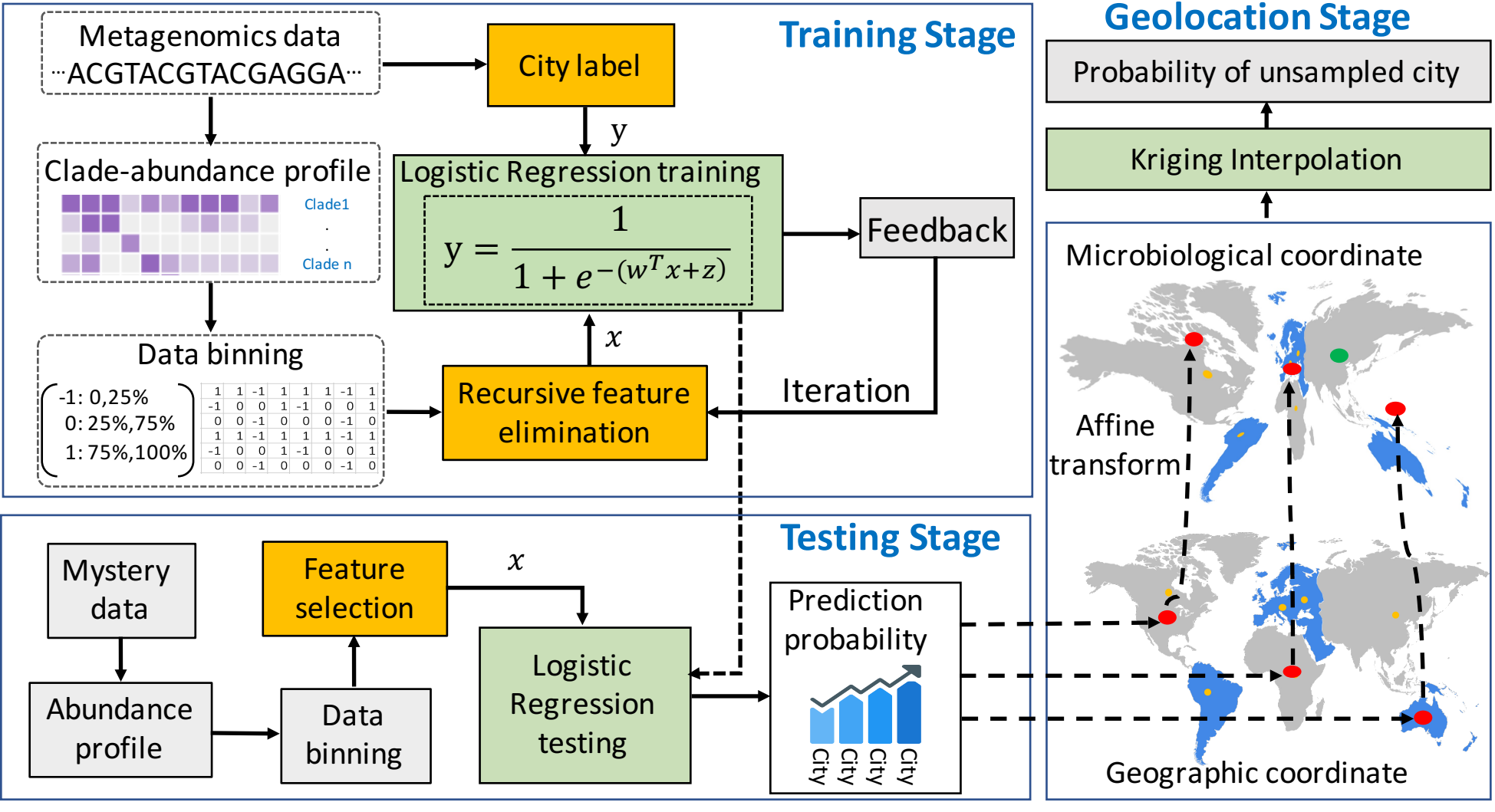
Training Stage
Metagenomics data
···ACGTACGTACGAGGA···
City label
y
Clade-abundance profile
Clade1
Clade n
Logistic Regression training
$y = \frac{1}{1 + e^{-(w^T x + z)}}$
Feedback
Data binning
-1: 0,25%
0: 25%,75%
1: 75%,100%
1 1 -1 1 1 1 -1 1
-1 0 0 1 -1 0 0 1
0 0 -1 0 0 0 -1 0
1 1 -1 1 1 1 -1 1
-1 0 0 1 -1 0 0 1
0 0 -1 0 0 0 -1 0
x
Recursive feature elimination
Iteration
Testing Stage
Mystery data
Feature selection
x
Abundance profile
Data binning
Logistic Regression testing
Prediction probability
City
City
City
City
Geolocation Stage
Probability of unsampled city
Kriging Interpolation
Microbiological coordinate
Affine transform
Geographic coordinate