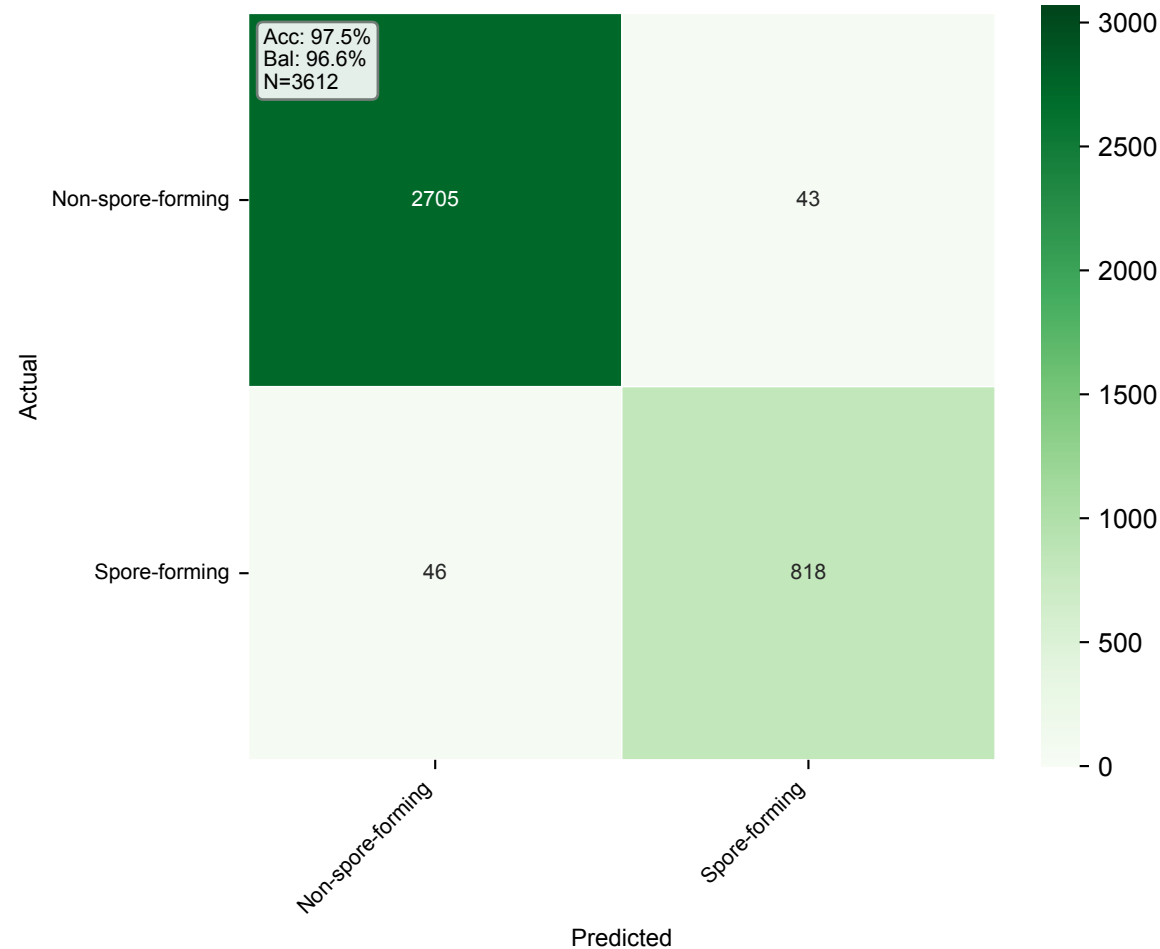
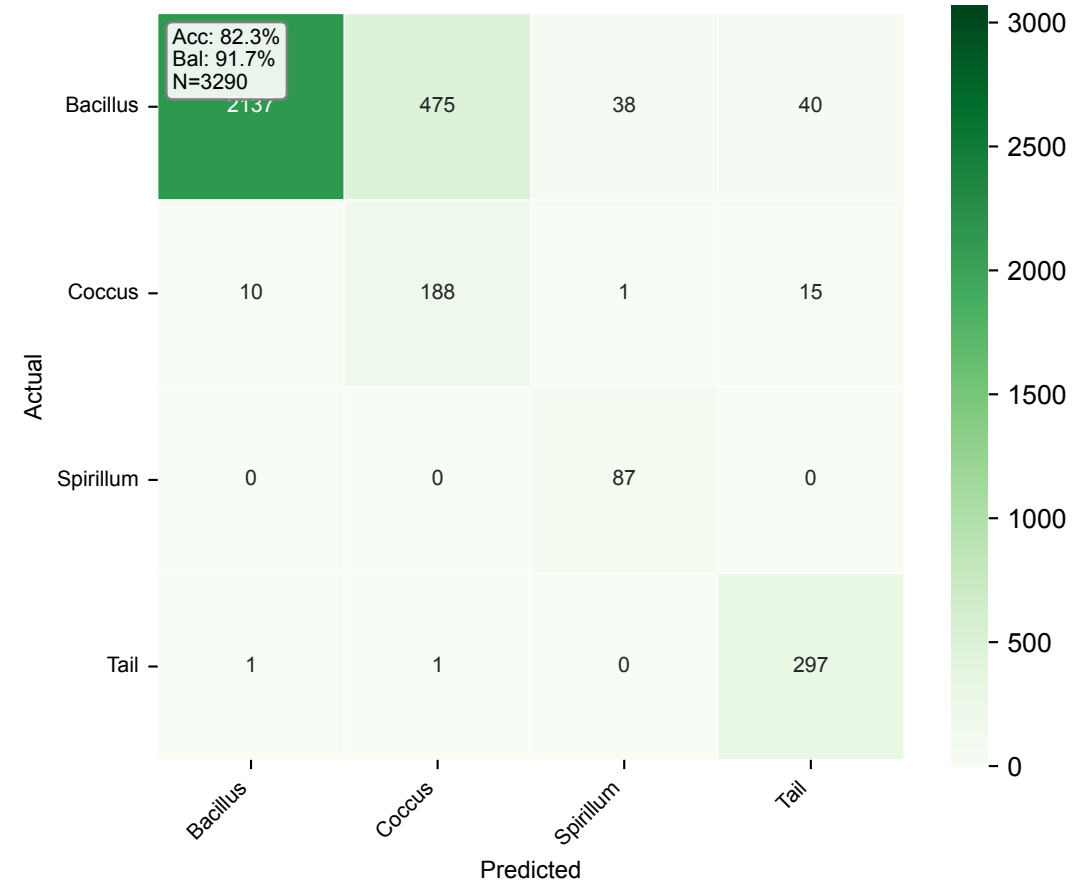


# Best Model Confusion Matrices for Each Phenotype

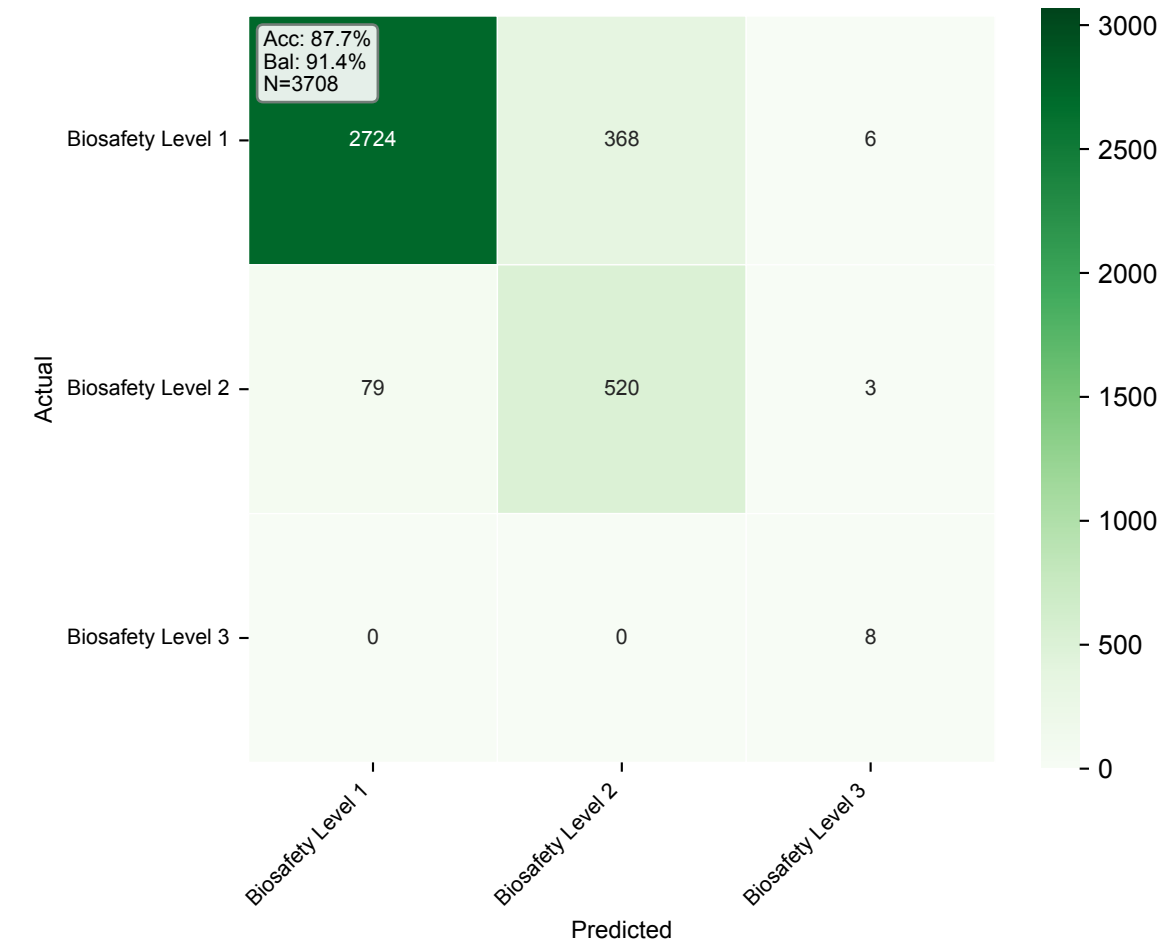
**Spore Formation**  
Model: gemini-2.5-pro



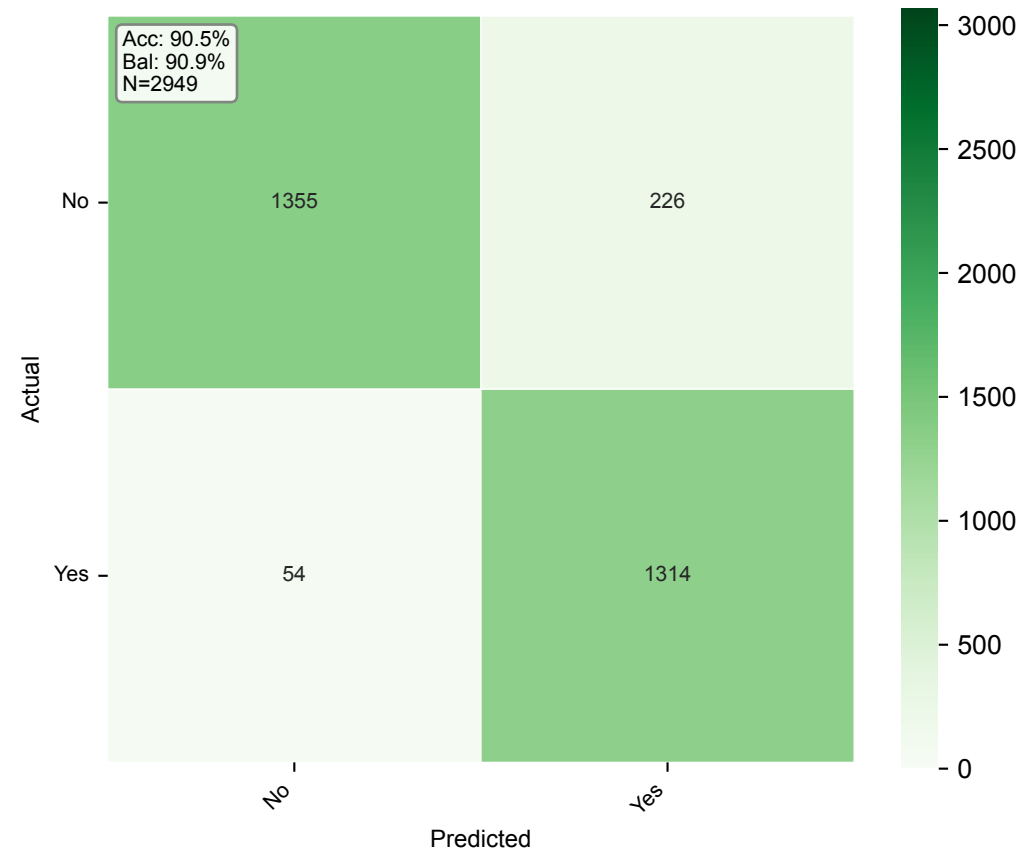
**Cell Shape**  
Model: gpt-4.1-nano



**Biosafety Level**  
Model: claude-3.5-sonnet

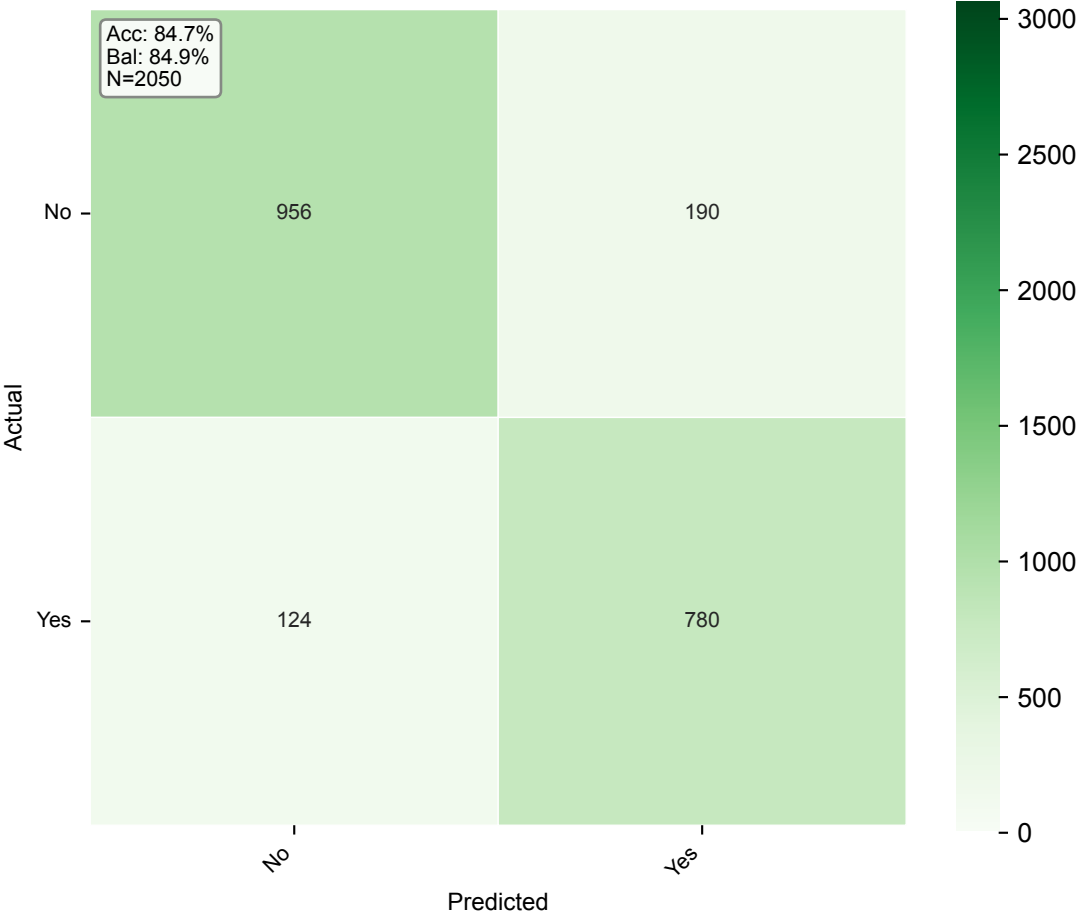


**Motility**  
Model: gpt-5

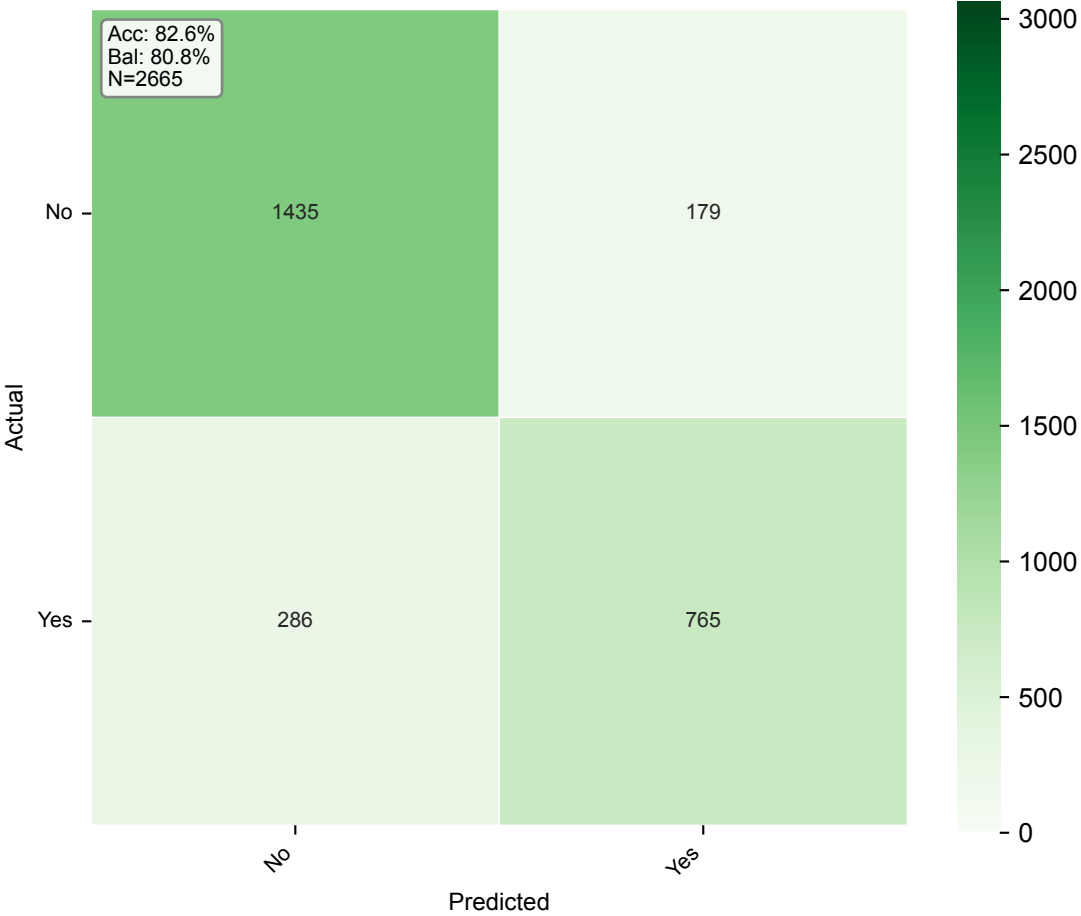


Best Model Confusion Matrices for Each Phenotype

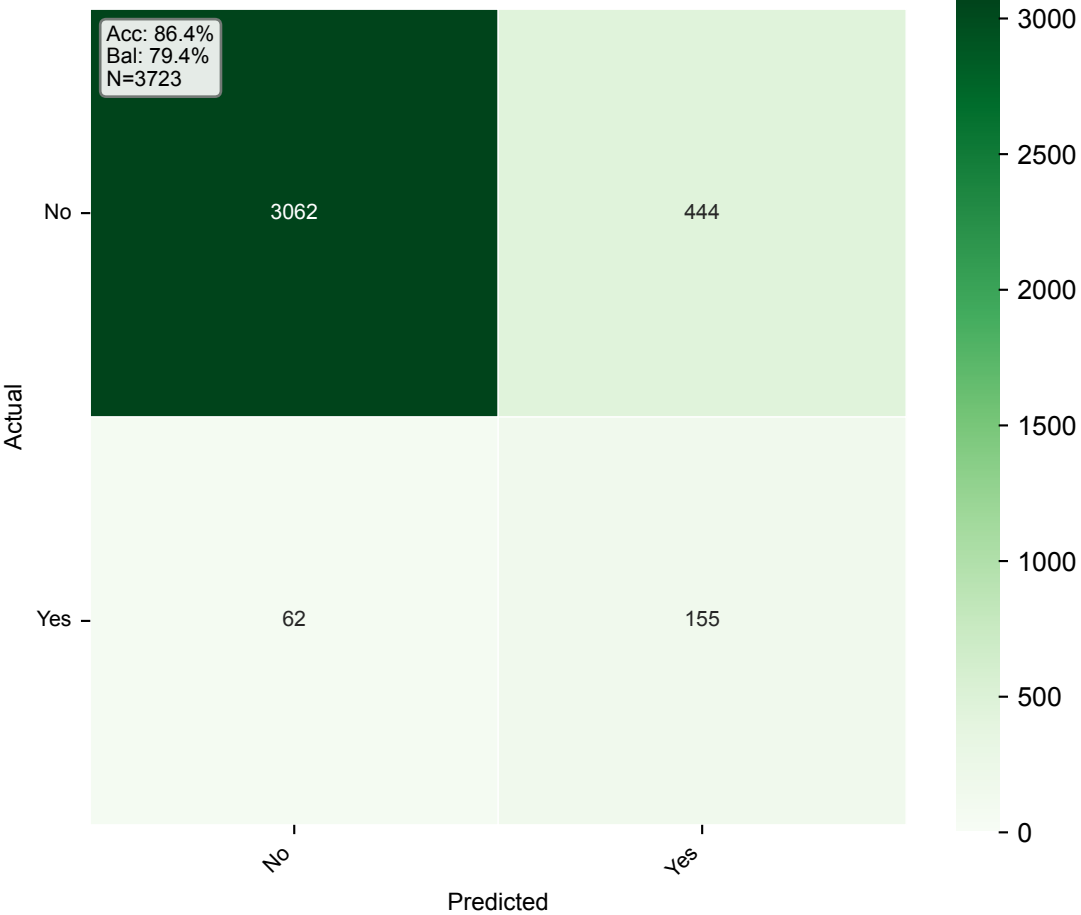
Animal Pathogenicity  
Model: gemini-flash-1.5



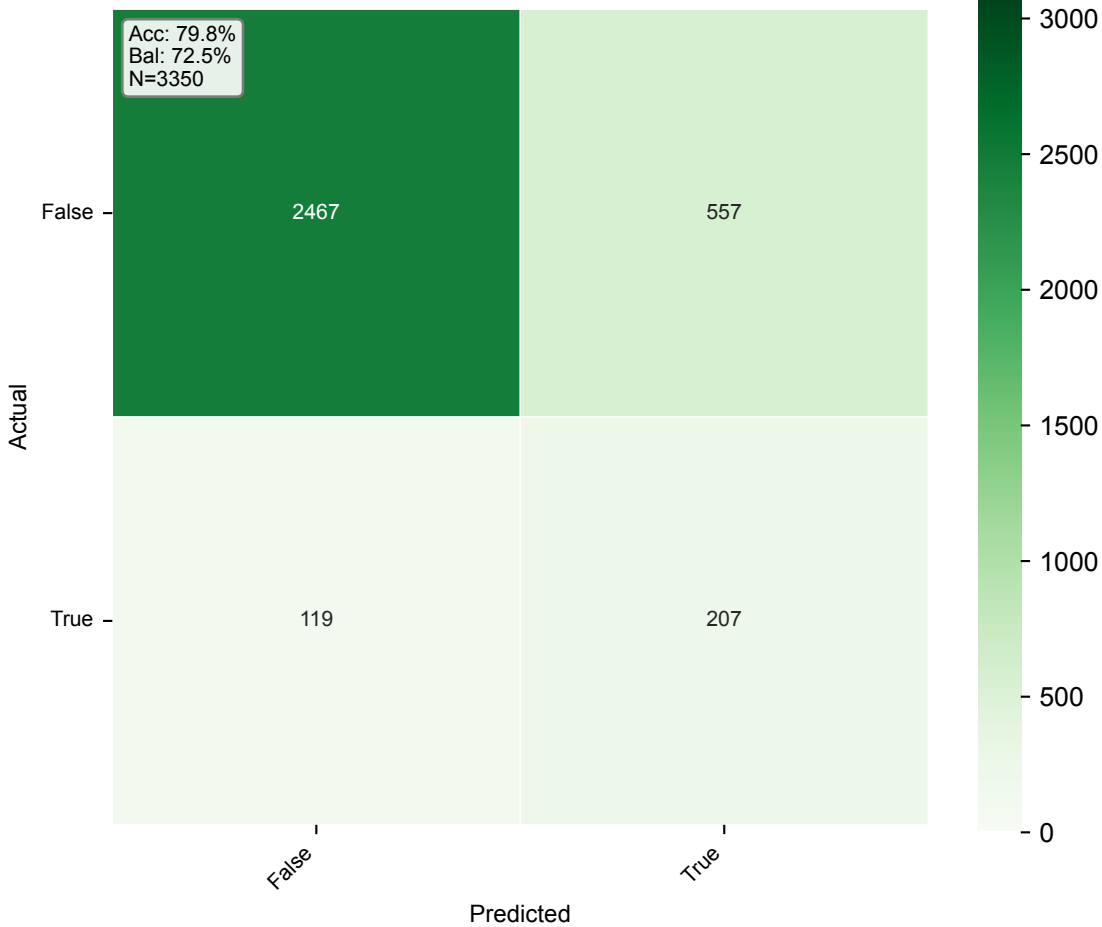
Host Association  
Model: gpt-4o



Plant Pathogenicity  
Model: gemini-pro-1.5

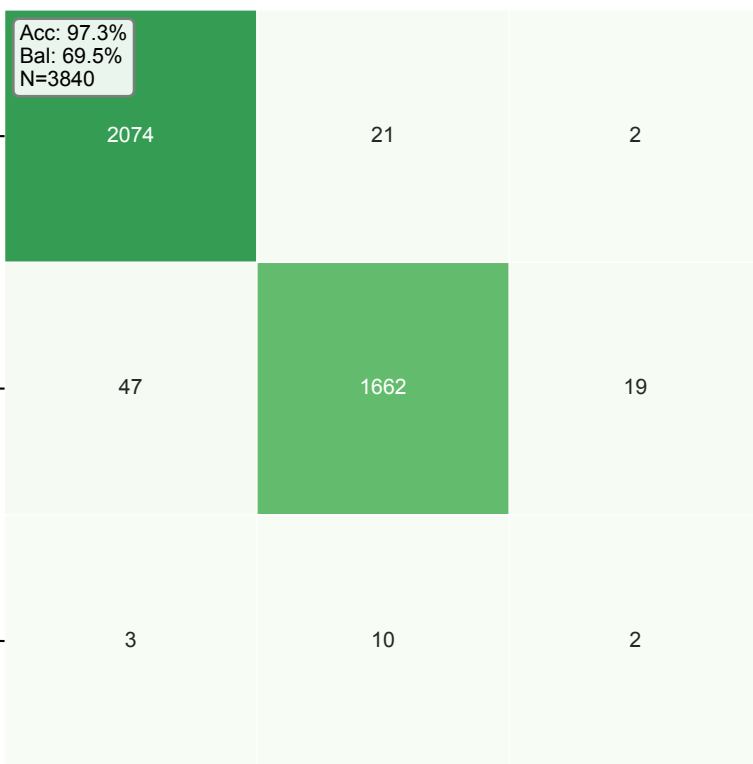


Extreme Environment Tolerance  
Model: deepseek-r1



# Best Model Confusion Matrices for Each Phenotype

**Gram Staining**  
**Model: gemini-2.5-pro**



Acc: 97.3%  
Bal: 69.5%  
N=3840

Gram Stain Negative

2074

21

2

Gram Stain Positive

47

1662

19

Gram Stain Variable

3

10

2

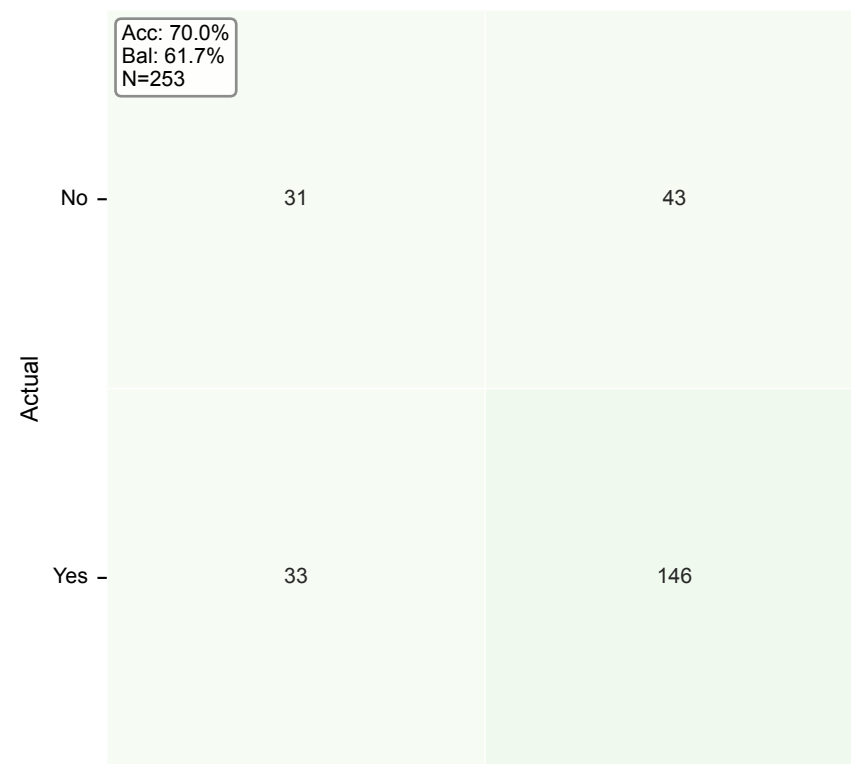
Gram Stain Negative

Gram Stain Positive

Gram Stain Variable

Predicted

**Biofilm Formation**  
**Model: gpt-4**



Acc: 70.0%  
Bal: 61.7%  
N=253

No

31

43

Yes

33

146

No

Yes

Predicted

Summary: Best Model Performance by Phenotype

Phenotype	Best Model	Accuracy	Balanced Accuracy	Sample Size
Spore Formation	gemini-2.5-pro	97.5%	96.6%	3,612
Cell Shape	gpt-4.1-nano	82.3%	91.7%	3,290
Biosafety Level	claude-3.5-sonnet	87.7%	91.4%	3,708
Motility	gpt-5	90.5%	90.9%	2,949
Animal Pathogenicity	gemini-flash-1.5	84.7%	84.9%	2,050
Host Association	gpt-4o	82.6%	80.8%	2,665
Plant Pathogenicity	gemini-pro-1.5	86.4%	79.4%	3,723
Extreme Environment Tolerance	deepseek-r1	79.8%	72.5%	3,350
Gram Staining	gemini-2.5-pro	97.3%	69.5%	3,840
Biofilm Formation	gpt-4	70.0%	61.7%	253