Kmer\_length: Length of kmers that reads are split into

FPR: false positive rate used to construct the bloom filter (affects bloom filter size + hash count based on equations for optimal m, k)

Query Threshold: threshold for determining read belongs to genome while querying sequence bloom tree

Accuracy: Fraction of time one of the genome matches is the genome where the read came from

AvgMatches: Average number of matching genomes per read

Bloom Filter Size: Number of bits in a single bloom filter bit vector

Bloom Tree Size: size of bloom tree object (just root node) in bytes (using sys.getsizeof() function)

* 2^n = whole size of bloom tree in bytes

Construction Time: time to construct bloom tree in seconds (using time.time())

Query Time: time to query 500 contaminant reads in seconds (using time.time())

Will create confusion matrix w/ optimal parameters

Parameter Optimization (on Contamination Data that Tree was Built with)

**FPR = [0.1, 0.3, 0.5]**

Kmer\_length: 30

**FPR: 0.1**

Query Threshold: 0.3

Accuracy: 1.0

AvgMatches: 1.048

avgMouseMatches: 0.0

Bloom Filter size: 120035574

Hash Count: 3

Bloom Tree size: 56

Construction Time: 623.1445491313934

Query Time: 263.23813796043396

Kmer\_length: 30

**FPR: 0.3**

Query Threshold: 0.3

Accuracy: 1.0

AvgMatches: 1.05

avgMouseMatches: 0.0

Bloom Filter size: 62764050

Hash Count: 1

Bloom Tree size: 56

Construciton Time: 324.816055059433

Query Time: 155.55537009239197

Kmer\_length: 30

**FPR: 0.5**

Query Threshold: 0.3

Accuracy: 1.0

AvgMatches: 5.0

avgMouseMatches: 5.0

Bloom Filter size: 36134308

Hash Count: 0

Bloom Tree size: 56

Construction Time: 225.54240322113037

Query Time: 100.3700201511383

**Query Threshold = [0.1, 0.3, 0.5]**

Kmer\_length: 30

FPR: 0.1

**Query Threshold: 0.1**

Accuracy: 1.0

AvgMatches: 1.056

avgMouseMatches: 0.0

Bloom Filter size: 120035574

Hash Count: 3

Bloom Tree size: 56

Construction Time: 604.3383688926697

Query Time: 243.45062613487244

Kmer\_length: 30

FPR: 0.3

**Query Threshold: 0.3**

Accuracy: 1.0

AvgMatches: 1.05

avgMouseMatches: 0.0

Bloom Filter size: 62764050

Hash Count: 1

Bloom Tree size: 56

Construction Time: 324.816055059433

Query Time: 155.55537009239197

Kmer\_length: 30

FPR: 0.1

**Query Threshold: 0.5**

Accuracy: 1.0

AvgMatches: 1.048

avgMouseMatches: 0.0

Bloom Filter size: 120035574

Hash Count: 3

Bloom Tree size: 56

Construction Time: 950.2536268234253

Query Time: 246.81621313095093

Performance Testing (on Testing Data)

Kmer\_length: 20

FPR: 0.1

Query Threshold: 0.3

Accuracy: 0.0

AvgMatches: 0.328

avgMouseMatches: 0.0

Bloom Filter size: 120044200

Hash Count: 3

Bloom Tree size: 56

Construciton Time: 593.7528820037842

Query Time: 22.30403995513916

Additional Random Tests

Kmer\_size = 30

1 hash function

m = 1000

kmer\_size = 30

threshold = 0.3

Accuracy: 1.0

AvgMatches: 5.0

Kmer\_size = 5

Fpr = 0.1

Threshold = 0.3

Accuracy: 1.0

AvgMatches: 5.0

avgMouseMatches: 5.0

BF size: 120057140

Hash Count: 3

Bloom Tree size: 56