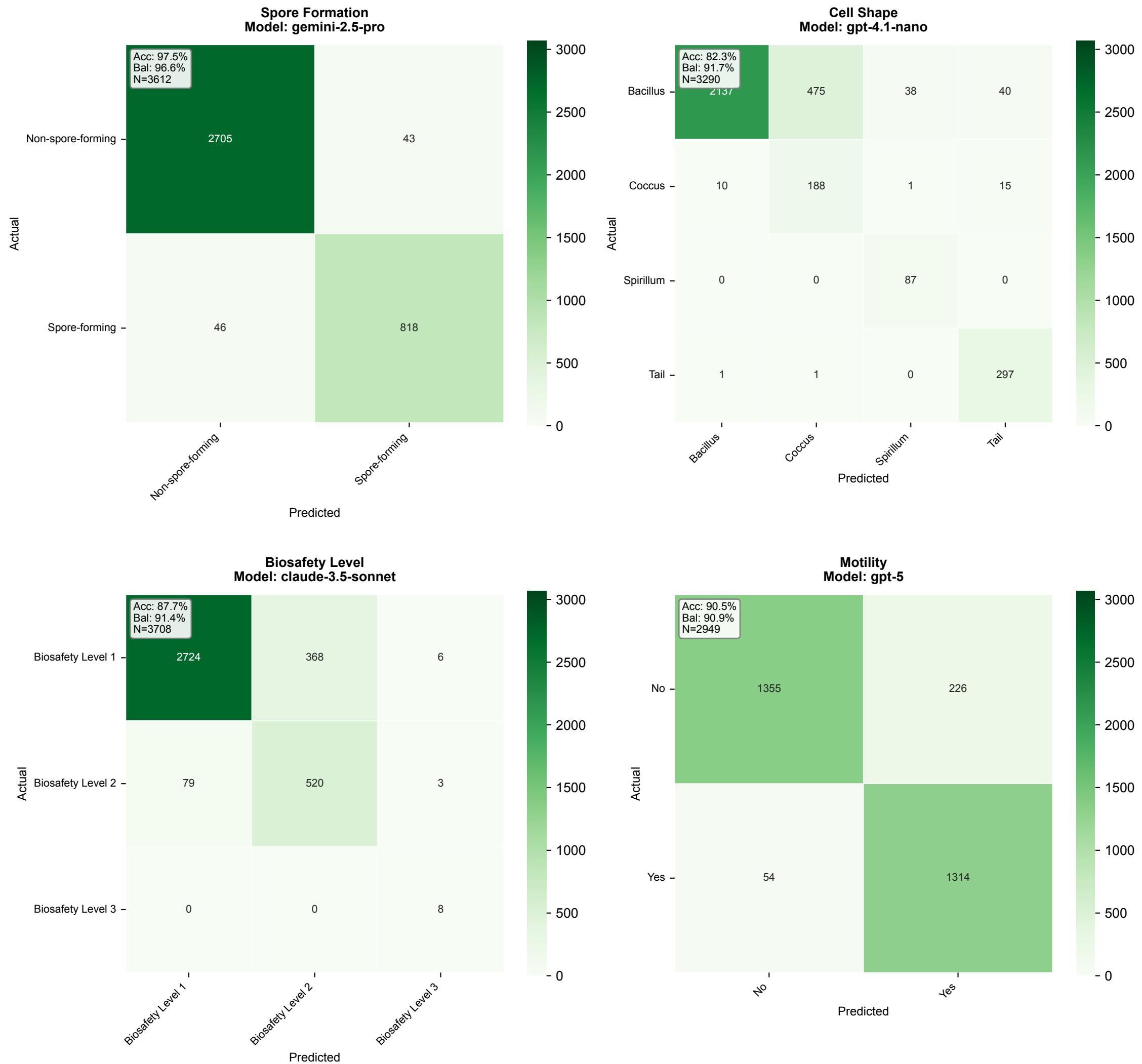
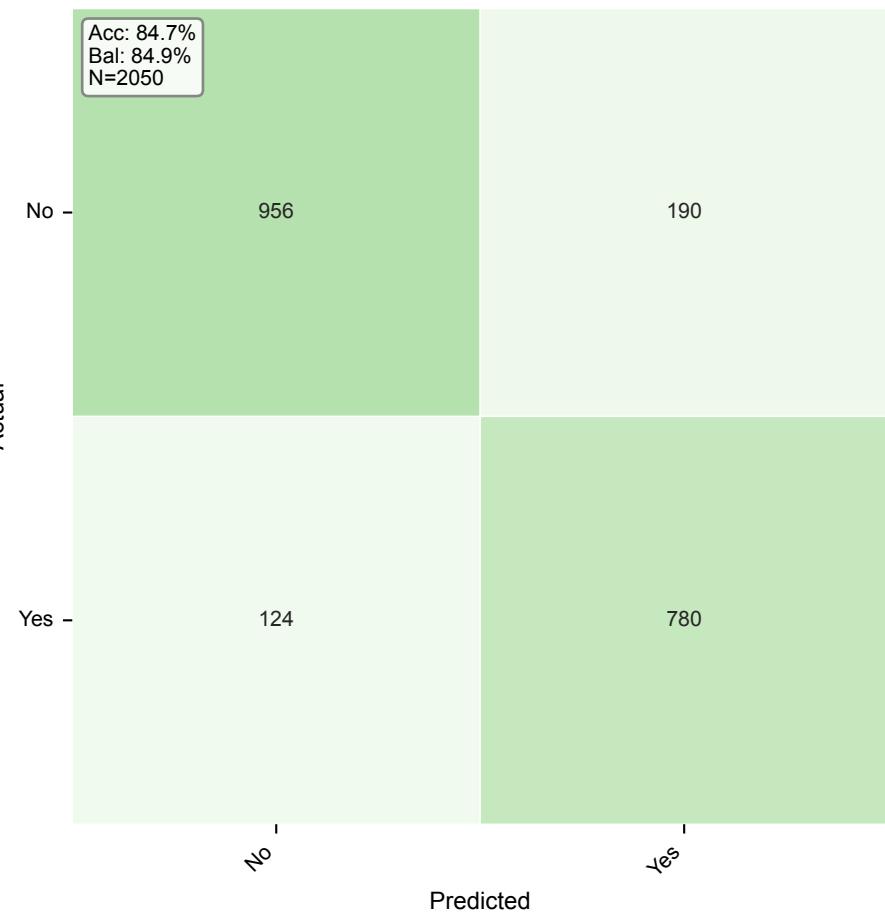


# Best Model Confusion Matrices for Each Phenotype

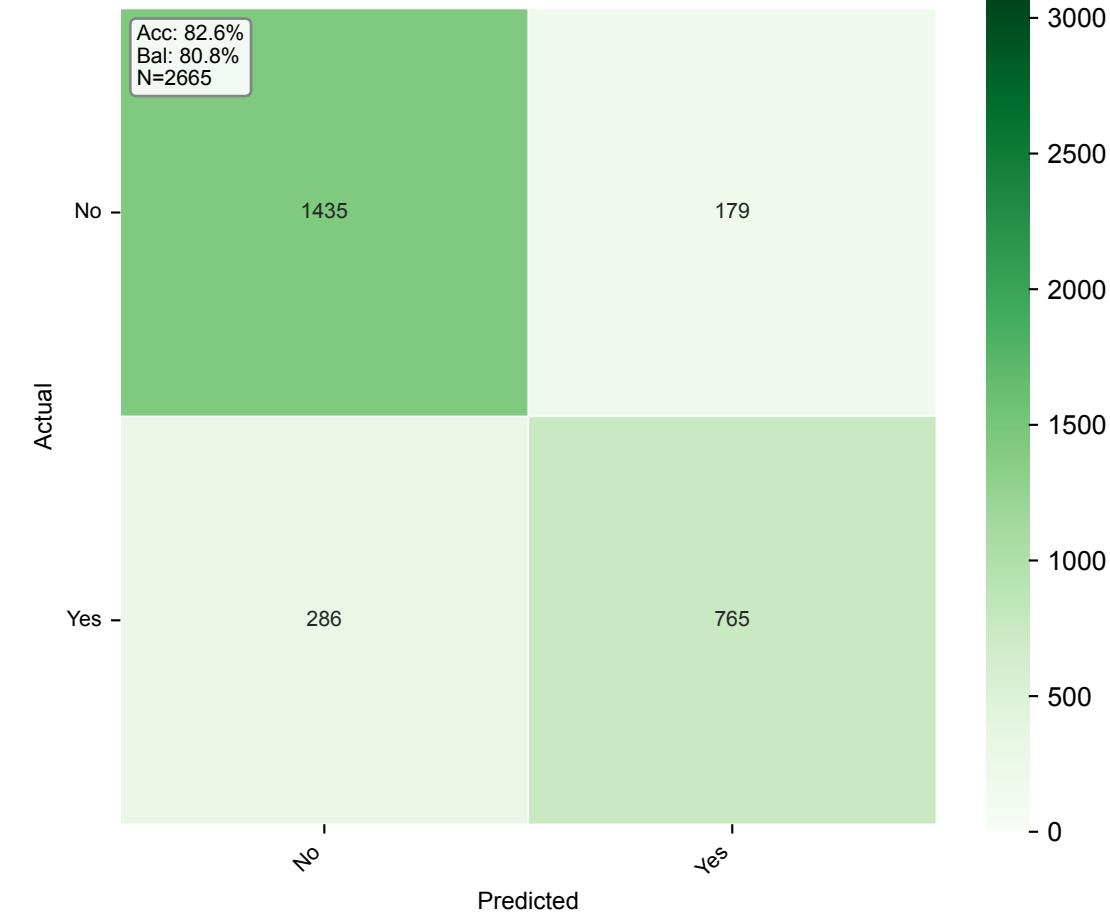


# Best Model Confusion Matrices for Each Phenotype

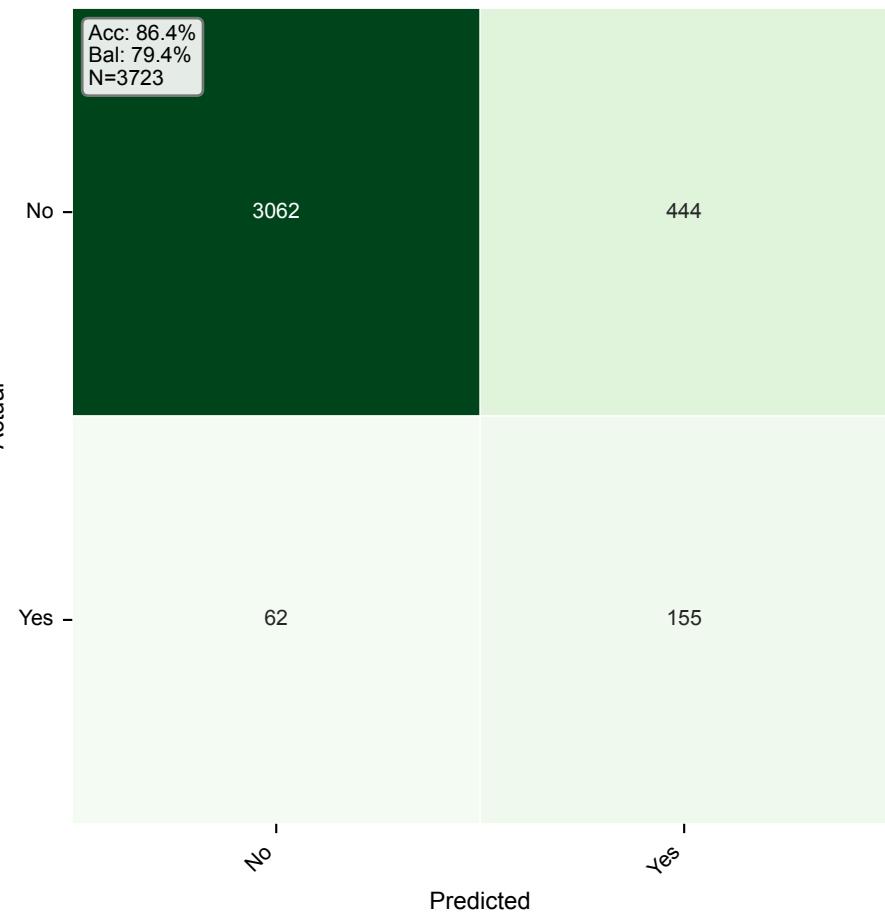
**Animal Pathogenicity**  
Model: gemini-flash-1.5



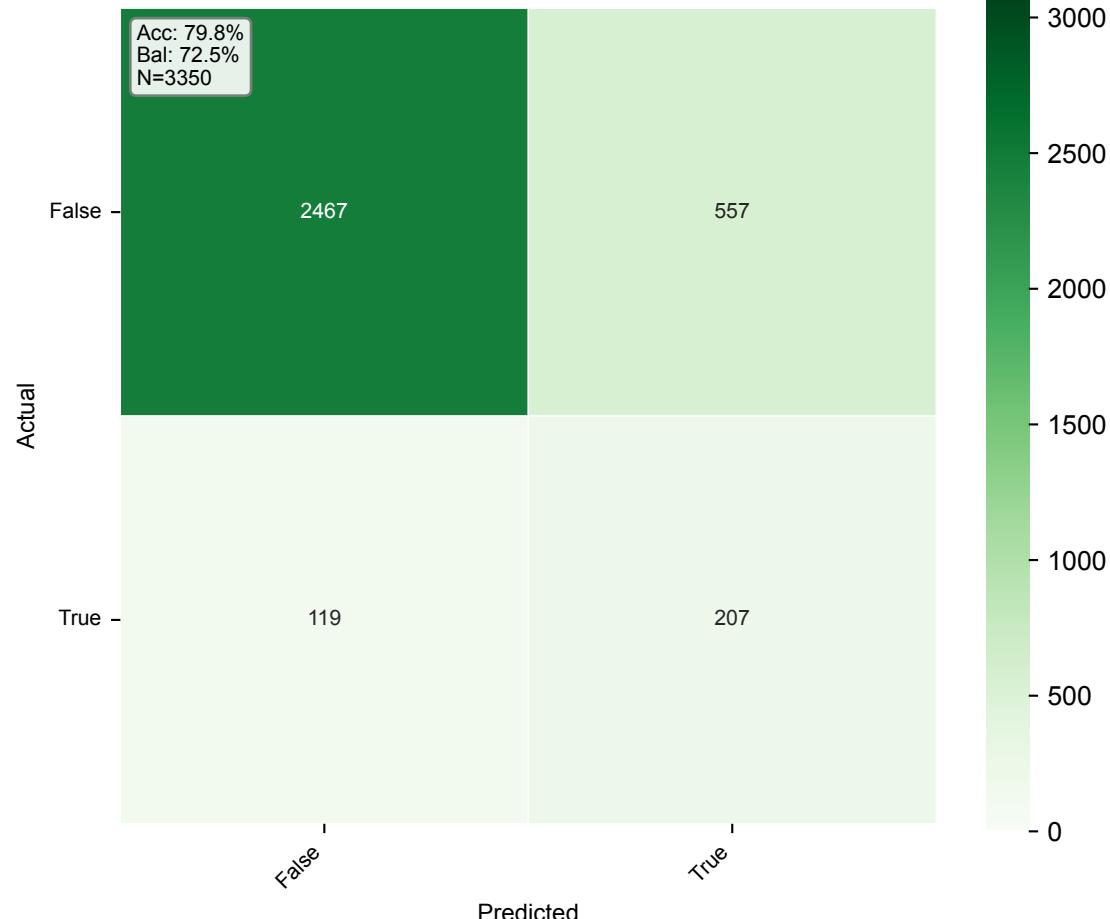
**Host Association**  
Model: gpt-4o



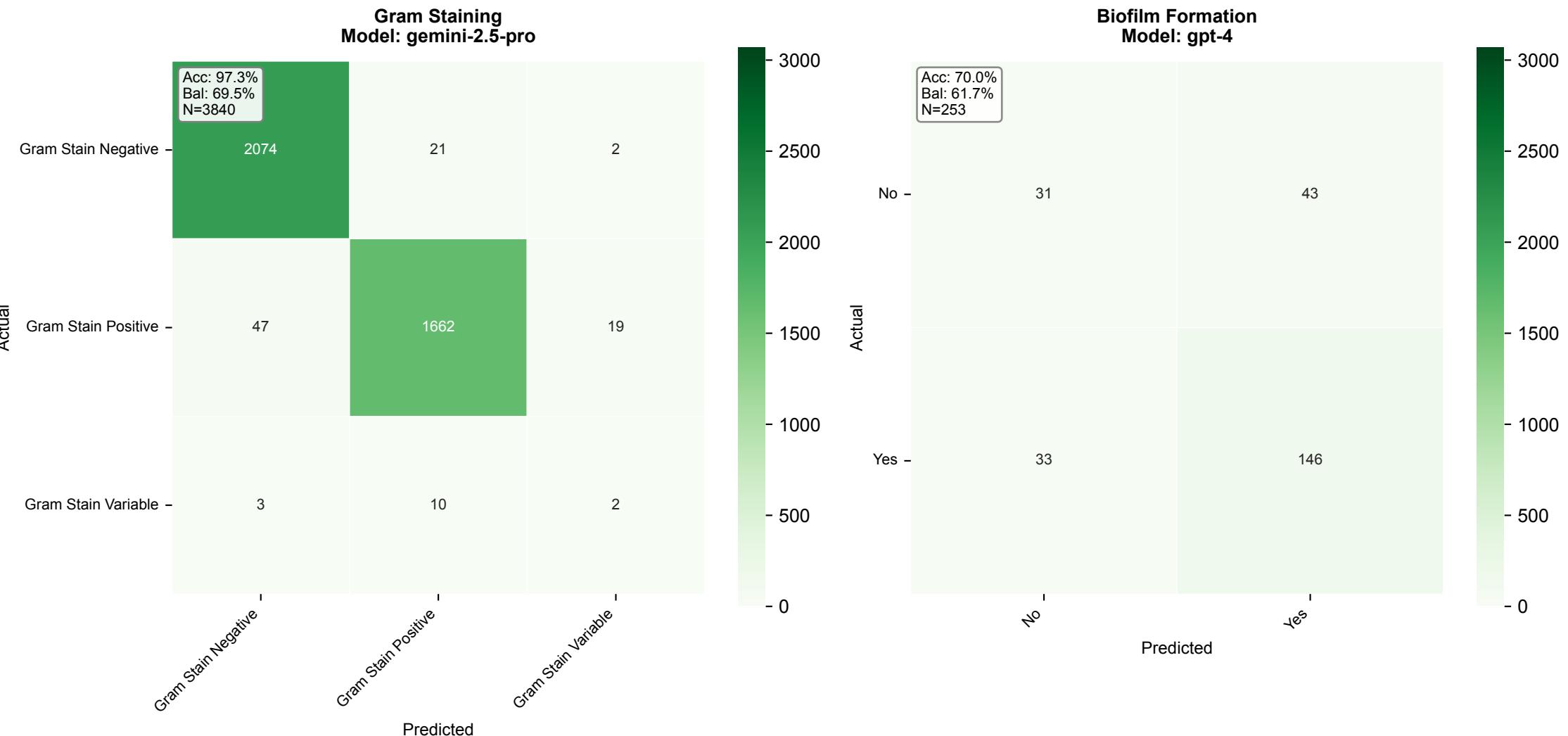
**Plant Pathogenicity**  
Model: gemini-pro-1.5



**Extreme Environment Tolerance**  
Model: deepseek-r1



# Best Model Confusion Matrices for Each Phenotype



## 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