Table 1:

| Analysis Phase | Research Question | Model ID | Statistical Method | Key Finding | Sample Size | Performance Metric¹*1* |
| --- | --- | --- | --- | --- | --- | --- |
| **Laboratory Development** | | | | | | |
| Discovery | Can immune genes predict infection costs? | **DISC-1** | Linear regression (PC1, PC2 → weight loss) | Significant but modest prediction | n = 136 | **R² = 0.106\*\*\*** |
| Optimization | Can machine learning improve prediction? | **DISC-2** | Random forest (19 genes → weight loss) | Substantial improvement achieved | n = 136 | **R² = 0.476\*\*\*** |
| Validation | Is the model reliable? | **DISC-3** | Train-test cross-validation | Strong predictive accuracy | n = 40 (test set)*2* | **r = 0.79\*\*\*** |
| **Cross-Population Translation** | | | | | | |
| Gene Validation | Which genes show consistent responses across populations? | **TRANS-1** | Linear regression per gene (lab vs field) | 3 genes cross-validated²*3* | n = 305 | **3/19 genes validated** |
| **Field Translation** | | | | | | |
| Detection | Does the model work in wild populations? | **FIELD-1** | Predicted vs. observed infection status | Successfully detects infection | n = 305 | **+1.15%\*\*\*** |
| Discrimination | Can it distinguish parasite species? | **FIELD-2** | Predicted loss by species identity | Species-specific responses | n = 169 | **E.f: +2.06%\*\* E.r: +1.25%\*\**4*** |
| Scaling | Does it correlate with infection severity? | **FIELD-3** | Predicted loss vs. parasite load | Scales with infection intensity | n = 185 | **R² = 0.114\*\*\*** |
| **Biological Validation** | | | | | | |
| Physiological relevance | Does it capture real health impacts? | **PROOF-1** | Predicted loss vs. body condition | Correlates with actual body weight | n = 336 | **β = -0.076\*** |
| Specificity | Is the response Eimeria-specific? | **PROOF-2** | Predicted loss vs. parasite community | Specific to Eimeria infections only*5* | n = 305 | **p < 0.001\*\*\*** |
| *1*Significance levels: \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001 | | | | | | |
| *2*Train-test validation: 70% training (n=96), 30% testing (n=40) | | | | | | |
| *3*Cross-validated genes: CXCL9 (both species), TICAM1 (E. falciformis), PRF1 (E. falciformis) | | | | | | |
| *4*E.f: Eimeria falciformis; E.r: E. ferrisi | | | | | | |
| *5*Parasite community model tested: Eimeria (significant), Aspiculuris, Syphacia, Trichuris, Mastophorus (all non-significant) | | | | | | |
| Framework demonstrates progression from basic linear prediction (R² = 0.106) through machine learning optimization (R² = 0.476) to comprehensive field validation with biological relevance. Cross-population translation validates 3/19 genes as conserved biomarkers. | | | | | | |