Table 1: AUC values for the best model of each classification scheme for both the observed (training) data and the generalized (testing) data. Results from all three different supervised learning approaches are shown here. AUC values range between 0.5 and 1.

	random forest		multinomial logistic regression		linear discriminate analysis	
Scheme	Observed	Generalized	Observed	Generalized	Observed	Geneeralized
Morph 1	0.63	0.73	0.75	0.79	0.75	0.80
Morph 2	0.61	0.58	0.76	0.77	0.76	0.77
Mito 1	0.63	0.62	0.75	0.63	0.75	0.63
Mito 2	0.77	0.67	0.80	0.64	0.80	0.63
Mito 3	0.56	0.64	0.71	0.74	0.71	0.73
Nuclear	0.56	0.67	0.74	0.62	0.74	0.77

Table 2: Results of bootstrap comparisons between the scheme with the highest mean AUC value and all other schemes, defined P(best - other > 0). An asterix indicates the best scheme. This was done for each of the three modeling techniques included in this study: random forest (RF), multinomial logistic regression (MLR), and linear discriminate analysis (LDA). Probabilities are the percent of comparisons that are greater than the observed difference in means.

Scheme	RF	MLR	LDA
Morph 1	$*$	$*$	1
Morph 2	0.79	0.55	1
Mito 1	0.89	0.94	1
Mito 2	0.82	0.57	$*$
Mito 3	0.79	0.69	0.73
Nuclear	0.79	0.96	0.96

Table 3: Results of comparisons between correctly and incorrectly classified observations from the testing data set, defined P(distance - simulated > 0). For each scheme, the classifications with at least 10 observations were tested. This was done for each of the three modeling techniques included in this study: random forest (RF), multinomial logistic regression (MLR), and linear discriminate analysis (LDA).

Scheme	Class	RF	MLR	LDA
Morph 1	CCR	0.77	1.00	1.00
	marm	0.87	0.93	1.00
Morph 2	CCR	0.88	1.00	1.00
	marm	1.00	1.00	0.98
Mito 1	CCR	0.94	0.99	0.98
	marm	0.99	0.98	0.47
Mito 2	marm	0.85	0.97	0.96
	pall	0.94	1.00	1.00
Mito 3	1	0.40	0.92	0.90
	3	0.97	0.99	1.00
Nuclear	marm	1.00	1.00	1.00
	pall	0.99	1.00	1.00

Table 4: Results from classification model estimates of the secondary, multi-species dataset. Models are random forest (RF), linear discriminate analysis (LDA), and multinomial logistic regression (MLR). Comparison of in-sample and out-of-sample AUC of the best performing model, along with the number of predictors. AUC values range between 0.5 (no better than random) and 1 (perfect classification).

	# predictors	In-sample AUC	Out-of-sample AUC
RF	11	1.000	0.999
MLR	6	1.000	0.998
LDA	10	1.000	1.000