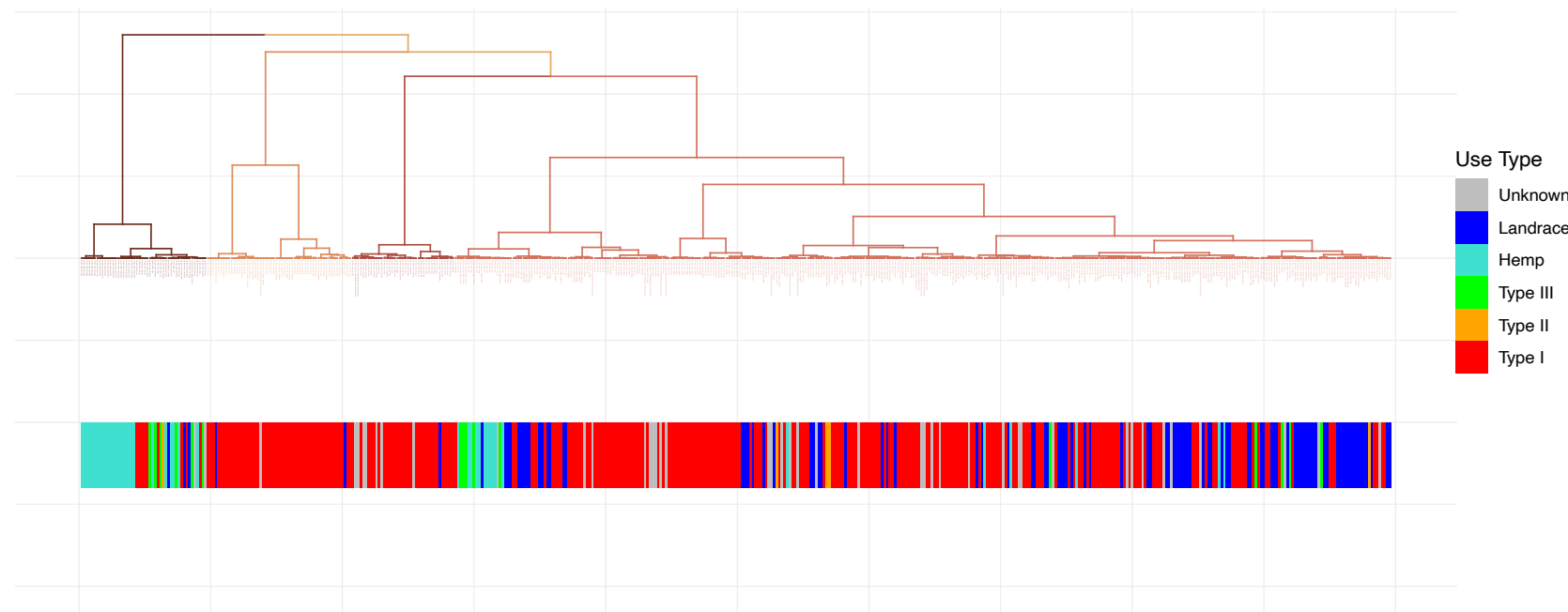
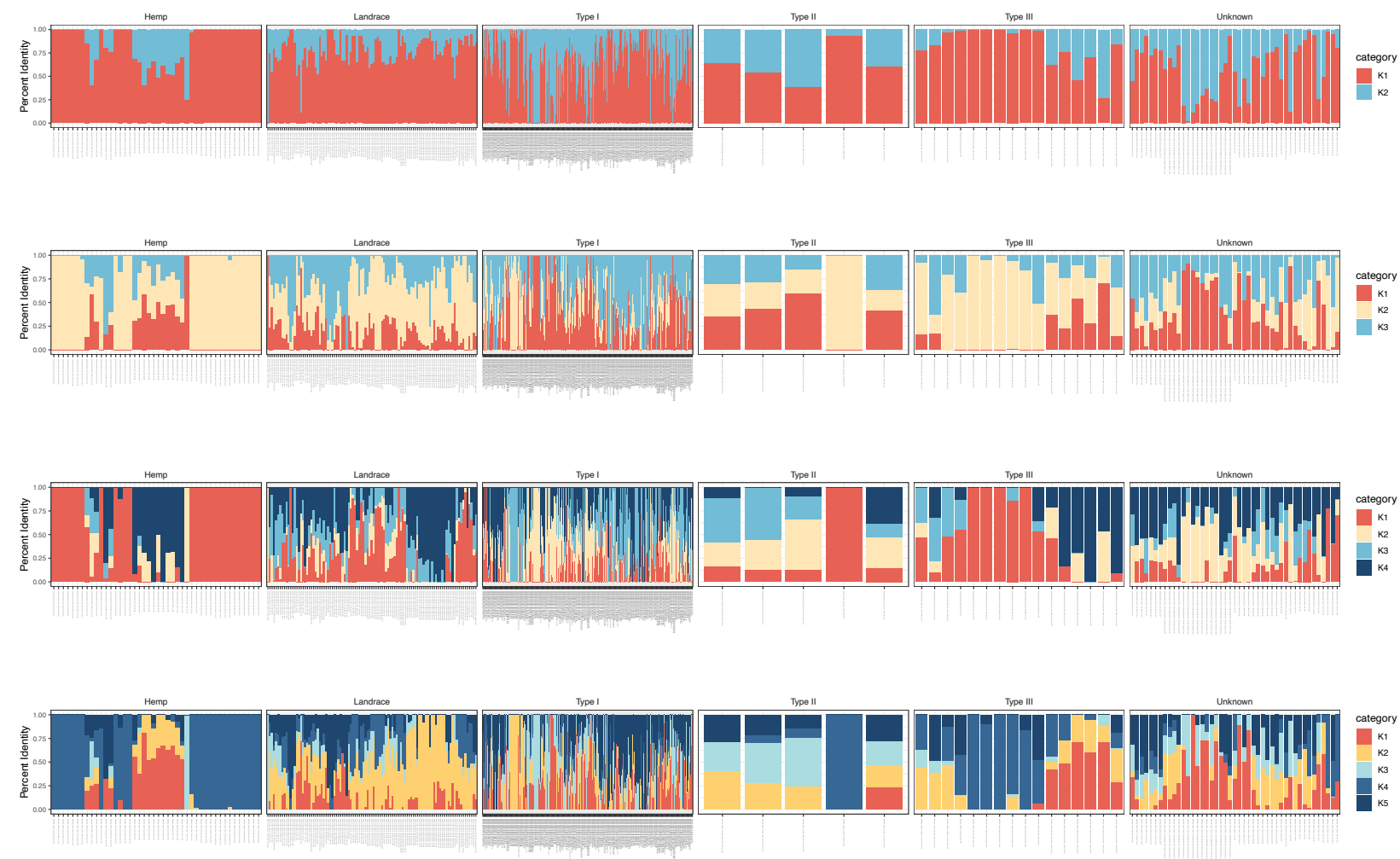


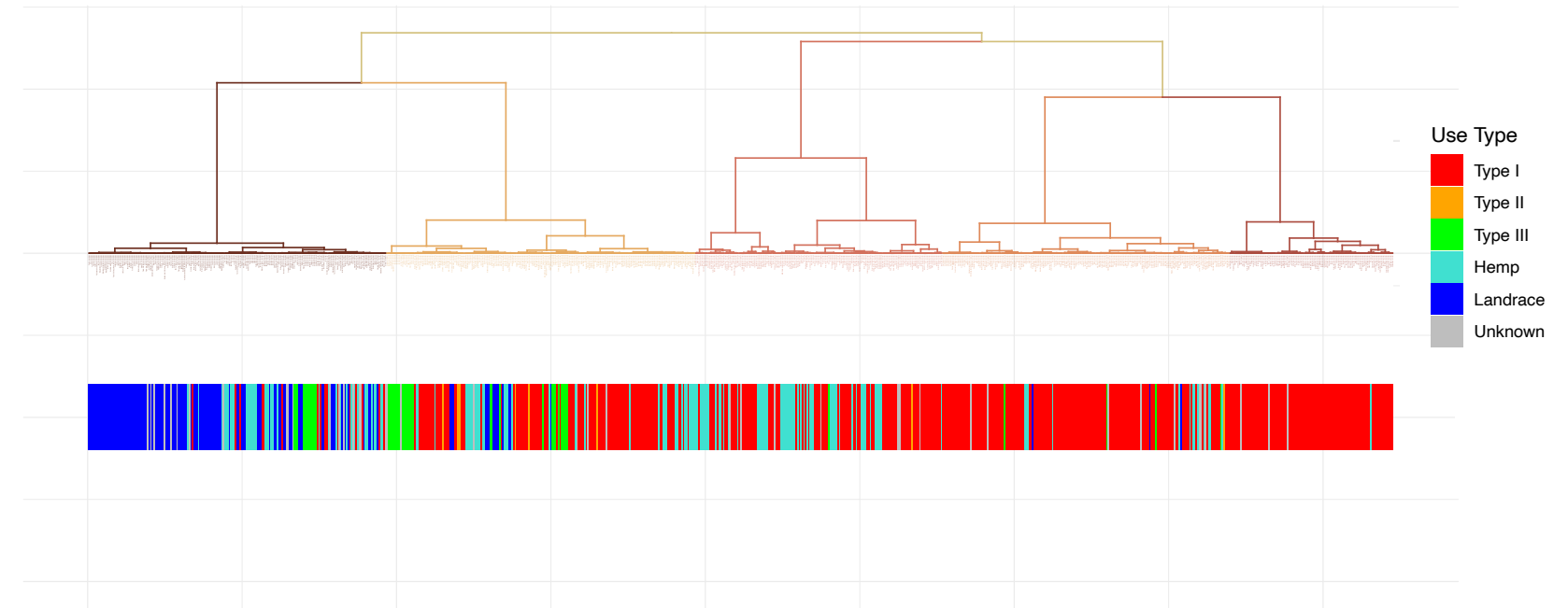
A



B



C



D

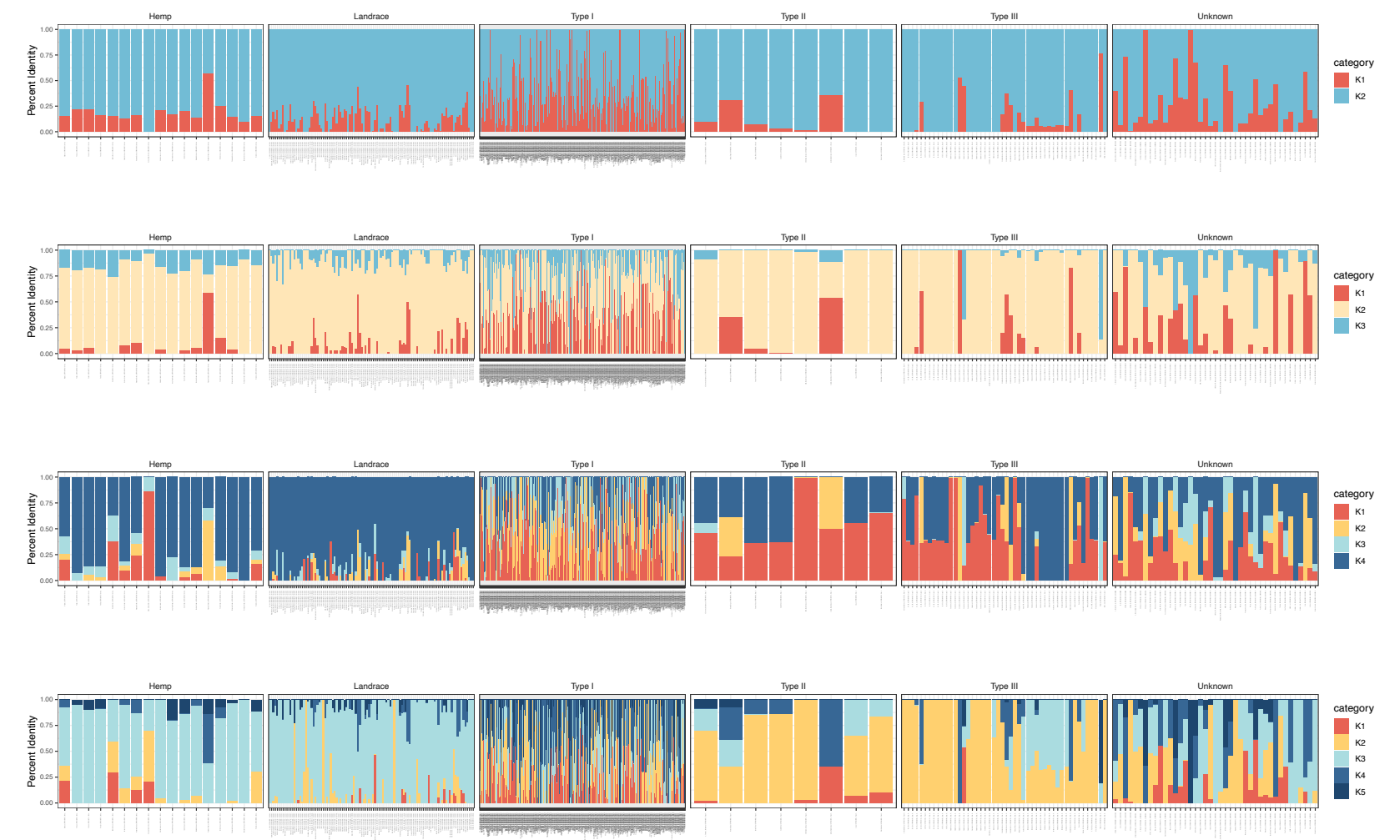
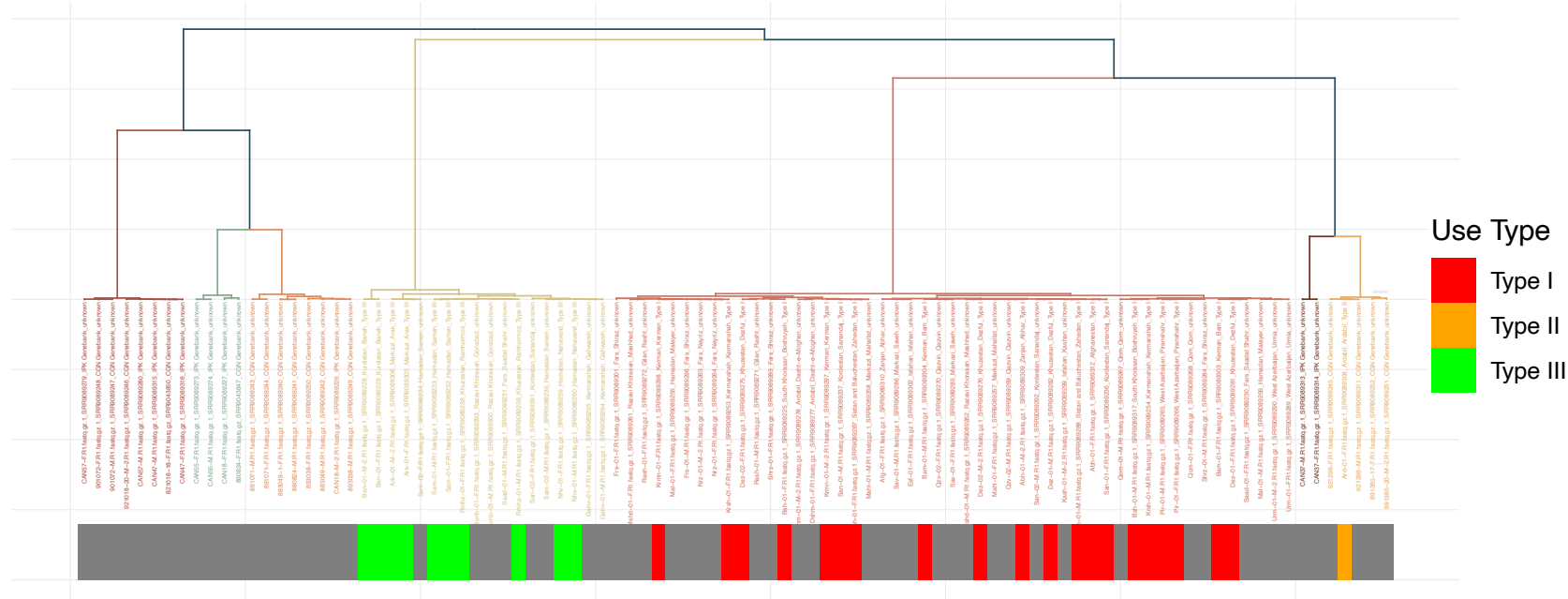
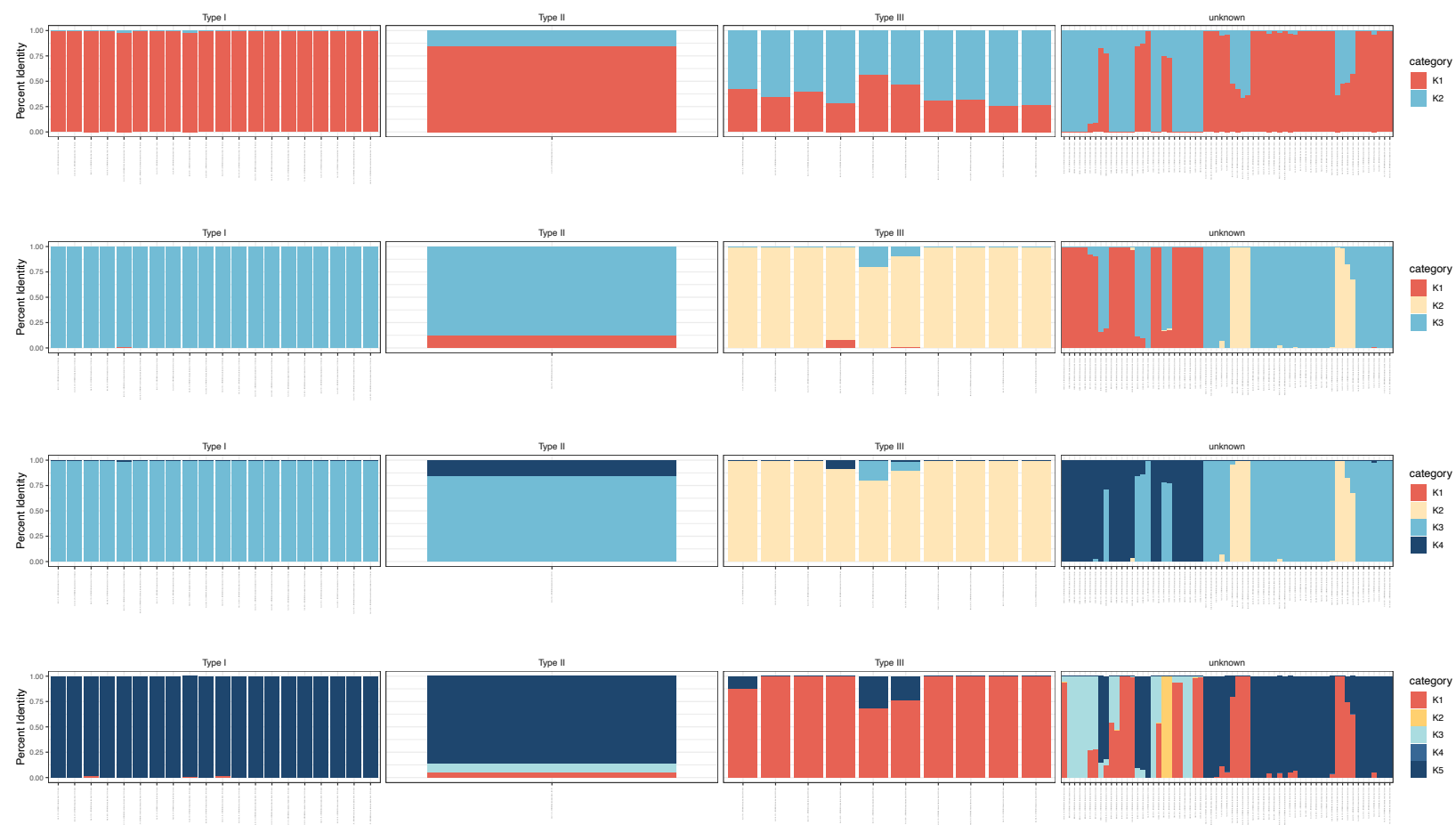


Figure 1 Examining hierarchical clustering on principal components (HCPC) and population structure in the LeafWorks Inc. (n=498) and Phylos Biosciences (n=845) datasets. In each case population genetic clustering was conducted based only on nuclear genetic SNPs while reported use-type within the dataset is below in solid bars to facilitate interpretation based upon community standards **(A)** Hierarchical cluster dendrogram from 520 nuclear SNPs for the LeafWorks Inc. dataset with use-type indicated below. Use-type are pictured below (Type I=288, Type II=5, Type III=16, Hemp=44, Landrace=101 and Unknown=44) **(B)** Visualization of population structure and admixture from 1,405 nuclear SNPs for the LeafWorks Inc. dataset using the fastSTRUCTURE software (k=2-5) with the optimal number of K being 4 using the silhouette method (**Fig. S9-10**) **(C)** Hierarchical cluster dendrogram from 292 nuclear SNPs for the Phylos Biosciences dataset with use-type indicated below. Use-type accessions include Type I=479, Type II=8, Type III=46, Landrace=127, Hemp=143 and Unknown=42 **(D)** Visualization of population structure and admixture from 385 nuclear SNPs for the Phylos Biosciences dataset using the fastSTRUCTURE software (k=2-5) with the optimal number of K being 3 using the Silhouette method (**Fig. S9-10**).

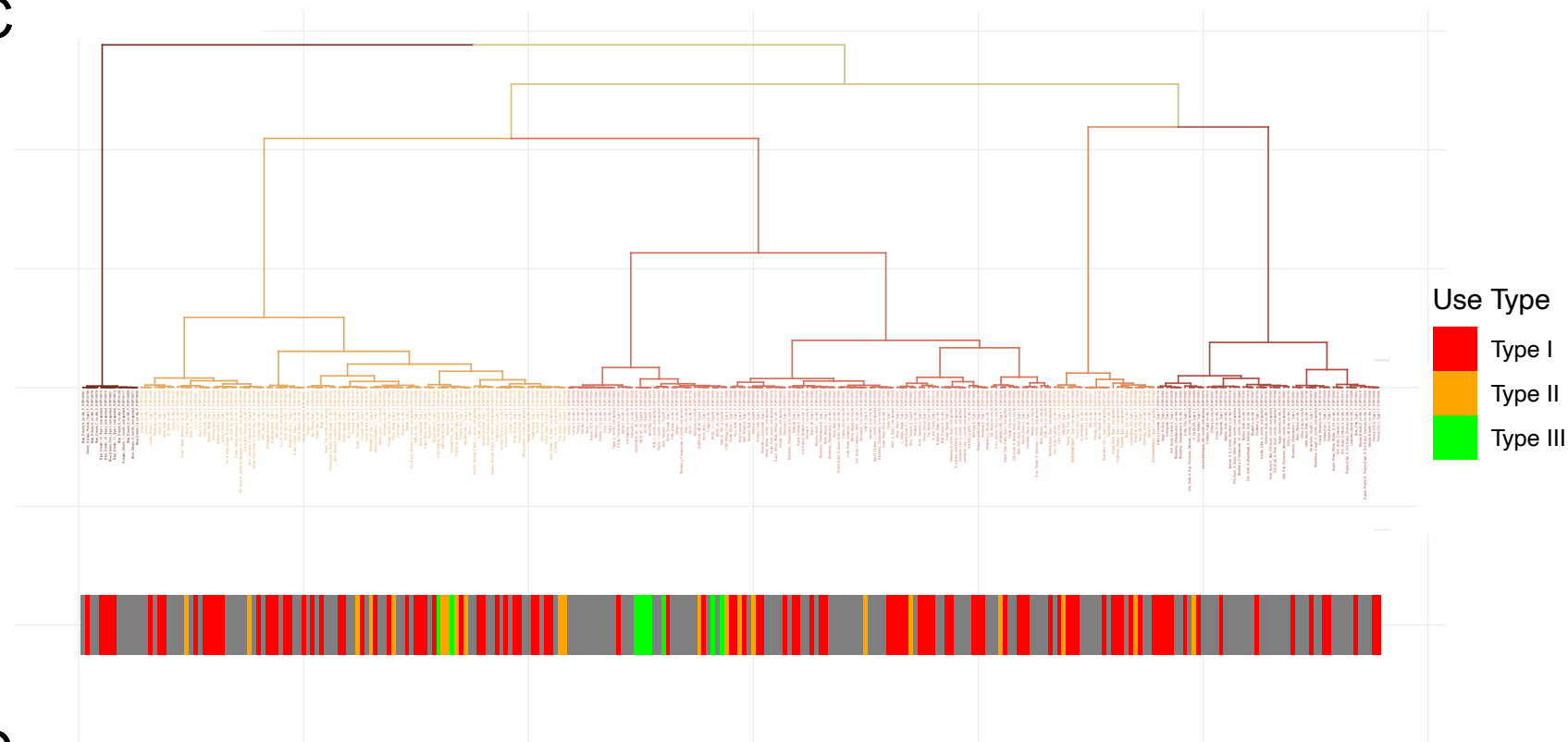
A



B



C



D

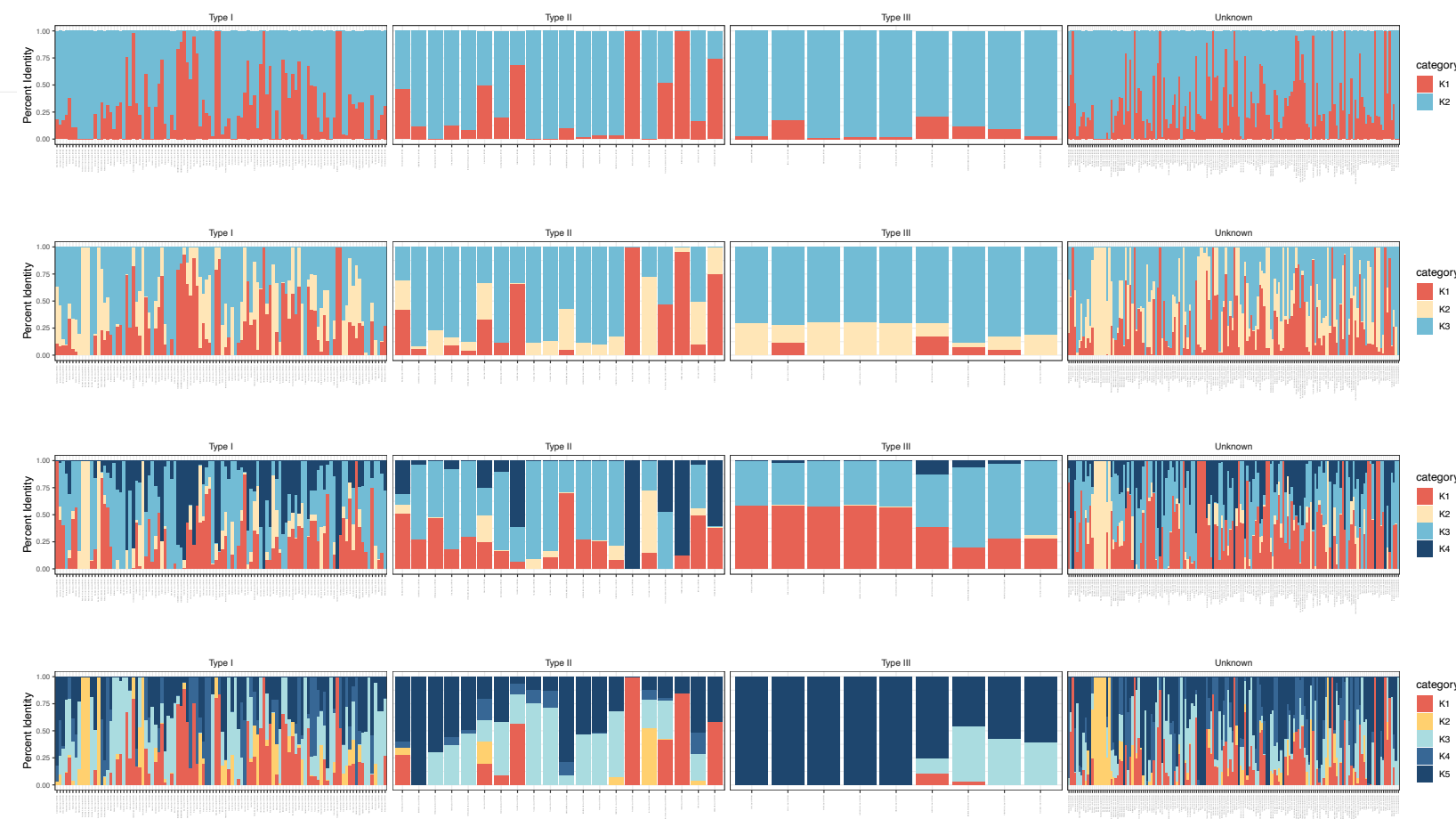
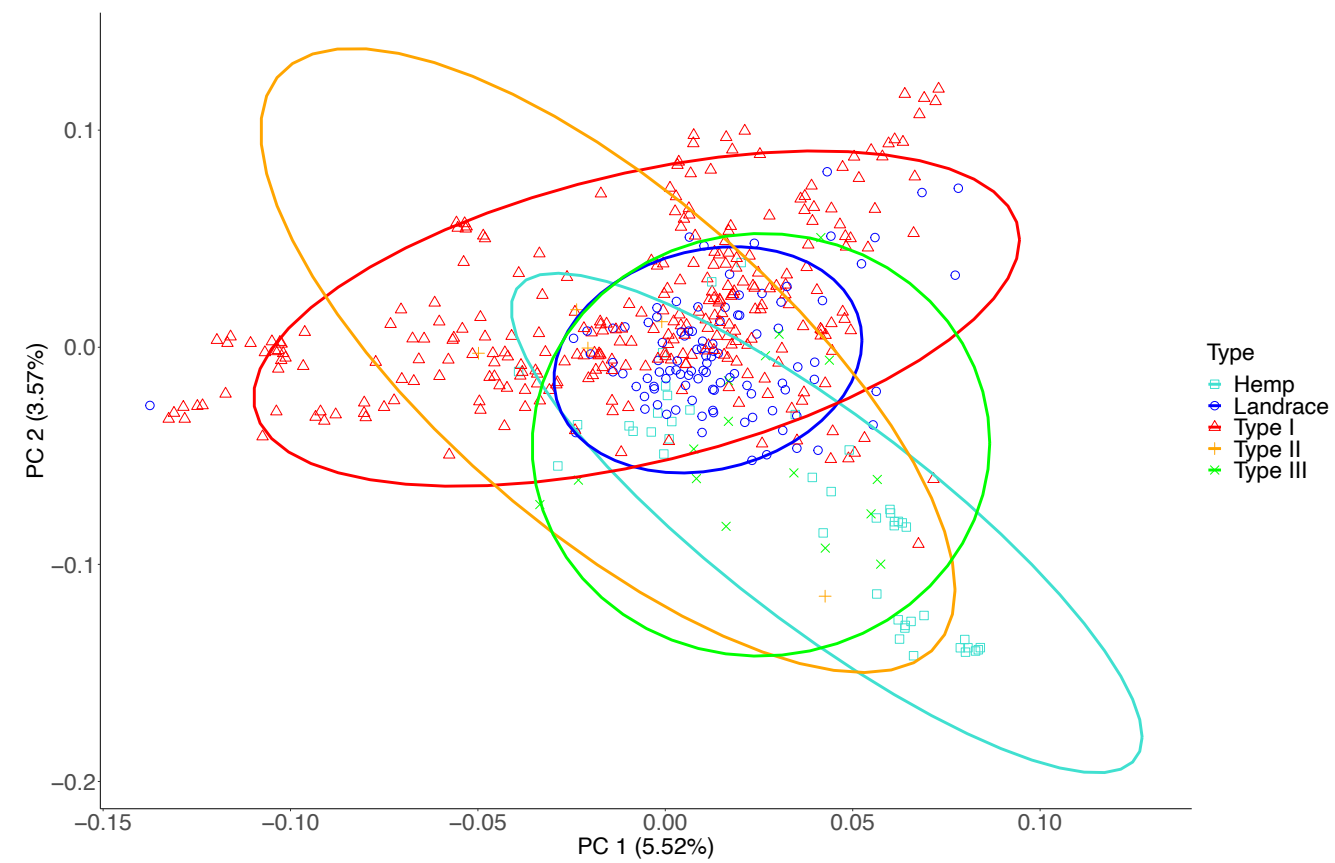
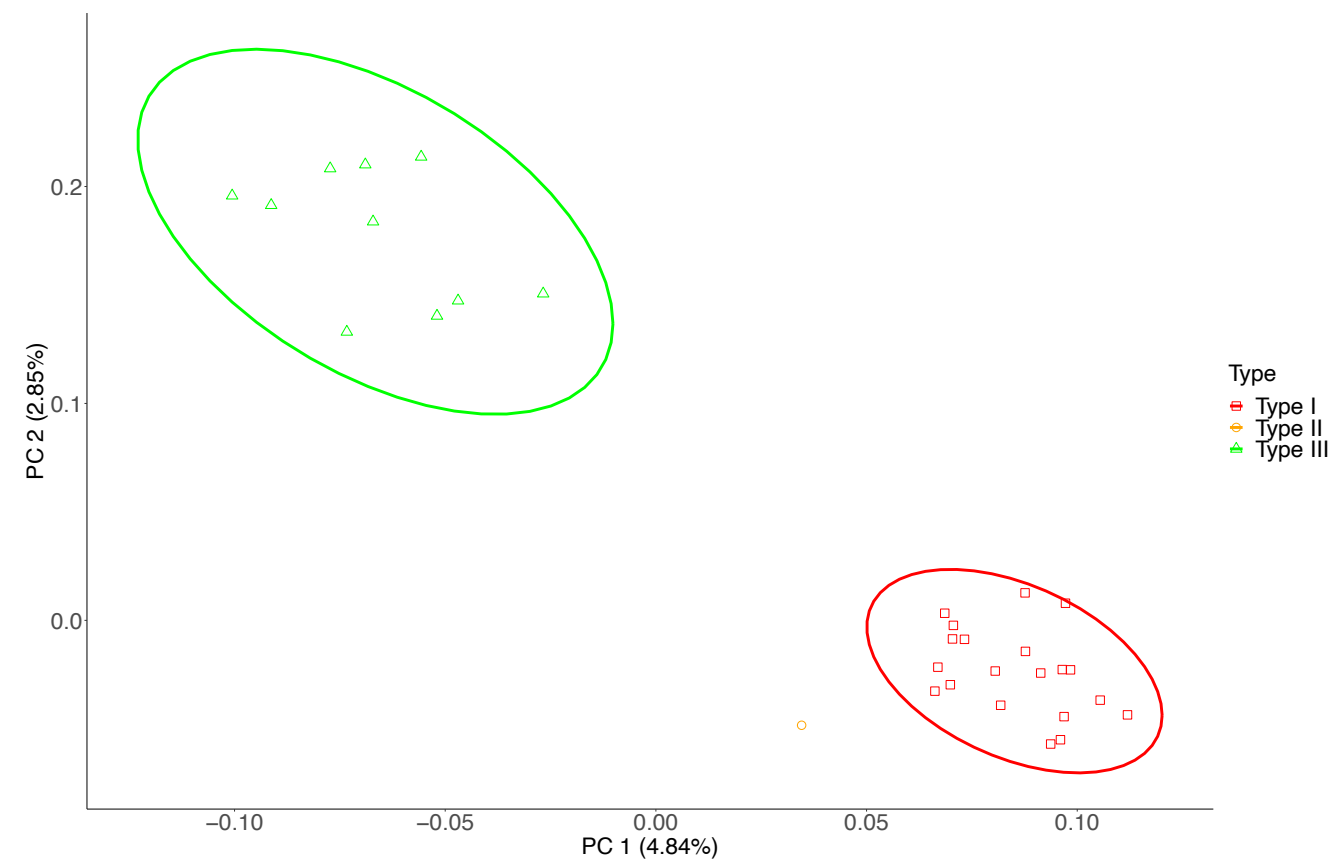


Figure 2 Examining hierarchical clustering and population structure in the Soorni et al., 2017 (n=94) and the Medicinal Genomics StrainSEEK V1 (n=289) datasets. In each case clustering was conducted based on nuclear genetic SNPs while reported use-type within the dataset is below in solid bars to facilitate interpretation based upon community standards **(A)** Hierarchical cluster dendrogram from 6,865 nuclear SNPs for the Soorni et al., 2017 dataset with use-type of each accession indicated below. Use-type are pictured below (Type I=20, Type III=10, Type II=1, Landrace=78 and Unknown=63) **(B)** Visualization of population structure and admixture from 33,629 nuclear SNPs for the Soorni et al., 2017 dataset using the fastSTRUCTURE software (k=2-5) with the optimal number of K being 3 using the silhouette method **(Fig. S9-10)** **(C)** Hierarchical cluster dendrogram from 5,045 nuclear SNPs for the Medicinal Genomics StrainSEEK V1 dataset with use-type indicated below. Use-type of accessions include Type I=108, Type III=9, Type II=17 and Unknown=155 **(D)** Visualization of population structure and admixture from 20,566 nuclear SNPs for the Medicinal Genomics StrainSEEK V1 dataset using the fastSTRUCTURE software (k=2-5) with the optimal number of K being 3 using the silhouette method **(Fig. S9-10).**

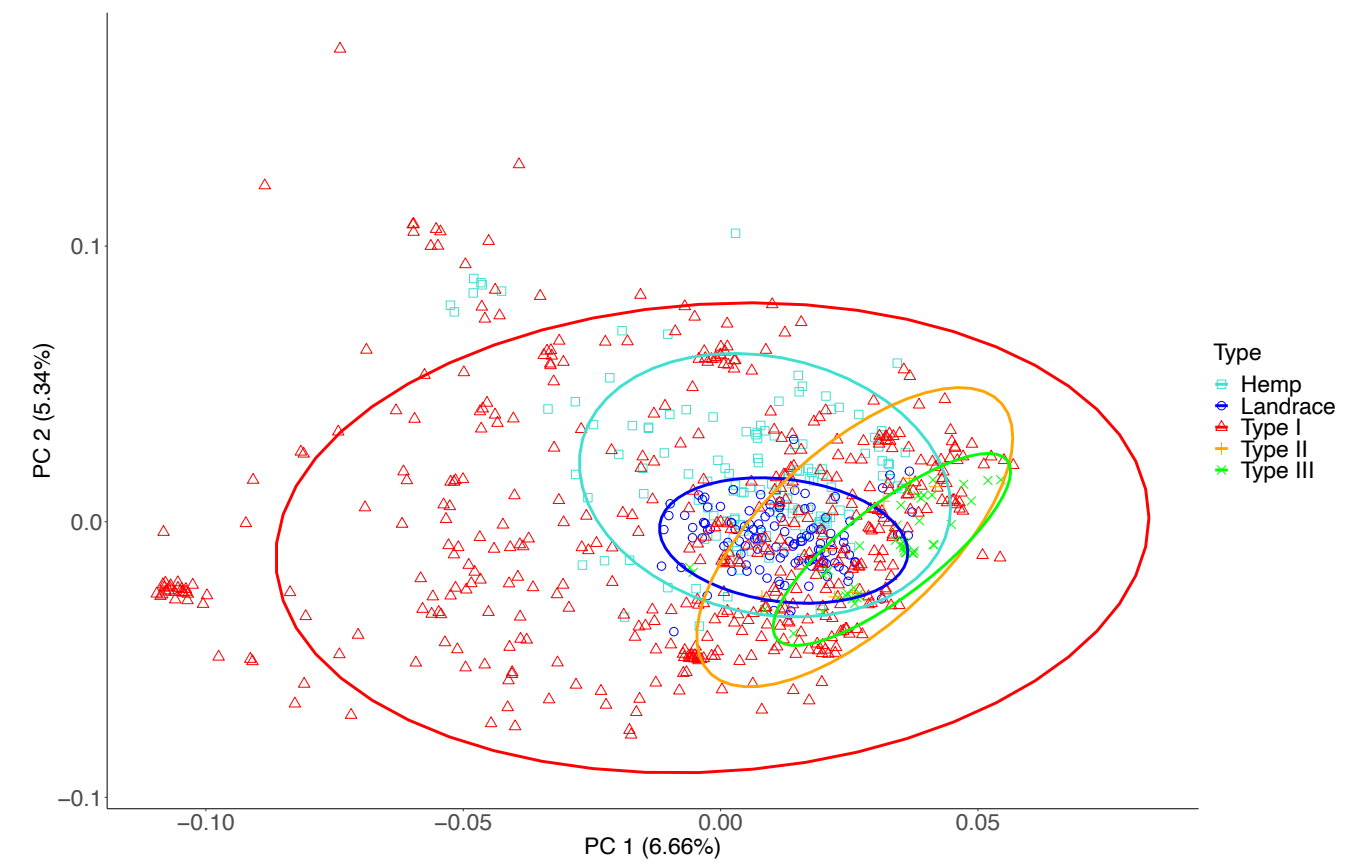
A



B



C



D

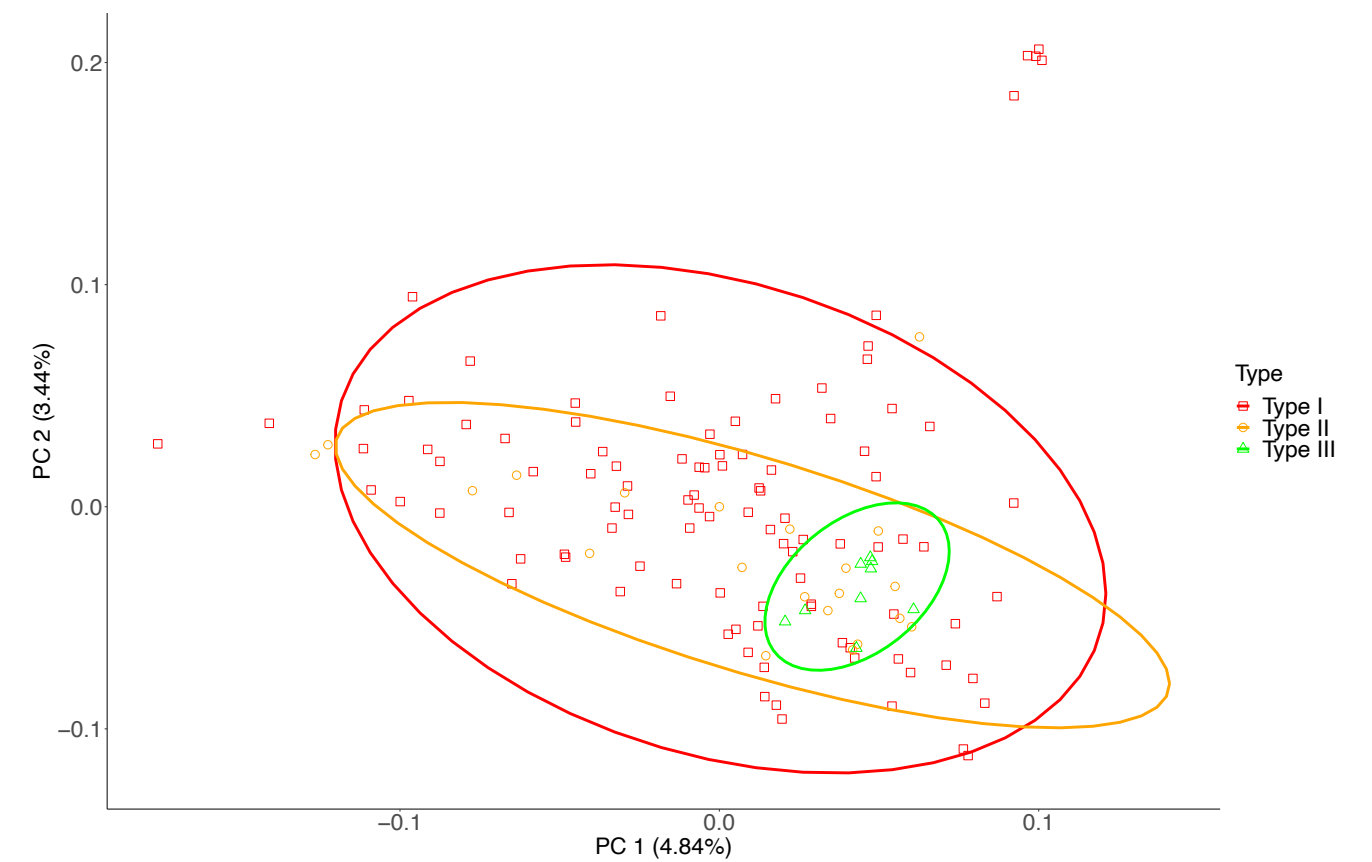
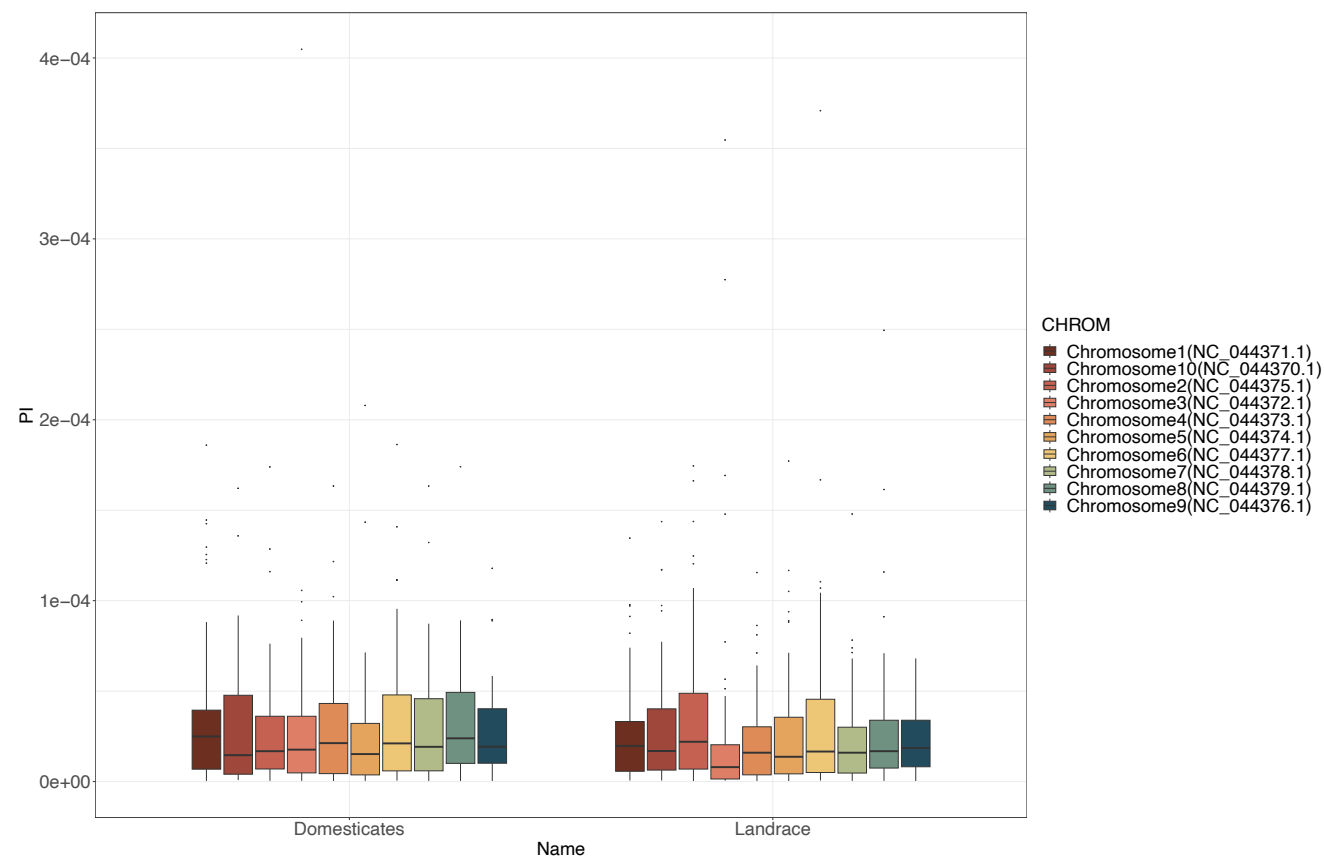
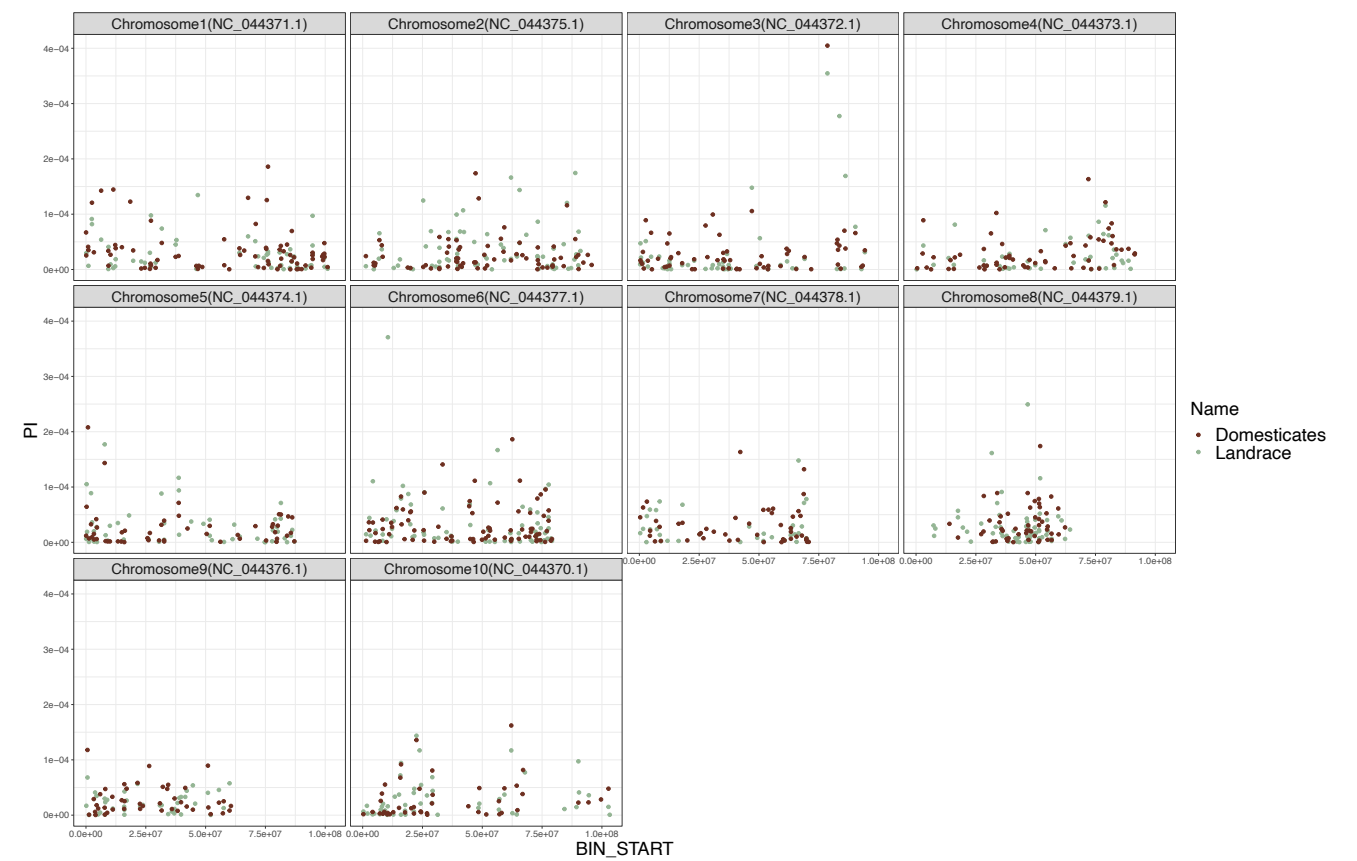


Figure 3 Examination of use-type association across datasets **(A)** Principal component analysis (PCA) from 520 nuclear SNPs for the LeafWorks Inc. dataset **(B)** PCA from 213 SNPs Phylos Biosciences(n=845) dataset **(C)** PCA from 6,865 nuclear SNPs for the Soorni et al., 2017 dataset where cannabinoid content could be determined due to recent publication for 31/94 samples. **(D)** PCA from 5,045 nuclear SNPs for the Medicinal Genomics StrainSEEK V1 dataset.

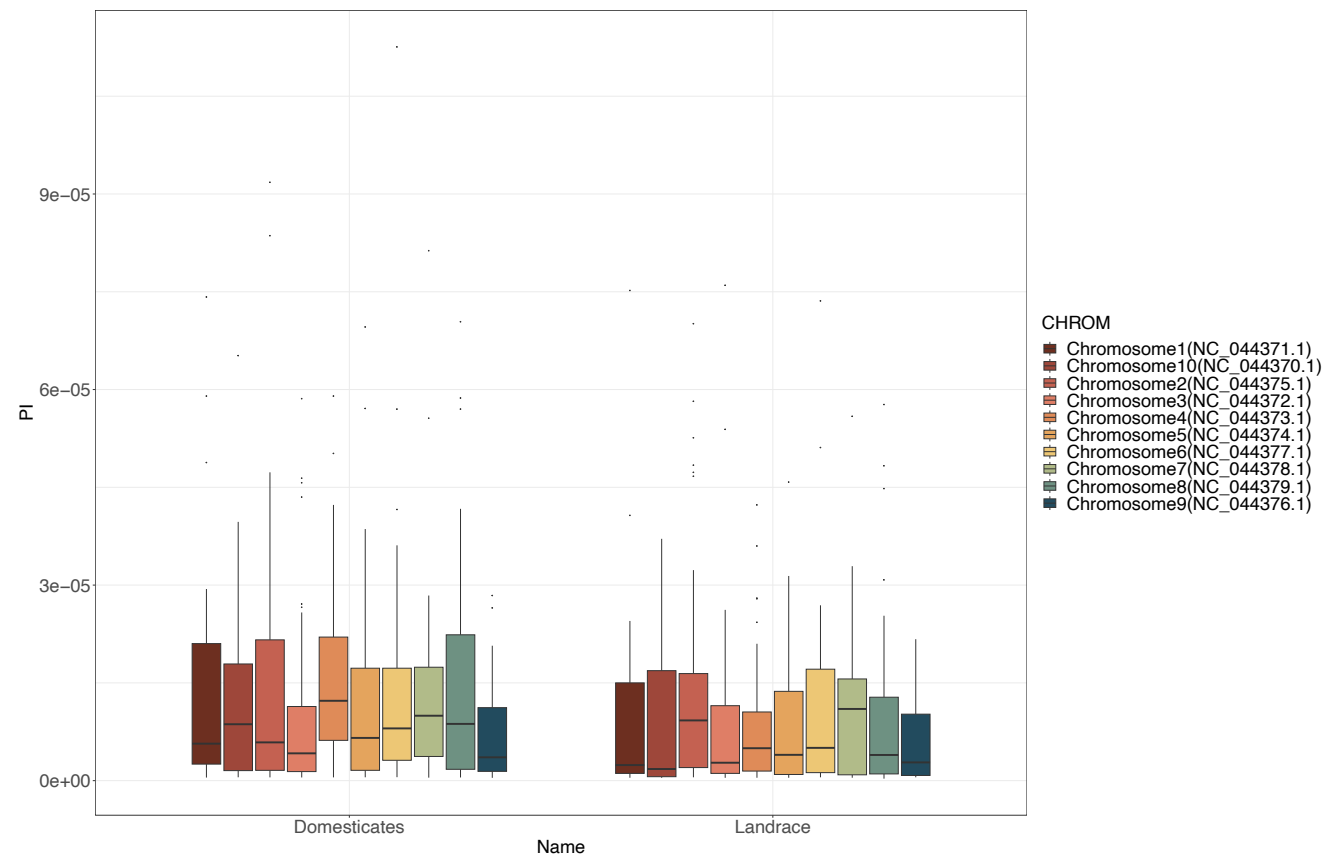
A



B



C



D

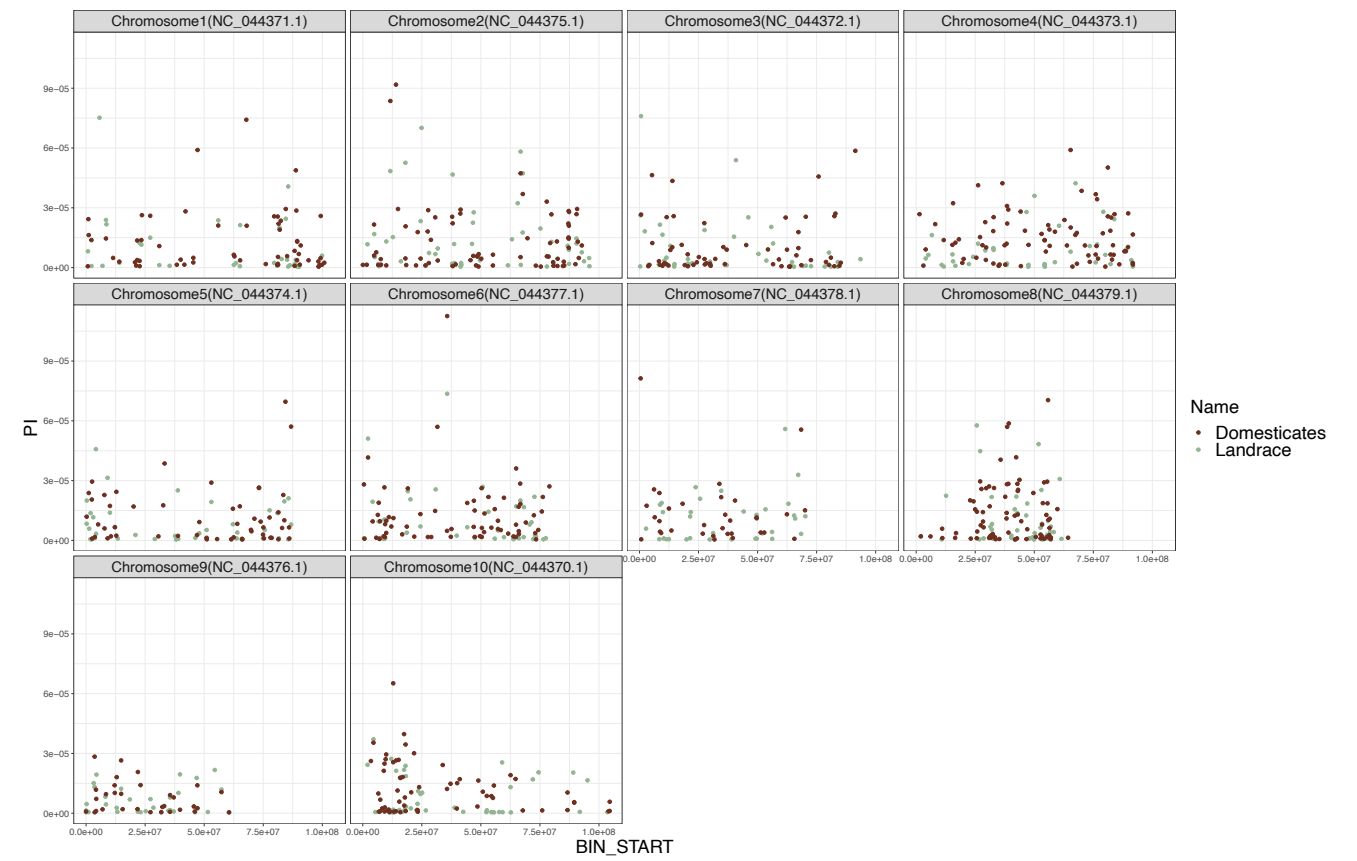
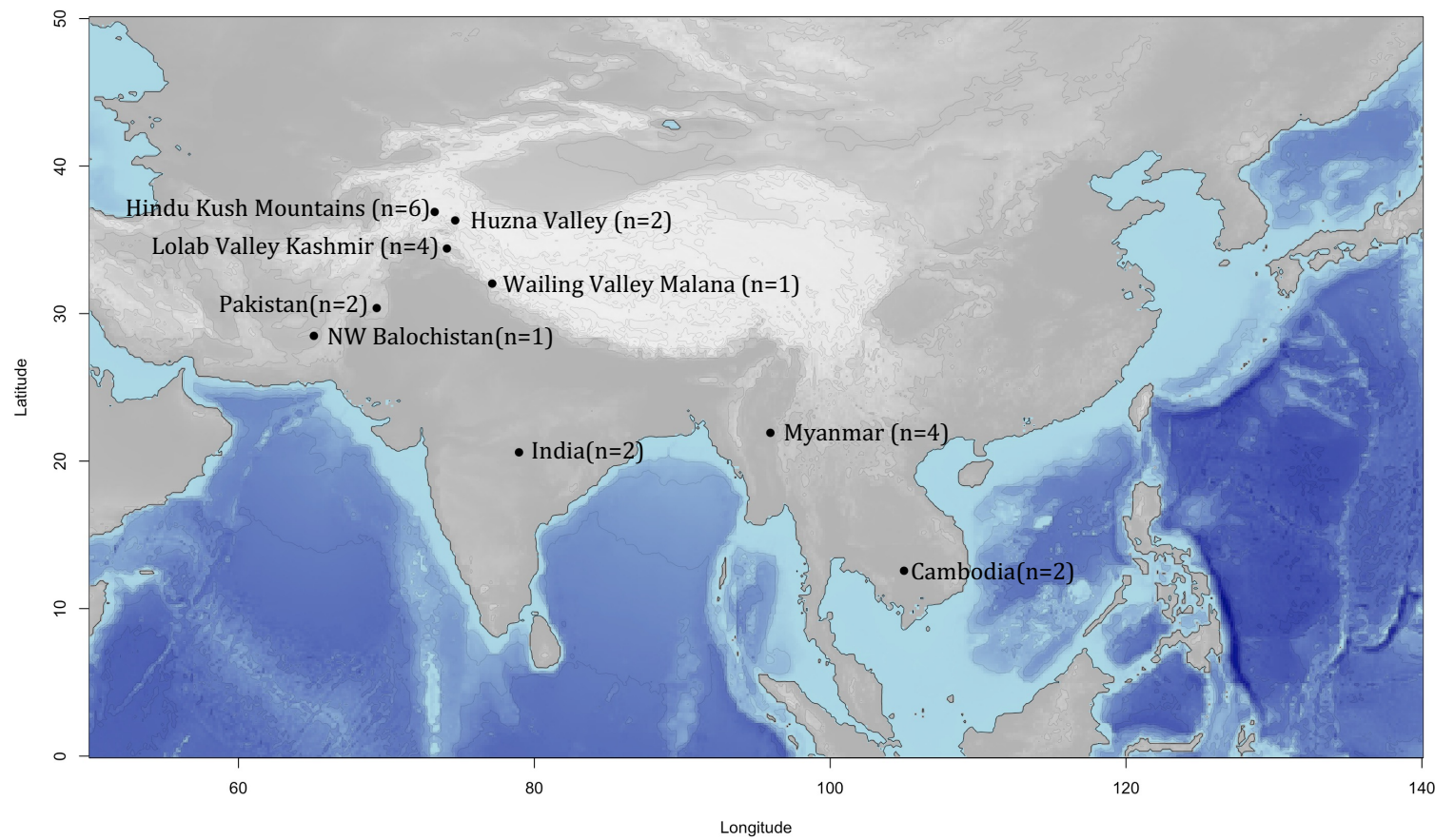
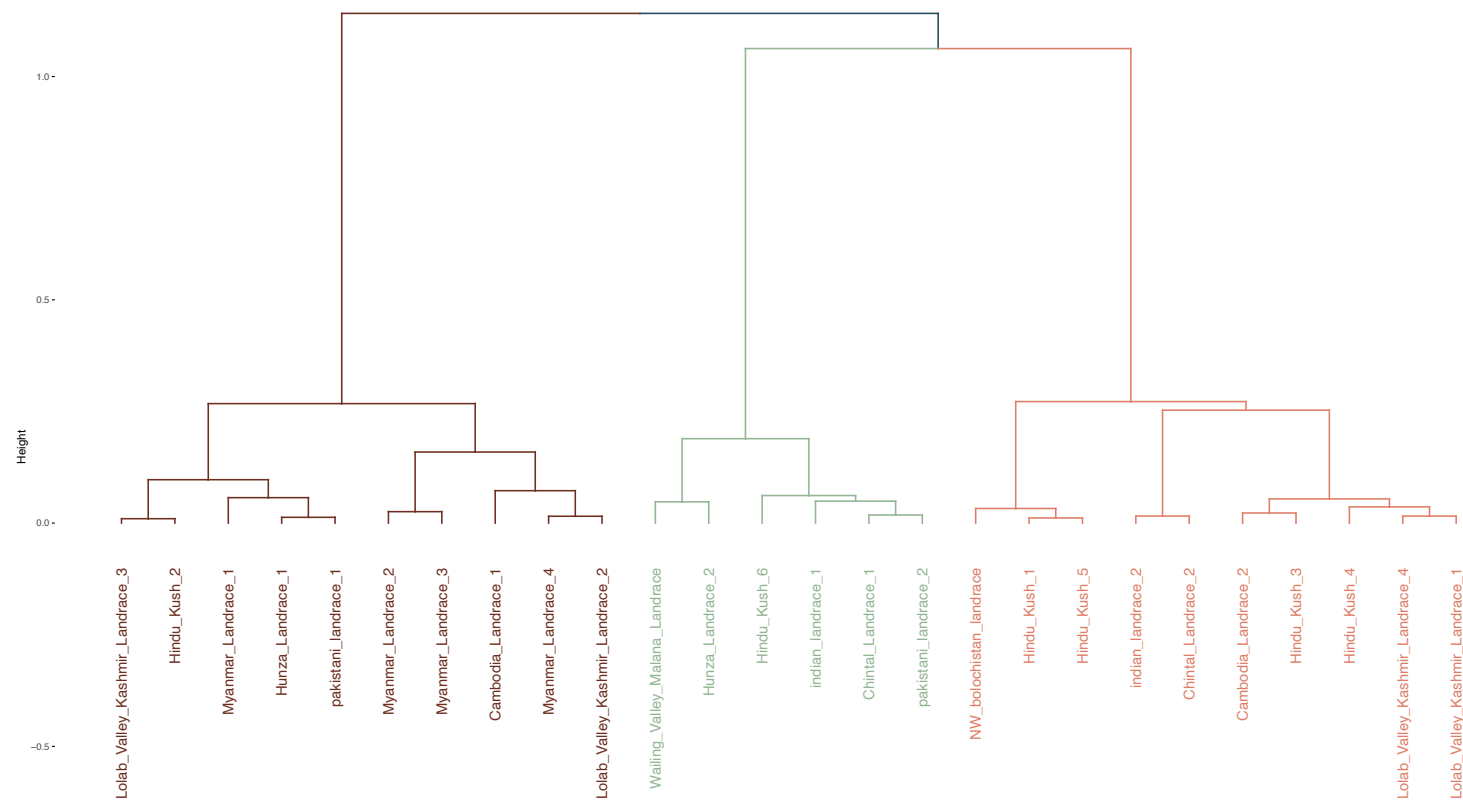


Figure 4 Nucleotide diversity as examined by a 10kb sliding window for landrace and domesticated partitions for the LeafWorks Inc. and Phylos Biosciences datasets **(A)** Nucleotide diversity by chromosome and **(B)** across chromosome length for Domesticated (n=397, 2,096 SNPs) and Landrace (n=101, 2,131 SNPs) samples for the LeafWorks Inc. dataset **(C)** Nucleotide diversity by chromosome and **(D)** across chromosome length for Domesticated (n=718, 749 SNPs) and Landrace (n=127, 566 SNPs) samples for the Phylos Biosciences dataset.

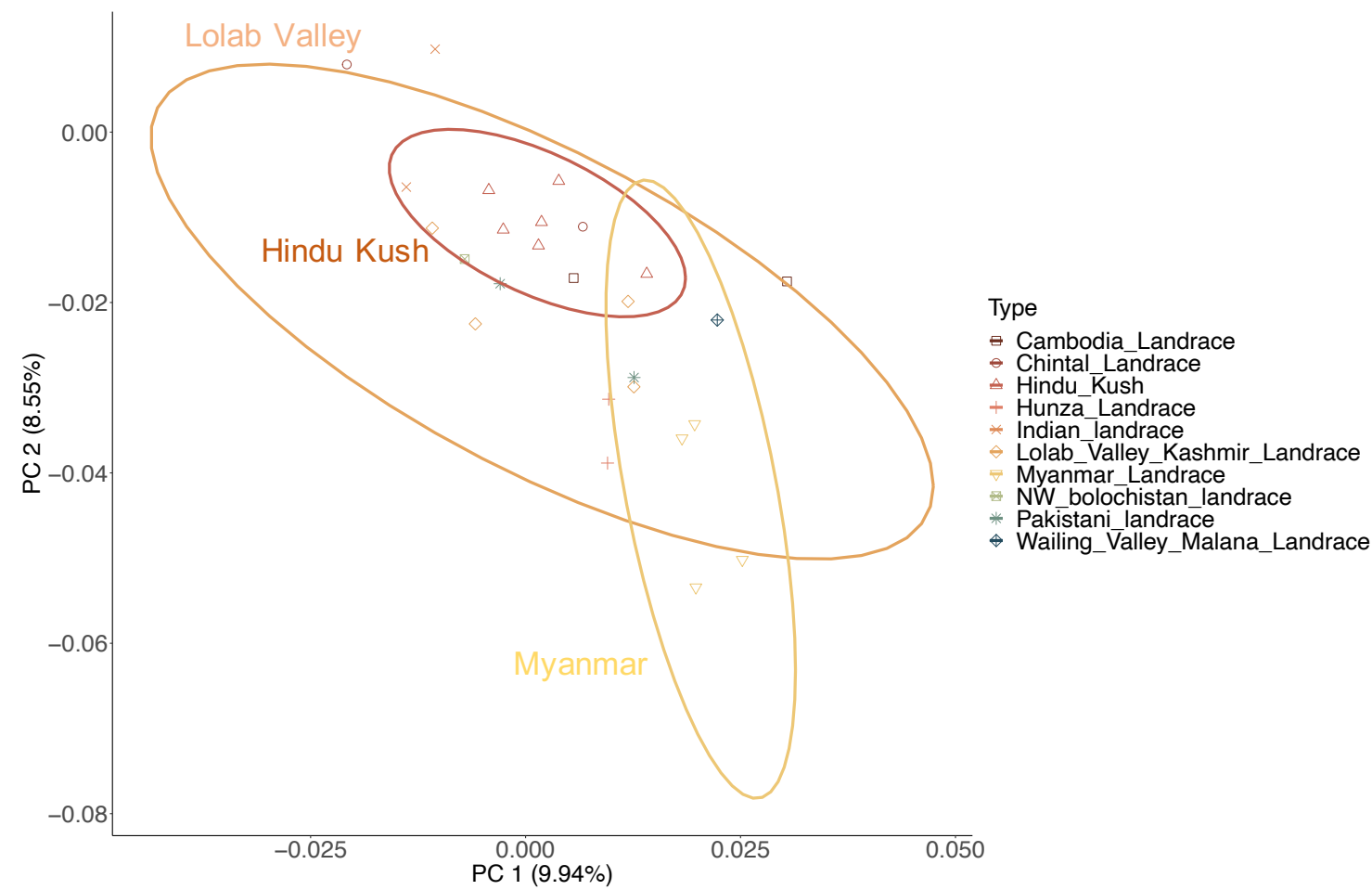
A



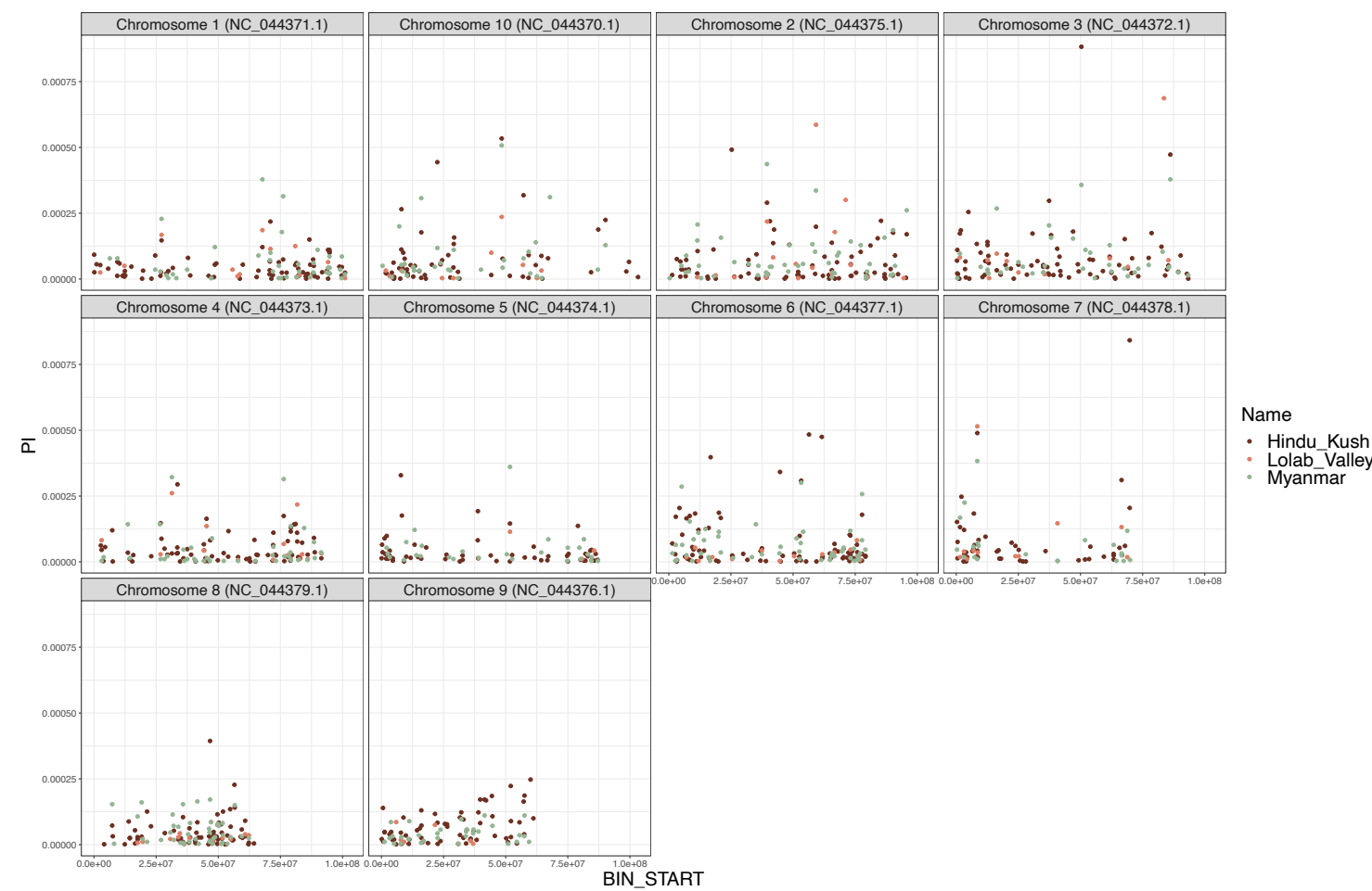
B



C



D



E

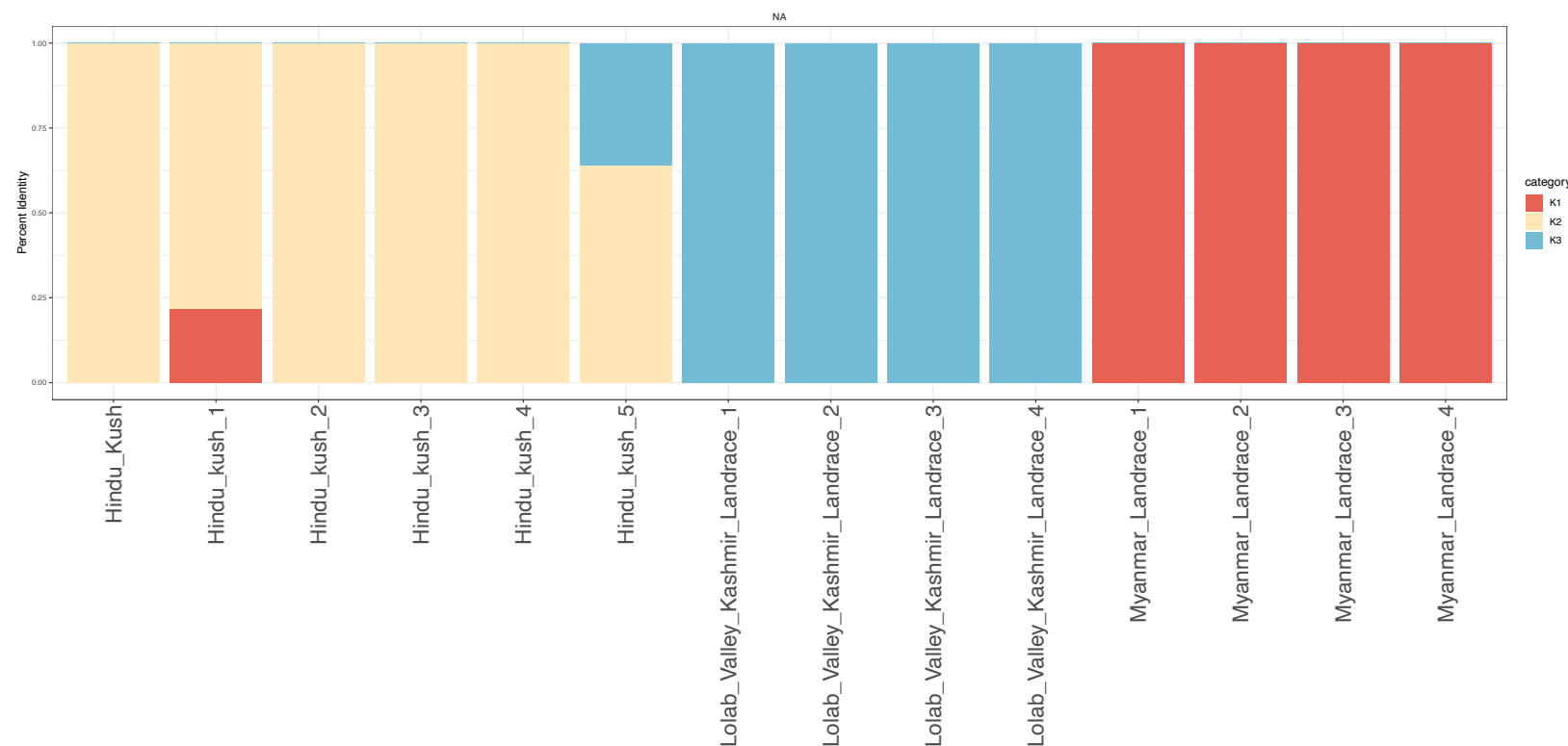


Figure 5 Landrace accessions from the LeafWorks Inc. dataset show separation between Indian and Myanmar populations **(A)** Map detailing the locations of landrace accessions, highlighted are the Hindu Kush Mountains, Lolab Valley and Myanmar **(B)** Hierarchical cluster dendrogram based on 304 SNPs (LD 0.2) across 26 samples of known and trusted origin **(C)** PCA based on 304 SNPs with geographical locations of samples as indicated **(D)** Nucleotide diversity comparison between Hindu Kush Mountains (n=6, 4,304 SNPs), Lolab Valley (n=4, 853 SNPs) and Myanmar (n=4, 2,204 SNPs) as examined by a 10kb sliding window **(E)** Visualization of population structure and admixture using the fastSTRUCTURE software (k=3) with the optimal number of K being 3 using the silhouette method.