

Baseline Subtraction	Tophat filter Factor: 0.02
Noise Reduction	Savitzky-Golay filter -Window length: 11 -Polynomial order: 3
Alignment	Allowed shift: Medium Mass tolerance for aligning the spectra: -Linear Mass tolerance: 2000 ppm
Peaks Finding	By threshold -Peak threshold: 0.01 -Constant Mass tolerance: 0.5Da -Linear Mass tolerance: 2000ppm -Analysis Data: Peak intensity

Table S1. Pre-treatment of raw data spectra for the biomarkers search.

Baseline Subtraction	Tophat filter Factor: 0.02
Noise Reduction	Savitzky-Golay filter -Window length: 11 -Polynomial order: 3
Alignment	Allowed shift: Medium Mass tolerance for aligning the spectra: -Linear Mass tolerance: 2000 ppm
Peaks Finding	By mass position -Constant Mass tolerance: 0.5Da -Linear Mass tolerance: 2000ppm
Merge	Mass tolerance for aligning the peaks - Constant Mass tolerance: 1Da - Linear Mass tolerance: 2000ppm
Normalization	-Total Ion Count (TIC)

Table S2. Parameters applied for the construction of the peak matrix.

A

Score: 100 %

PC 0 (46.73%)

Actual / Predicted	Deneoformans	Hybrid/Neoformans	% Correct
Deneoformans	7	0	100 %
Hybrid/Neoformans	0	58	100 %
	100 %	100 %	100 %

Confusion Matrix

k-fold Cross Validation

10-fold cross validation

Actual / Predicted	Deneoformans	Hybrid/Neoformans	% Correct
Deneoformans	7	0	100 %
Hybrid/Neoformans	0	58	100 %
	100 %	100 %	100 %

Confusion Matrix

B

Score: 98.46 %

From / to	Deneoformans	Interspecies hybrids	Neoformans	% Correct
Deneoformans	7	0	0	100 %
Interspecies hybrids	0	23	1	95.83 %
Neoformans	0	0	34	100 %
				98.46 %

Confusion Matrix

k-fold Cross Validation

10-fold cross validation

From / to	Deneoformans	Interspecies hybrids	Neoformans	% Correct
Deneoformans	6	1	0	85.71 %
Interspecies hybrids	0	23	1	95.83 %
Neoformans	0	0	34	100 %
				96.92 %

Confusion Matrix

Table S3. Score and validation of (A) the SVM analysis applied to the two-step model with 5 biomarkers and (B) PLS-DA applied to the one-step model with 5 biomarkers. K=10