- **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 [15]:
```
from pretrained.predictor import SingleKModel, MultiKModel, OneTestKModel, BatchSingleKModel
import metrics
```

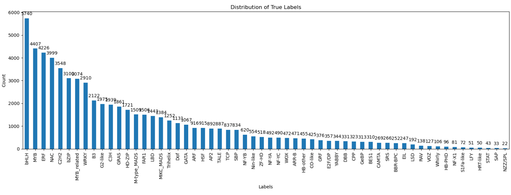
- **Predict**

1- Using single K Model

In [3]:
```
kmodel = BatchSingleKModel(kmer_size=5, batch_size=2000)
kmodel.set_load_config("../data/testset-full/k5/testset.csv", format="csv", type='kmer_file')
genboard = kmodel.predict()
```

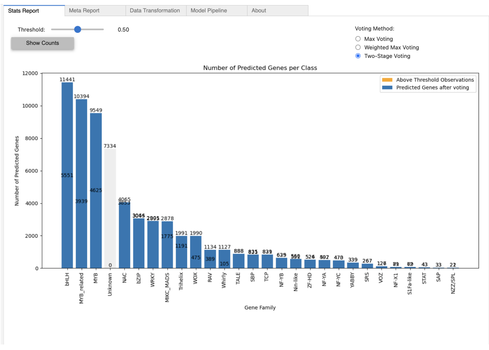
```
Batch Predictions: 33it [39:45, 72.28s/it]
```

In [4]:
```
metrics.plot_testset(true_label_df_path="../data/testset-full/k5/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/k5/true_labels.csv')['true_label'].values
with open('../data/testset-full/k5/class_mapping.json', 'r') as json_file:
    class_mapping = json.load(json_file)
class_mapping['Unknown'] = 0
```

In [16]:
```
genboard.show_eval_metric(
    true_label=true_label,
    class_mapping_rules=class_mapping,
    voting_method="Two-Stage Voting",
    voting_threshold=0.95,
    binary_class_threshold=0.5,
    components=['confusion_matrix', 'general_accuracy', 'accuracy_per_family']
)
```

Overall Accuracy

kmer_size=5

Score	0.50

Accuracy per Gene Family

Gene Family	Accuracy	Precision	Recall	F1 Score
AP2	**0.99**	0.51	0.95	0.67
ARR-B	**0.99**	0.53	0.98	0.68
BBR-BPC	**1.00**	1.00	0.98	0.99
C2H2	**0.99**	1.00	0.84	0.91
CAMTA	**1.00**	1.00	1.00	1.00
CPP	**1.00**	1.00	0.97	0.99
Dof	**1.00**	1.00	1.00	1.00
EIL	**1.00**	1.00	0.98	0.99
FAR1	**1.00**	0.97	0.94	0.95
GATA	**1.00**	1.00	0.98	0.99
GRF	**1.00**	1.00	1.00	1.00
HB-PHD	**1.00**	1.00	1.00	1.00
HD-ZIP	**1.00**	0.95	1.00	0.97
HSF	**1.00**	1.00	1.00	1.00
LFY	**1.00**	1.00	0.94	0.97
M-type_MADS	**0.99**	0.81	0.97	0.88
MYB	**0.99**	0.94	0.98	0.96
NAC	**1.00**	0.99	0.96	0.98
NF-YA	**1.00**	0.98	1.00	0.99
NF-YC	**1.00**	1.00	0.96	0.98
Nin-like	**1.00**	1.00	0.95	0.98
S1Fa-like	**1.00**	1.00	0.96	0.98
SBP	**1.00**	1.00	0.98	0.99
STAT	**1.00**	1.00	1.00	1.00
TCP	**1.00**	1.00	0.99	1.00
VOZ	**1.00**	1.00	0.99	1.00
WRKY	**1.00**	1.00	0.99	1.00
YABBY	**1.00**	1.00	0.98	0.99
bHLH	**1.00**	0.99	0.96	0.98

Gene Family	Accuracy	Precision	Recall	F1 Score
ARF	**1.00**	1.00	1.00	1.00
B3	**0.99**	0.93	0.87	0.90
BES1	**1.00**	1.00	0.99	1.00
C3H	**1.00**	1.00	0.88	0.93
CO-like	**1.00**	0.99	1.00	0.99
DBB	**1.00**	0.53	0.98	0.69
E2F/DP	**1.00**	1.00	0.99	1.00
ERF	**1.00**	0.96	1.00	0.98
G2-like	**0.99**	0.81	0.99	0.89
GRAS	**1.00**	1.00	0.98	0.99
GeBP	**1.00**	0.69	0.96	0.81
HB-other	**0.98**	0.22	0.80	0.35
HRT-like	**1.00**	1.00	1.00	1.00
LBD	**1.00**	0.98	0.98	0.98
LSD	**1.00**	0.91	1.00	0.95
MIKC_MADS	**0.99**	0.78	1.00	0.87
MYB_related	**0.97**	0.68	0.87	0.76
NF-X1	**1.00**	1.00	0.98	0.99
NF-YB	**1.00**	0.97	0.99	0.98
NZZ/SPL	**1.00**	1.00	0.95	0.98
RAV	**1.00**	0.35	1.00	0.52
SAP	**1.00**	0.97	0.97	0.97
SRS	**1.00**	1.00	1.00	1.00
TALE	**1.00**	1.00	1.00	1.00
Trihelix	**1.00**	0.99	0.94	0.97
WOX	**1.00**	0.99	1.00	0.99
Whirly	**1.00**	1.00	0.99	1.00
ZF-HD	**1.00**	0.98	0.99	0.99
bZIP	**1.00**	1.00	0.98	0.99

Confusion Matrix

Save Report

In [ ]:
```
```