-i}, \forall i \geq j$). The derived lineage matrix **R** encodes only the subtree subtending the branch where the mutation occurred (red lightning symbol), following the same scheme as **F**. Note that the **W** matrix is a redundant encoding of the *n*-1 coalescent times (t_1,t_2, …,t_{n-1}), which contains information roughly equivalent to the original SIA ARG features (Hejase et al. 2022).



675 **Supplementary Figure 2. Selection coefficient inference performance of SIA**
676 **models.** Raw data used to plot **Figs. 3B** and **3D** are presented in (**A**) and (**B**),
677 respectively. Performance of SIA models in the simulation experiment of failure to account
678 for background selection (**C**) and in the simulation experiment of demographic model mis-
679 specification (**D**) is presented in terms of mean and standard deviation of the absolute
680 error (top) as well as the distribution of raw error (bottom). Statistical significance (*) of
681 the difference between the absolute error of the standard model and that of the domain-
682 adaptive model is evaluated with Welch's *t*-test. See **Fig. 1C** for definition of the model
683 labels.



Supplementary Figure 3. Sweep classification performance of domain-adaptive SIA models trained with imbalanced target domain data. The classification performance of domain-adaptive SIA models trained with different proportions of sweep vs. neutral examples in the target domain is shown in the form of precision-recall curves (**A**) and the value of the area under precision-recall curve (AUPRC) (**B**). The dashed line in **(B)** indicates AUPRC of the standard model.



690 **Supplementary Figure 4**. **Recombination rate inference performance of ReLERNN**
691 **models.** Raw data used to plot **Figs. 4A** and **4B** are presented in (**A**) and (**B**),
692 respectively. Performance of ReLERNN models in the simulation experiment of failure to
693 account for background selection (**C**) and in the simulation experiment of demographic
694 model mis-specification (**D**) is presented in terms of mean and standard deviation of the
695 absolute error (top) as well as the distribution of raw error (bottom). Statistical significance
696 (*) of the difference between the absolute error of the standard model and that of the
697 domain-adaptive model is evaluated with Welch's *t*-test. See **Fig. 1C** for definition of the
698 model labels.