BUILDING A BIOSIGNATURE ROCK SAMPLE LIBRARY AND DEVELOPING AUTOMATED CLASSIFIERS. Virginia C. Gulick[1], Sascha T. Ishikawa, Patrick M. Freeman[1,2], Timothy Johnsen[1,3], Jason Angell[1,4,] Paige Morkner[1,5], and Job Bello[5]. [1]NASA Ames/SETI Institute (NASA Ames Research Center, MS 239-20, Moffett Field, CA 94035, Virginia.C.Gulick@nasa.gov), [2]UC- Santa Cruz, [3]UC-Irvine, [4]CalPoly-San Luis Obispo, [5] EIC Laboratories, Norwood, MA.

Introduction: Identifying minerals, organics and other potential biosignatures within individual rock samples is an important part of both terrestrial field and planetary surface exploration studies. To help with this effort, Raman spectrometers are being increasingly utilized. Raman spectroscopy is a nondestructive method to acquire spectra relatively quickly with little or no sample preparation, making it ideal for use on future planetary missions. This method provides unique fingerprint spectra of the sample being examined. However, because the resulting spectra from rocks may contain a variety of minerals and biosignatures, in addition to sources of noise including background and mineral fluorescence, it can be difficult for a field scientist to immediately identify the individual spectral constituents. Consequently, rock samples are often collected and re-analyzed back in the lab under controlled conditions and compared with a library of known mineral and biosignature spectra.

Automatically Classifying Minerals: Our group has previously developed an automated mineral classifier to autonomously identify pure minerals from Raman spectra [1]. We utilized Principle Component Analysis (PCA) to reduce the dimensionality of the data prior to training with an Artificial Neural Network (ANN) implemented by the Multi Layer Perceptron (MLP) algorithm to automatically classify some of the major rock forming minerals, including quartz, olivine, potassium feldspar, plagioclase, mica, and pyroxene. Using Raman Spectra of mineral samples from our library and our 852 nm laser excitation instrument, our classifier performed with an overall accuracy of ~83%. Quartz and olivine returned an accuracy of 100%. Using the online RRUFF mineral database (http://rruff.info) of these same minerals, our classifier performed with an overall accuracy of ~80%, while quartz, olivine and pyroxene returned an accuracy of 100%.

Since then, we used this same classifier on mineral spectra common to sedimentary rocks using the RRUFF database. We selected sulfates, carbonates, oxides, feldspars, quartz, and micas. Our classifier returned an overall accuracy of ~92%. See Table 1 for results from individual minerals.

Having demonstrated the efficacy of our mineral identification system, we are now set to develop automated algorithm(s) that will enable individual minerals and biosignatures to be automatically identified and classified from rock spectra. Such automated algorithms can aid in a variety of applications and can contribute to rock identification systems when couled with image analysis algorithms [2,3] to classify rock samples.

Table 1: Automated classification of minerals common to sedimentary rocks using spectra from the RRUFF database. On average, 92.3% of the classifications were correct.

	SUL	CAR	OXI	MCA	QTZ	FEL	% Correct
SUL	172	11	5	0	0	0	91.5
CAR	17	153	1	2	0	0	88.4
OXI	0	0	13	4	0	0	76.5
MCA	1	2	12	80	1	0	83.3
QTZ	0	0	0	0	21	0	100.0
FEL	0	2	0	0	0	289	99.3

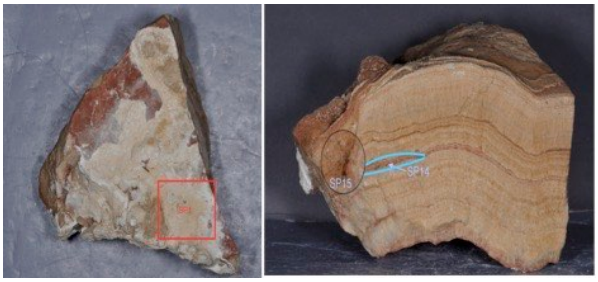
Building the Spectral Library of Rock and Biosignature Samples: However the development of automated classifiers for multi-minerals, mineral mixtures and biosignature spectra requires establishing a library of biosignature samples as well as analyzed rock samples containing multiple minerals and biosignatures. Such a custom library is required since no publically available online sample library currently exists. To this end, we have been building such a sample library.

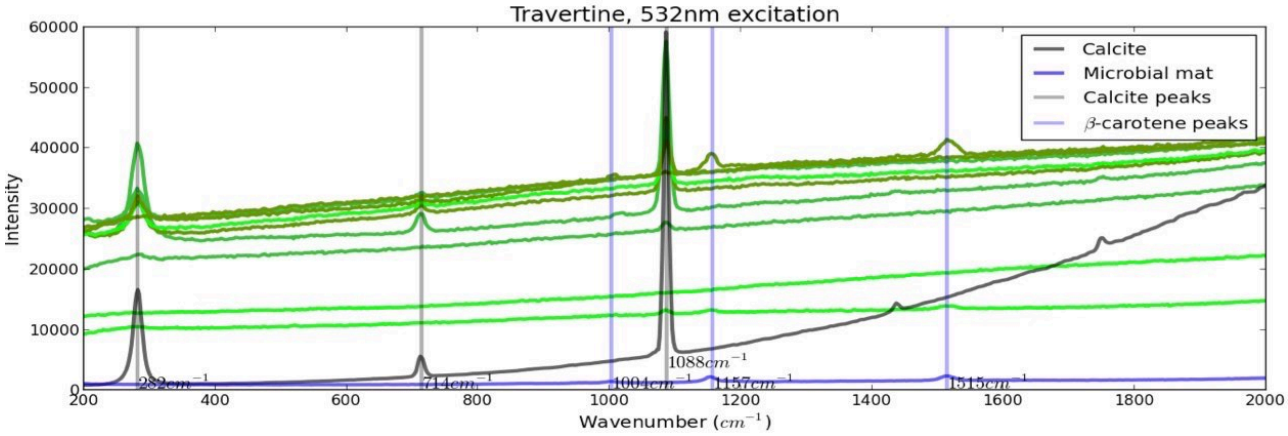
Currently our sample library encompasses over 1200 samples that have undergone careful hand sample analysis and macro imaging under uniform lighting conditions. We are in the process of acquiring Raman spectra of these samples with our in house dual excitation (532nm and 785nm) Raman instrument provided by EIC Laboratories. Approximately 700 of these samples are igneous rocks, with the remainder being of sedimentary and metamorphic origin. We also have a wide variety of common rock forming minerals that we have analyzed with our Raman instrument. More recently, we have begun to collect a variety of sedimentary samples containing biosignatures, including travertine from California hot springs containing beta-carotene and other biosignatures, gypsums and salts from the Atacama Desert, palmitic acid and gypsum and anhydrite containing palmitic acid, diatomite, and microbial mats.

Locating Biosignatures in Rock:

An important component of sample analysis on future missions will be location selection on the sample. Prepared samples are mostly homogeneous, but natural

Figure 1: Travertine sample from Travertine Springs, CA. Image on left shows region where no biosignatures were identified. Image on right contained beta-carotene peaks in two protected regions circled. Plot shows several Raman spectra of the sample (green) compared with β-carotene peaks in a microbial mat (blue) and calcite (gray) spectra. Distinct calcite peaks (gray vertical lines) as well as distinct β-carotene *peaks (blue vertical lines) in the sample demonstrate the ability to detect both minerals and biosignatures in the same spectra. (L. Bebout (NASA-ARC) provided microbial mat*





rock samples are more heterogeneous and non-uniform. Analysis of the travertine sample shown in Figure 1 revealed ß-carotene peaks in the more protected locations on the rock, while large flatter areas appeared devoid of spectral signatures pointing to life.

Developing Multi-Mineral, Mixed Mineral and Biosignature Classifiers.

As we move away from classifying pure minerals, we need to be able to classify individual minerals and biosignatures from spectra taken from rocks.

Therefore, we performed a preliminary test of our mineral classifier to compare them with results from our hand sample analysis. We randomly selected 22 rock samples from our collection and acquired several spectra from each. We tested the ability to detect the presence or absence of plagioclase, potassium feldspar, quartz, and pyroxene in each spectra. This simple test yielded an overall success rates of 85%, with 91.3% of the plagioclase, 87% of the K-feldspar, 83% of the quartz, and 78% of the pyroxene being correctly identified.

However, we have also been evaluating numerous other machine learning approaches (MLA) in combination with a wider range of pre-processing techniques on acquired spectra to improve results. We also included cross-validation and voting schemes. We use a variety of combinations which we call trainers. A trainer is composed of various pre-processing methods and one select MLA. Pre-processing techniques we have found to be the most effective are Saviztky Golay Polynomial Fit (to deal with noise), Truncating Overly-Noisy Wave Numbers (to remove incoherent data), Peak Finding + Baseline Removal (to help with fluorescence), Normalizing Data (since different spectra yield different relative intensities), and Squashing Data (square-rooting normalized data to help strengthen the weaker Raman peaks). The most successful Machine Learning Algorithms (MLA) so far, from various testing and reasoning, are Adaptive Boosted Trees, Multilayer Perceptron, Gradient Boosted Trees, Bayes, Decision Trees, (Extremely) Random Trees, and Support Vector Machines. Each trainer yields a different perspective on the data. The next step is to determine a set of uncorrelated, significant trainers.

To accomplish this, we have built an ensemble of trainers. Each ensemble is a different combination of trainers. We then determine which of these ensembles produce the most accurate results, therefore yielding the highest success percents. Several methods are used to avoid overfitting. Tests are currently underway, and results are pending.

References: [1] Ishikawa, S.T. and V.C. Gulick (2013) *Computers and Geosciences, doi: 10.1016/ j.cageo.2013.01.011.* [2] Freeman et al., 2014. LPSC,2739. [3] Valenzuela et al., 2015, LPSC 3009.