Table S1. Optimized hyperparameters for the ML algorithms applied to the *M. canis* dataset.

model_name	RF	SVC	GB	LR	KNN
preprocess_fn	baug10	baug10	baug10	baug10	baug10
model	bootstrap=True	C=0.1	ccp_alpha=0.0	C=0.1	algorithm='auto'
	ccp_alpha=0.0	break_ties=False	criterion='friedman_mse'	class_weight=None	leaf_size=30
	class_weight=None	cache_size=200	init=None	dual=False	metric='minkowski'
	criterion='gini'	class_weight=None	learning_rate=0.1	fit_intercept=True	metric_params=None
	max_depth=None	coef0=0.0	loss='deviance'	intercept_scaling=1	n_jobs=None
	max_features='auto'	decision_function_shape='ovr'	max_depth=3	l1_ratio=None	n_neighbors=1
	max_leaf_nodes=None	degree=3	max_features='log2'	max_iter=100	p=2
	max_samples=None	gamma='scale'	max_leaf_nodes=None	multi_class='auto'	weights='uniform'
	min_impurity_decrease=0.0	kernel='linear'	min_impurity_decrease=0.0	n_jobs=None	
	min_impurity_split=None	max_iter=-1	min_impurity_split=None	penalty='l2'	
	min_samples_leaf=1	probability=False	min_samples_leaf=1	random_state=None	
	min_samples_split=2	random_state=None	min_samples_split=2	solver='lbfgs'	
	min_weight_fraction_leaf=0.0	shrinking=True	min_weight_fraction_leaf=0.0	tol=0.0001	
	n_estimators=2	tol=0.001	n_estimators=7	verbose=0	
	n_jobs=None	verbose=False	n_iter_no_change=None	warm_start=False	
	oob_score=False		presort='deprecated'		
	random_state=None		random_state=None		
	verbose=0		subsample=1.0		
	warm_start=False		tol=0.0001		
			validation_fraction=0.1		
			verbose=0		
			warm_start=False		
threshold	20	20	20	20	20
component_min	0	0	0	0	0
kfold	5	5	5	5	5
pc	0	0	0	0	0
mcc_cv	0.84	0.77	0.77	0.774	0.66
acc_cv	0.95	0.93	0.93	0.937	0.88
mcc_fit	1	0.95	0.95	0.952	1
acc_fit	1	0.98	0.98	0.989	1
pca					
idx	36531	36339	36434	36699	36692

Table S2. Optimized hyperparameters for the ML algorithms applied to the *M. gypseum* dataset.

model_name	SVC	LR	GB	RF	KNN
preprocess_fn	baug10	baug10	baug10	baug10	baug10
model	C=0.1	C=0.1	ccp_alpha=0.0	(bootstrap=True	algorithm='auto'
	break_ties=False	class_weight=None	criterion='friedman_mse'	ccp_alpha=0.0	leaf_size=30
	cache_size=200	dual=False	init=None	class_weight='balanced'	metric='minkowski'
	class_weight=None	fit_intercept=True	learning_rate=0.01	criterion='gini'	metric_params=None
	coef0=0.0	intercept_scaling=1	loss='deviance'	max_depth=9	n_jobs=None
	decision_function_shape='ovr'	l1_ratio=None	max_depth=3	max_features='auto'	n_neighbors=1
	degree=3	max_iter=100	max_features=10	max_leaf_nodes=None	p=2
	gamma='scale'	multi_class='auto'	max_leaf_nodes=None	max_samples=None	weights='uniform'
	kernel='linear'	n_jobs=None	min_impurity_decrease=0.0	min_impurity_decrease=0.0	
	max_iter=-1	penalty='l2'	min_impurity_split=None	min_impurity_split=None	
	probability=False	random_state=None	min_samples_leaf=1	min_samples_leaf=1	
	random_state=None	solver='lbfgs'	min_samples_split=2	min_samples_split=2	
	shrinking=True	tol=0.0001	min_weight_fraction_leaf=0.0	min_weight_fraction_leaf=0.0	
	tol=0.001	verbose=0	n_estimators=10	n_estimators=5	
	verbose=False	warm_start=False	n_iter_no_change=None	n_jobs=None	
			presort='deprecated'	oob_score=False	
			random_state=None	random_state=None	
			subsample=1.0	verbose=0	
			tol=0.0001	warm_start=False	
			validation_fraction=0.1		
			verbose=0		
			warm_start=False		
threshold	30	30	30	30	30
component_min	0	0	0	0	0
kfold	5	5	5	5	5
pc	0	0	0	0	0
mcc_cv	0.80	0.80	0.75	0.69	0.61
acc_cv	0.94	0.93	0.92	0.90	0.86
mcc_fit	0.95	1	0.45	1	1
acc_fit	0.98	1	0.64	1	1
pca					
idx	36339	36699	36485	36586	36692