- **Setup**

In [1]:
```
import sys, os
current_directory = os.getcwd()
root_directory = os.path.abspath(os.path.join(current_directory, os.pardir))
sys.path.append(root_directory)
```

- **Import utils**

In [9]:
```
from pretrained.predictor import SingleKModel, MultiKModel, OneTestKModel
import metrics
```

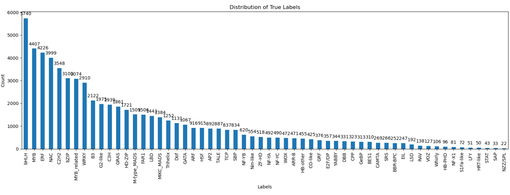
- **Predict**

1- Using single K Model

In [3]:
```
kmodel = SingleKModel(kmer_size=3)
kmodel.load("../data/testset-full/k3/testset.csv", format="csv", type='kmer_file')
genboard = kmodel.predict()
```

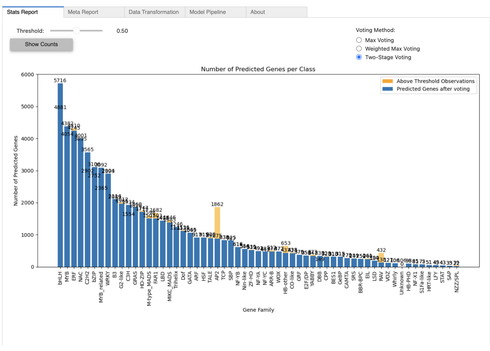
```
Predicting: 100%|██████████| 58/58 [05:27<00:00,  5.64s/it]
```

In [4]:
```
metrics.plot_testset(true_label_df_path="../data/testset-full/k3/true_labels.csv", class_mapping_df
```



In [5]:
```
genboard.display()
```

```
Tab(children=(VBox(children=(HBox(children=(VBox(children=(FloatSlider(value=0.5, continuous_update
=False, des…
```



In [6]:
```
genboard.prediction.shape
```

Out[6]:
```
(64091, 58)
```

- **Confusion report for validation**

In [7]:
```
import json
import pandas as pd

true_label = pd.read_csv('../data/testset-full/k3/true_labels.csv')['true_label'].values
with open('../data/testset-full/k3/class_mapping.json', 'r') as json_file:
    class_mapping = json.load(json_file)
class_mapping['Unknown'] = 0
```

In [8]:
```
genboard.show_eval_metric(
    true_label=true_label,
    class_mapping_rules=class_mapping,
    voting_method="Two-Stage Voting",
    voting_threshold=0.5,
    binary_class_threshold=0.5,
    components=['confusion_matrix', 'general_accuracy', 'accuracy_per_family']
)
```

Overall Accuracy

kmer_size=3

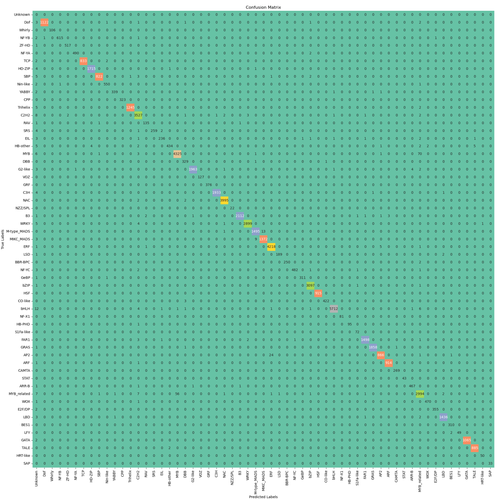
Score	0.99

Accuracy per Gene Family

Gene Family	Accuracy	Precision	Recall	F1 Score
AP2	**0.98**	0.47	0.98	0.64
ARR-B	**1.00**	0.90	0.97	0.93
BBR-BPC	**1.00**	1.00	0.99	1.00
C2H2	**0.99**	0.98	0.80	0.88
CAMTA	**1.00**	0.98	0.99	0.99
CPP	**1.00**	0.98	0.99	0.99
Dof	**1.00**	0.99	0.99	0.99
EIL	**1.00**	0.67	0.71	0.69
FAR1	**1.00**	0.88	0.98	0.93
GATA	**1.00**	1.00	0.98	0.99
GRF	**1.00**	1.00	0.99	1.00
HB-PHD	**1.00**	0.97	0.99	0.98
HD-ZIP	**1.00**	0.98	0.99	0.99
HSF	**1.00**	1.00	1.00	1.00
LFY	**1.00**	0.98	0.86	0.92
M-type_MADS	**0.99**	0.86	0.91	0.88
MYB	**0.99**	0.97	0.89	0.93
NAC	**1.00**	1.00	0.97	0.98
NF-YA	**1.00**	0.99	1.00	0.99
NF-YC	**1.00**	0.99	0.89	0.93
Nin-like	**1.00**	0.96	0.95	0.95
S1Fa-like	**1.00**	0.99	0.99	0.99
SBP	**1.00**	1.00	0.94	0.97
STAT	**1.00**	1.00	1.00	1.00
TCP	**1.00**	0.99	0.99	0.99
VOZ	**1.00**	1.00	1.00	1.00
WRKY	**1.00**	0.99	0.99	0.99
YABBY	**1.00**	0.92	1.00	0.96
bHLH	**0.99**	1.00	0.85	0.92

Gene Family	Accuracy	Precision	Recall	F1 Score
ARF	**1.00**	0.99	1.00	1.00
B3	**1.00**	0.93	0.92	0.93
BES1	**1.00**	1.00	1.00	1.00
C3H	**0.99**	0.99	0.80	0.88
CO-like	**1.00**	0.96	0.99	0.98
DBB	**1.00**	0.95	0.43	0.59
E2F/DP	**1.00**	0.99	0.98	0.99
ERF	**1.00**	0.98	1.00	0.99
G2-like	**1.00**	0.96	0.97	0.96
GRAS	**1.00**	0.99	0.97	0.98
GeBP	**1.00**	0.93	0.94	0.94
HB-other	**0.99**	0.46	0.66	0.55
HRT-like	**1.00**	0.98	1.00	0.99
LBD	**1.00**	0.99	0.98	0.99
LSD	**1.00**	0.98	0.94	0.96
MIKC_MADS	**1.00**	0.94	0.99	0.96
MYB_related	**0.98**	0.90	0.69	0.78
NF-X1	**1.00**	0.95	1.00	0.98
NF-YB	**1.00**	1.00	0.99	0.99
NZZ/SPL	**1.00**	0.56	1.00	0.72
RAV	**1.00**	0.31	0.99	0.48
SAP	**1.00**	0.94	1.00	0.97
SRS	**1.00**	1.00	0.93	0.96
TALE	**1.00**	0.96	1.00	0.98
Trihelix	**1.00**	0.99	0.90	0.94
WOX	**1.00**	0.99	1.00	0.99
Whirly	**1.00**	0.98	1.00	0.99
ZF-HD	**1.00**	0.96	0.99	0.98
bZIP	**0.99**	1.00	0.88	0.94

Confusion Matrix



Save Report

In [ ]:
```
```