Figure S5  Heterogeneity of accuracies across population sub-groups
In each panel, principal component (PC) 2 vs. PC 1 of genotypic data is plotted to show population sub-groups. Each panel corresponds to a dataset-imputation method combination. Individuals are color coded according to their overall average imputation accuracy on an individual genotype basis, $R_i^2$. The colors yellow, orange, red, pink, purple, blue, and black correspond to $0.9 \leq R_i^2 \leq 1$, $0.8 \leq R_i^2 < 0.9$, $0.7 \leq R_i^2 < 0.8$, $0.6 \leq R_i^2 < 0.7$, $0.5 \leq R_i^2 < 0.6$, $0.4 \leq R_i^2 < 0.5$, $0.3 \leq R_i^2 < 0.4$ respectively. Panels A-D correspond to the Cornell winter wheat (WW) data imputed with k nearest neighbors imputation (kNN; A), singular value decomposition imputation (SVD; B), random forest imputation (RFI, C), expectation maximization imputation (EMI, D). Panels E-H correspond to the CIMMYT elite spring wheat data imputed with kNN (E), SVD (F), RFI (G), EMI (H). Panels I-L correspond to the CIMMYT drought tolerant maize data imputed with kNN (I), SVD (J), RFI (K), EMI (L). Panels M-P correspond to the North American barley data imputed with kNN (M), SVD (N), RFI (O), EMI (P).