This is a preprint, the final version is subject to change, of the American Mineralogist (MSA)
Cite as Authors (Year) Title. American Mineralogist, in press.
(DOI will not work until issue is live.) DOI: http://dx.doi.org/10.2138/am-2015-5185 4/1

Revision 1

Provenance Determination of Sapphires and Rubies using Laser-Induced Breakdown Spectroscopy and Multivariate Analysis

Kristen A. Kochelek[1], Nancy J. McMillan[1], Catherine E. McManus[2], and David L. Daniel[3]
[1]Department of Geological Sciences, New Mexico State University
[2]Materialytics, LLC
[3]Department of Economics, Applied Statistics, & International Business,
New Mexico State University

Abstract

Determination of gem provenance is a topic of research in the gemological community for financial, security, and societal reasons. Laser-Induced Breakdown Spectroscopy (LIBS) and multivariate analysis have the potential to revolutionize the field of gem provenance. This study acquired LIBS spectra from 569 rough sapphire and ruby specimens from 21 localities in 11 countries. The spectra were analyzed using the multivariate technique PLSR (Partial Least Squares Regression) in separate algorithms for sapphires and rubies. Each algorithm consists of a series of PLS models. Each model compares the spectra from a locality of interest to the spectra from all other localities in the database. Success rates, as determined by the percent of correct provenance identifications, are 98.9% (sapphire) and 96.0% (ruby) for country of origin and 97.9% (sapphire) and 95.4% (ruby) for deposit of origin. Individual deposits are not recognized by the concentrations of a few elements; rather, the unique geochemical signature of each deposit consists of the ratios of many elements, primarily Ca, Zr, Fe, Ba, Mt, Ti, Sr, Si, Cr, H, C, and Li, some of which may reside in inclusions. This work demonstrates that determination of country or deposit of origin can be related to a quantitative measure with a high level of success.

Keywords: ruby, sapphire, provenance, laser-induced breakdown spectroscopy, chemometrics

Introduction

The relationship between gem provenance (i.e., country or deposit of origin) and monetary value, combined with the difficulties in provenance determination, have fueled decades of gemological research and limited transparency in the gem industry. In 1990, Gemological Institute of America (GIA) chairman Richard T. Liddicoat explained the technical challenges that exist for determining the source of a gem with any amount of certainty (Liddicoat, 1990). Despite his

concerns with the ability and need to accurately determine provenance, demand from both the gem trade and the public for provenance certification has grown (Rossman, 2009; Shor and Weldon, 2009).

Demand has also risen for gem provenance determination because of socio-political concerns. In 2003, the United States banned trade with Myanmar through the Burmese Freedom and Democracy Act, citing concerns of human rights abuses in the country (US Dept. of State, 2008a). In a 2008 amendment, the Tom Lantos Block Burmese Jade Act of 2008 closed a loophole in the legislation, specifying that any rubies mined in the country could not be legally imported into the United States for commercial purposes (US Dept. of State, 2008b). In order to enforce sanctions such as these, a reliable method of determining the provenance of gems is required.

Gemstones from different localities possess different characteristics. Provenance has been determined using a combination of inclusion analysis, trace element chemistry, special characteristics, and internal growth structures (Gübelin, 1953; Gübelin and Koivula, 1986; Ward, 1995; Hughes, 1997; Abduriyim and Kitawaki, 2006b; Devouard and Notari, 2009). To gather this information, one must send a gem in question to a laboratory where a series of tests are performed. Despite the amount of information obtained from these tests, it is cautioned that declarations of provenance be regarded only as the professional opinion of the laboratory or gemologist (Ward, 1995; Abduriyim and Kitawaki, 2006b). It is evident that an objective and reliable technique needs to be developed to determine provenance.

Rubies and sapphires are among the most valuable gemstones. This pilot study uses spectra acquired with Laser-Induced Breakdown Spectroscopy (LIBS) from 569 ruby and sapphire specimens from 21 localities in 11 countries to demonstrate a highly successful method for determining the provenance of corundum samples. An attempt was made to characterize each deposit with analysis of 30 samples; however, sample set size ranged from four to 40 samples per deposit. The method utilizes the multivariate statistical analysis technique PLSR (Partial Least Squares Regression) to build a series of models that sequentially compares spectra from one region to all other regions. This method has proven successful in distinguishing between

bacterial pathogens (Multari et al., 2010) and individual limestone beds (McMillan et al., 2012). Success rates, as determined by the percent of correct provenance identifications, are greater than 95% for both country of origin and deposit of origin. LIBS analysis is also minimally destructive (similar to Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry, or LA-ICP-MS), simple to use, and portable. This study demonstrates the potential for LIBS analysis to quantitatively and reliably introduce a new level of transparency to the gem industry.

Background

Geochemical analysis of gemstones enhances techniques for provenance determination by adding to the traditional methods of observations of optical and mineralogic features of the specimen. These observations include index of refraction; UV-visible, infrared, and Raman spectroscopy; fluid and mineral inclusions; and the presence or absence of crystallographic features unique to a specific deposit (Gübelin, 1953; Gübelin and Koivula, 1986; Ward, 1995; Hughes, 1997; Abduriyim and Kitawaki, 2006b; Devouard and Notari, 2009). Because these optical and mineralogic features may not be observed in every specimen from a locality, this method has proven to be useful for some samples but not for all (Muhlmeister et al., 1998).

Early analytical methods applied to ruby and sapphire provenance were PIXE (Particle-Induced X-ray Emission) and ED-XRF (Energy-Dispersive X-ray Fluorescence). Both techniques are non-destructive and determine the concentrations of several elements in the gemstone. For instance, Calligaro et al. (1999) were able to positively identify the origin of 41 of 64 rubies set on an elaborate necklace using multivariate analysis on PIXE analysis of V, Cr, Fe, Ti, and Ga. A database of 200 analyses of rubies from nine countries was used as a comparative database. Muhlmeister et al. (1998) demonstrated through ED-XRF analysis that chemical characteristics of rubies vary from one source to another. The ternary diagram Fe-V-Ga was used to discriminate rubies from basalt-hosted, metasomatic, and marble-hosted hosts; however, rubies from various localities exhibit significant overlap on the triangular diagram. Similarly, Schwarz and Schmetzer (2001) developed the use of ED-XRF data and the Cr/Ga vs. Fe/Cr diagram to distinguish rubies from Myanmar (Mong Hsu and Mogok), Madagascar (Vatomandry) and Thailand-Cambodia. Rubies from these three areas show good separation on this two-variable diagram.

The development of the Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) from traditional ICP-MS techniques allowed for minimally destructive trace element analysis (Gray, 1985; Günther and Kane, 1999). Rankin et al. (2003) applied LA-ICP-MS data to the Cr/Ga vs. Fe/Cr diagram of Schwarz and Schmetzer (2001) to discriminate between the east African localities Longido and Chimwdzulu. Saminpanya et al. (2003) proposed the use of the Cr_2O_3/Ga_2O_3 vs. Fe_2O_3/TiO_2 diagram to differentiate between rubies and sapphires and between metamorphic and magmatic corundum samples. This discrimination diagram was employed by Abduriyim and Kitawaki (2006b), Sutherland et al. (2009) and Sutherland and Abduriyim (2009) to determine provenance of various suites of gem corundums. Peucat et al. (2007) proposed that the Fe vs. Ga/Mg diagram is useful for discriminating between magmatic and metamorphic blue sapphires in a study with 114 samples from 18 locations. Guillong and Günther (2001) applied multivariate statistics to LA-ICP-MS analysis of seven elements (Mg, Si, Ti, V, Cr, Fe, and Ga) of 25 sapphires from five locations to demonstrate provenance determination. Separation of the groups was fairly good, but some overlap of fields existed. Similarly, Pornwilard et al. (2011) applied multivariate analysis to 10 elements determined by LA-ICP-MS (B, Si, Zn, Ga, Sn, V, Mg, Ti, Cr, and Fe), using 58 samples from six countries. Samples from Kenya and Nigeria were easily distinguished. However, samples from Madagascar and Tanzania were indistinguishable from each other, although together they were distinguished from the other groups. The same was true for samples from Cambodia and Thailand. Although high-precision LA-ICP-MS analysis has proven to be a minimally destructive technique for provenance determination, the fields for many of the countries of origin overlap on these diagrams. All of these diagrams use only a few (three to ten) variables to model a complex chemical system, resulting in insufficient definition of each provenance group. In addition, the difficulty and expense of analysis, combined with the cost of the samples, restricts such studies to relatively few samples (fewer than 15) typically used to characterize each location. This sample size may be insufficient to fully characterize the chemical variability in the deposit, suggesting that the overlap in compositions might be greater if a larger sample set were to be utilized.

The laser fluorination technique to analyze oxygen isotopes developed by Sharp (1990) made *in situ* analysis of a small amount of sample possible; modifications of this technique were made

for use with corundum (Yui et al., 2003; Giuliani et al., 2005). Oxygen isotopes have been used to determine the geologic origin of rubies (Giuliani et al., 2005), and have proved useful in determining the source lithology of xenocrystic or placer corundum specimens. Marble-hosted corundum samples have high $\delta^{18}O$, reported by Guiliani et al. (2005) to range from +16.3 ‰ to +23‰ and by Yui et al. (2008) from +21.6 ‰ to +25.7 ‰. Rubies and sapphires from basalts, igneous rocks, and non-carbonate metamorphic rocks have lower $\delta^{18}O$, ranging from +1.3 ‰ to +13.9 ‰ (Yui et al., 2003; Guiliani et al., 2005; Yui et al., 2006; Giuliani et al., 2009; Vysotsky et al., 2009). Ruby and sapphire specimens from the same placer deposit have been shown to have distinct $\delta^{18}O$ values (Yui et al., 2006), suggesting different source locations for the placer. Overall, oxygen isotope analysis of corundum has been able to set some constraints on the source lithology of rubies and sapphires, despite significant overlap in isotopic composition of samples from various igneous and non-carbonate metamorphic lithologies (Guiliani et al., 2005). However, the isotopic compositions of corundum samples are not diagnostic of locality, limiting the use of this technique for provenance determination.

Traditional techniques for determining ruby and sapphire provenance, including observation of inclusions and other mineralogical features, trace element analysis by XRF, PIXE, and LA-ICP-MS, and oxygen isotope analysis have not yet been able to provide a unique fingerprint for each corundum location. In addition, analytical techniques like LA-ICP-MS require skilled technicians in a laboratory environment. In contrast, Laser-Induced Breakdown Spectroscopy (LIBS) is a rapid and inexpensive technique that is simple to operate, portable, and does not require special laboratory conditions. Thus, provenance of gems using LIBS could be determined at border crossings, ports and airports, gemological laboratories, gem studios, and galleries.

Materials and Methods

Specimen Acquisition. A large sample set of rough corundum was acquired from vendors at the Tucson Gem and Mineral Show and mining companies (Table 1). Heat-treated, chemically-treated, and cut gemstones were not used in this study. Because acquisition of samples with known provenance for database development is critical to the study, gems were purchased only from vendors who were able to cite a chain of possession and were knowledgeable about the

mines of the region. Every precaution was taken to ensure the reliability of the vendors. For the purposes of this study, corundum with deep pink, fuchsia, maroon, and purple-red colorations were considered rubies, and all other colors were considered sapphires. Countries were characterized by 30 samples, if possible, to capture the chemical variability within the deposits with statistical significance. Five sapphire samples from Sri Lanka and two rubies from Jagdelek, Afghanistan, were originally included; however, there was insufficient information to fully characterize the deposits, and provenance determination failed for these localities. Although several of the data sets have fewer than 30 samples (Psudipada, India; Umba Valley, Tanzania; Mysore, India; Hunza Valley, Pakistan; and Longido, Tanzania; Table 1), provenance determinations from these localities yielded acceptable accuracies. However, this work suggests that a large number of samples ($\geq$ 30) need to be used to characterize the chemical fingerprint from a gem locality.

Sample Analysis. Samples were analyzed by Laser-Induced Breakdown Spectroscopy (LIBS) (Fig. 1). In LIBS, a laser beam is pulsed on a small area of the sample to create a plasma, a weakly ionized collection of ions, atoms, and free electrons (Cremers and Radziemski, 2006). The plasma contains atoms ablated from the surface of the sample and atoms in the local environment (air or a gas selected for analysis). Due to the high temperature of the plasma (> 10,000 K), electrons are excited to upper orbitals. As the plasma cools, each electron decays to a lower-energy orbital, and the difference in energy is emitted as a photon with a discrete wavelength that corresponds to the energy difference between orbitals. The photons are collected by a lens into seven optic fibers. A spectrometer, which separates the light into individual wavelengths, is attached to each optic fiber. The spectra from all seven spectrometers are compiled to form a single, continuous spectrum (Fig. 2). The intensities of peaks at specific wavelengths are proportional to the number of photons emitted by a specific orbital transition from a specific element (Cremers and Radziemski, 2006). For example, the 396.1 nm peak for Al is the result of an excited electron in the 4s orbital decaying to the stable 3p orbital. This is statistically the most intense Al peak. Because many electron orbital transitions occur for each element, a LIBS spectrum consists of many peaks for each element. In this work, peaks are identified using the Handbook of Basic Atomic Spectroscopic Data published by NIST (Sansonetti and Martin, 2005).

The Ocean Optics © 2500+ LIBS instrument was used for this study (Fig. 1). A 1064 nm Nd:Yag laser was pulsed at 120 mJ with a Q-switch delay of 1.5 μs for each analysis. The Q-switch delay is the interval of time between plasma formation and the collection of light. A delay in collection is necessary because at the instant of plasma formation, a continuum is formed from bremsstrahlung processes and recombination events (Cremers and Raziemski, 2006). During these events, photons are emitted as electrons collide in the plasma. Photons emitted from these collisions do not provide elemental data, so it is necessary to delay collection until these events are no longer significant in the signal.

Analysis protocol was designed to maximize spectral intensity while minimizing damage to the sample. Samples were analyzed inside a sample chamber in an argon atmosphere so that the laser energy is used to excite atoms from the sample instead of atmospheric nitrogen and oxygen. At each analyzed spot, a cleaning shot was taken, followed by an analysis shot. The cleaning shots were at the same energy level as the analysis shots; the data simply were not recorded. Ablation craters were 10-100 μm in diameter.

Each specimen was analyzed at 30 spots; these spectra were averaged to create one spectrum per specimen. The Al peaks at 226.9, 236.7, 237.3, 257.5, 281.6, 308.2, 394.4, and 396.1 nm were removed from the spectra to allow multivariate analysis to focus on trace element distributions. All wavelengths > 680 nm except for the 854.2 nm Ca peak for rubies (Fig. 2) were also removed in order to eliminate Ar peaks from the analysis. The 854.2 nm Ca peak was retained in the ruby spectra because it was observed to vary between deposits.

Chemometric Analysis. Chemometric analysis is a method of processing chemical data using multivariate statistics in an attempt to recognize 'hidden phenomena' (Esbensen, 2007). Hidden phenomena are correlations that are not intuitive, or obvious from a simple two-dimensional, or bivariate, graph. In this study, The Unscrambler ® was the multivariate data analysis software used to model LIBS data.

Partial Least Squares Regression (PLSR) is a multivariate modeling technique that relates two data matrices by multiple linear regressions (Esbensen, 2007). In this study, one matrix (X) contains the LIBS spectra; the other matrix (Y) contains a single integer variable, called the provenance variable, that takes the value of 1 if the specimen came from the locality under scrutiny and 0 if the specimen came from any other locality. PLSR models the X and Y matrices interdependently during calibration, reducing the influence of large variations in X that are irrelevant to Y (Wold et al., 2001). Regression coefficients, the characteristics of X that are relevant to Y, are used to predict Y-values (provenance variables) from X-values (LIBS spectra). This is a powerful technique because it uses all channels of the LIBS spectra, which record emission from nearly all elements from the periodic table.

In this work, PLSR models are generated and tested in four steps (Fig. 3): 1) training of a model; 2) determination of the Value of Apparent Distinction (VAD) that separates two groups of samples; 3) prediction of provenance variables for validation samples; and 4) calculation of success rate by comparing the predicted provenance to known provenance of validation samples. In order to validate a model, it is critical to have spectra from a set of samples of known provenance that can be treated as unknowns. To accomplish this, half of the samples from each locality were used to calibrate the models and the remaining half were used in validation to verify the model and calculate success rates.

A PLSR model is trained by correlating matrices X (spectra) and Y (provenance variables) using multiple linear regressions. The model relates pertinent aspects of the spectra to provenance variables, 1 or 0, assigned to two groups of samples in the model. The provenance variable "1" is assigned to the spectra from the provenance of interest and the variable "0" is assigned to a group containing all other spectra. The model consists of a set of regression coefficients, one for each variable, which relate the spectral data to the provenance variables. This process is illustrated in Figure 3A, in which the model is represented by a PLSR score plot. The axes of the score plot are PLSR Component 1 (PC1) and PLSR Component 2 (PC2). These components are the axes of linear regressions used in the model and consist of contributions from many of the variables in the model. Each sample plots with a score, or value, along the principal component; samples that plot near each other are spectrally, and thus compositionally, similar. Ideally, the

provenance groups would plot in mutually exclusive spaces on this diagram, suggesting that their differences in composition can be completely resolved.

In detail, modeling was performed using the program The Unscrambler®. The NIPALS algorithm was applied with 9 PCs; no weighting was applied to variables. The model is mean-centered.

Once a PLSR model has been trained, the model is applied to the training spectra, predicting whether each sample in the training suite is from the location being tested for or from any other location. The predicted value should be close to 1 for stones from the location being examined, and should be close to 0 for stones not from that location (Fig. 3B). While 0.5 is the midpoint between the two target values of the provenance variable, if the distributions of the provenance variable values for the two populations (i.e., being from the location or not) are not symmetric and identical distributions, the midpoint may not provide a good cutoff point, or Value of Apparent Distinction (VAD), for predicting whether the stone is from the location being considered. This second step of setting the VAD is illustrated in Figure 3B, where the calculated provenance predicted values are plotted for samples used to train the Kenya ruby model. Note that the Kenya samples, in closed diamonds, have provenance variables greater than the VAD of 0.227 and the samples from all other locations, in open squares, have provenance variables less than the VAD.

In considering the determination of a good VAD, the sample distributions of the predicted values were often not symmetric and were quite variable (Fig. 4). Hence, a nonparametric kernel density estimator (Silverman, 1986) was chosen to obtain a density estimate of the predicted values for each group (Fig. 4). Kernel densities depend on the selection of a bandwidth parameter; numerous bandwidth selection strategies were investigated by running simulations to determine how well the kernel density estimators using various bandwidth selection methods reproduced the original distribution. The data used for this process simulated a normal distribution, but specifically targeted the sample sizes and variances present in the training data set. Visually, the biased cross-validation method worked the best, giving fewer false multimode densities. At its worst, use of this bandwidth selection method tracked the original distribution

similar to other methods, and at its best it tracked it substantially better. Generally, all the methods resulted in similar VADs, even when they resulted in substantially different looking densities.

For any specified VAD, the small area under the right density curve occurring to the left of the VAD is the estimate of how often misclassification would occur for the population of rubies from the location under consideration (dark shaded area in Fig. 4). Similarly, the area under the left density occurring to the right of the VAD estimates the misclassification rate for the population not from the location (light shaded area in Fig. 4). Once density estimates were obtained, numerical root-finding and numerical integration of the densities were used to solve for the VAD between 0 and 1 that minimized the sum of these two misclassification rates. For the ruby study, VADs ranged from 0.141 to 0.512, and for the sapphire study they ranged from 0.359 to 0.604.

In the validation step, the model uses the regression coefficients calculated in the training step to predict Y-values (provenance variables near 0 and 1) from a new set of validation spectra from samples not used in the training. For instance, in Figure 3C, provenance variables are calculated for five samples, two of which are greater than the VAD of 0.227 and three of which are below the VAD. Samples with predicted provenance variables greater than the VAD are assigned to the location being examined; those with predicted provenance variables less than the VAD are assigned to the group of all other locations.

The last step is to calculate the success rate of the model (Fig. 3D). This is illustrated for the Kenya ruby model in Figure 5. Spectra from Kenya samples all have predicted provenance variables greater than the VAD of 0.227; they are all correctly identified as Kenyan. All but one of the other spectra, from India, Greenland, Madagascar, Myanmar, Pakistan, Tanzania, and the USA, have predicted provenance variables less than 0.227; they are correctly identified as belonging to this large group. One sample in this group was incorrectly identified as Kenyan. Because 163 of 164 samples were correctly identified, the model has a 99.4 % success rate.

In this study, a series of PLSR models are used to determine provenance. Each model compares spectra from one location of interest (provenance integer variable = 1) to spectra from all other models (provenance integer variable = 0). Separate matching algorithms for rubies and sapphires were developed by constructing a sequential arrangement of models that separate spectra from a single locality from those from all other localities. These are illustrated by flow charts and corresponding score and validation plots (Figs. 6 & 7 for sapphires and Figs. 7 & 8 for rubies). A similar method was developed by Multari et al. (2010) for the identification of bacterial strains using LIBS data and by McMillan et al. (2012) for identification of individual limestone beds. In the algorithms, the first model (India in the sapphire model) uses prediction spectra from all locations. Each spectrum is determined to be similar either to India or to all the other localities in the model. The spectra from this first locality are then removed from all subsequent models. Thus, the second model sapphire (Sapphire Branch) does not include any spectra from India. The number of spectra in each model decreases and the spectra become more chemically similar to each other as one moves through the flow chart. In some cases, it was not possible to separate spectra from a single country, and the algorithm branches, as in the Sapphire Branch model. The algorithms are discussed in detail in the Results section.

To determine the order of the models, average spectra of each country were visually inspected for unique characteristics. Localities with the most distinct spectra were separated first, allowing the model to focus on small spectral difference between chemically similar samples later in the algorithm. By reducing the number of samples for similar spectra, subtle differences were emphasized in the later models.

Success rates can be determined for individual models as well as the entire algorithm. Table 2 reports the percentage of correct predictions for each model. These success rates reflect only the models that predict provenance and not models that cause branching. Thus, a second method for calculating success rates counts the total number of correct predictions for all models as a percent of the total number of spectra. These overall success rates are shown in Figures 6 & 8.

Results

Sapphire Provenance Algorithm. The sapphire provenance algorithm is illustrated in a flow chart in Figure 6; the score and validation plots are presented in Figure 7. Results are shown in Table 2. Score plots show the variation in composition in PC1-PC2 space for the samples used in the model. Ideally, the two groups of samples should not overlap, but in practice, groups with some overlap can still be modeled with high success. Validation plots present the calculated provenance variables for samples used to validate the model. Ideally, all samples in the provenance being modeled should have values greater than or equal to the VAD; samples from all other locations should have values less than the VAD. The success rate of the model is calculated as the percent of samples whose provenance was correctly identified.

The first locality to be identified in the sapphire algorithm is Psudipada, India. This model is 100% successful, correctly identifying the provenance of 118 of 118 samples. The second model, Sapphire Branch, separates samples from Afghanistan and Vietnam from the remaining localities (Tanzania, Australia, Madagascar, and USA). It successfully predicts the origin of 102 of 103 samples (99.0% success). The Afghanistan model compares samples from Afghanistan to those from Vietnam. This model is 96.7% successful, correctly identifying 29 of 30 samples. Continuing on the main stem of the algorithm, the Tanzania model identifies samples as being from Tanzania (Songea or Umba Valley) or as being from Australia, Madagascar, or the USA. It has a success rate of 97.3% (72 of 74 samples correctly identified). The Tanzania Deposit model attempts to predict whether samples are from Songea or the Umba Valley in Tanzania. It is not successful, with only 50% success rate. Note that this study only has ten samples from Songea and five from Umba Valley. This is an excellent example of the importance of having a large number of samples when attempting to determine provenance by chemical fingerprinting.

The Australia model compares sapphires from Australia to those from Madagascar and Montana (USA). The model is 100% successful, with a total of 66 spectra. The Madagascar model compares sapphires from Madagascar and Montana, USA. It, too, is 100% successful; all 45 spectra were correctly identified. The final model, Madagascar Deposit, distinguishes between blue and white sapphires from Madagascar. The model correctly identifies 29 of 30 samples (96.7% success).

Ruby Provenance Algorithm. Results for the ruby provenance algorithm are presented in a flow chart (Fig. 8), in score and validation plots (Fig. 9), and in Table 2. The ruby algorithm contains complexity not encountered in the sapphire algorithm, probably because the chemical variation within rubies is less than that within sapphires. The first model distinguishes rubies from Kenya from all other localities. It is 98.2% successful, correctly determining the provenance of 161 of 164 samples. Rubies from Mysore, India, that were obtained free from matrix are separated in the second model, India I. This model correctly predicts the locality of 137 of 149 samples (91.9% success). The algorithm bifurcates in the next model, Ruby Branch I. It separates the remaining samples into two groups: 1) USA (Wyoming), Pakistan, and Tanzania; and 2) Greenland, Madagascar, Myanmar, and India (samples in matrix). This model is 99.3% successful, correctly identifying 138 of 139 samples.

Rubies from Wyoming, Pakistan, and Tanzania comprise the algorithm on the right-hand side of the flow chart (Fig. 8). These sets are chemically similar; however, it is possible to separate them with high success rates by allowing the Wyoming samples to be present in the Ruby Branch II model, but not interpreting their provenance assignments. In the Ruby Branch II model, two groups are determined: 1) Wyoming & Pakistan, and 2) Wyoming & Tanzania. This model is 92.9% successful and correctly identified 26 of 28 samples (the Wyoming samples were not counted). The Pakistan model then discriminates between Wyoming and Pakistan rubies, at a 100% success rate (26 of 26 correct provenance determinations). Similarly, the Wyoming model discriminates between Wyoming and Tanzanian rubies. It also is 100% successful, identifying 42 of 42 samples correctly. Thus, Wyoming rubies can be identified through two routes: the Pakistan model and the Wyoming model. The final model on the right-hand side of the flow chart, the Tanzania Deposit model, discriminates between two Tanzanian regions: Winza and Longido. The model was 86.4% successful and correctly determined the provenance of 19 of 22 samples. This demonstrates that ruby deposits approximately 450 km apart have sufficiently distinct LIBS signatures to be identified with a high success rate.

The left-hand side of the ruby algorithm flow chart concerns rubies from Greenland, Madagascar, Myanmar, and India (Mysore, samples in matrix). The Greenland model is 98.9% successful, positively identifying 92 of 93 samples. The Madagascar model separates samples

from Madgascar at a 97.3% success rate; 73 of 75 samples were positively identified. The Madagascar Deposit model then attempts to discriminate between a set of Madagascar rubies free of matrix and rubies in matrix (amphibolite and biotite gneiss). This model is 100% successful, correctly identifying 32 of 32 samples. The Myanmar model is able to positively identify rubies of Myanmar at a success rate of 86.7% (39 of 45 samples). The final model is the India model, which discriminates Mysore rubies in fuchsite matrix from Mysore rubies in biotite gneiss matrix. This model has a success rate of only 83.3%, correctly identifying 25 of 30 rubies.

Success rates are reported in this work in two ways. First, the success rates for each individual model are shown on Figures 6 & 8 and in Table 2. This calculation simply considers the number of samples involved in the validation model and represents the percentage of samples for which provenance was correctly identified. Success rates for individual models range from 100% to 50%. Although the method of expressing success in terms of individual models is straightforward, it does not take into account every model in the algorithm. For instance, the sapphire algorithm branches in a model called Sapphire Branch. The ruby algorithm has two such branches; the success of all models subsequent to the branches depends on the success of the branching model. An alternative method is to calculate the success rate of the total algorithm as the percent of total correct provenance determinations relative to the total number of determinations. Using this method, the overall success rates for sapphire provenance determination are 98.9% for determination of country of origin and 97.9% for determination of deposit of origin. Overall success rates for rubies are similar: 96.0% for country of origin and 95.4% for deposit of origin.

Necessity of large sample set. These results demonstrate the need for a large sample set when attempting to determine mineral provenance. Table 2 lists success rates for provenance determination for each locality and the number of samples used in each model. Success rates are > 85%, with two exceptions. One exception is distinguishing between Mysore rubies in fuchsite matrix and Mysore rubies in biotite gneiss matrix. The success rate for this model is only 83.3% success rate, suggesting that the two groups of rubies are similar in composition. In algorithms such as these, materials near the end of the flow chart are chemically similar to each other and

thus more difficult to distinguish from each other. The other failure in this study is the ability to distinguish between the Tanzania sapphire localities Songea and Umba Valley. Note, however, that the sample set contains only ten samples from Songea and five from Umba Valley (Table 1). Because half of the samples were used to calibrate the model and half to verify the model, only eight samples were used in the prediction step, and only half of them were correctly identified (Table 2). This suggests that many samples are required to fully capture the chemical variability of a deposit. Clearly, the five samples from Umba Valley were not sufficient. Success rates for localities with at least 30 samples range between 83.3% and 100%.

Discussion

Provenance determination is a difficult issue to solve. Issues such as the large number of samples necessary to conduct rigorous determinations, the cost and complexity of analytical instrumentation, and the skill and experience required to make high-quality gemological observations are obstacles in the attempt to provide quantifiable determinations of country of origin or deposit of origin. This study demonstrates that chemometric analysis of LIBS spectra from a large number of samples makes accurate, rapid provenance determination possible, with success rates generally higher than 95%. In addition, LIBS technology is relatively inexpensive, rugged, portable, and simple to use. The analyst needs no intense training in spectroscopy or gemology.

The method described here could be fully automated and implemented at critical points in the gem stream, such as border control stations, ports of entry, gem studios, and galleries. Prior to widespread application, several issues need further research. The effect of heat and chemical treatment on provenance analysis with LIBS spectra and multivariate analysis has not been studied; this pilot study used only untreated, uncut samples. Sample library size is critical to the success of any gem provenance project; ultimately, one would like to build a model with at least 30 samples from every important corundum deposit. In addition, the number and spatial distribution of LIBS analytical spots on cut gems would be limited to a few spots on the stone's girdle; the model would have to be tested using these analytical protocols. It is possible that more specimens from each locality would be necessary to capture the compositional diversity of each deposit with fewer spectra per sample.

One goal of provenance determination is to gain an understanding of the chemical differences between samples from various locations. In chemometric analysis, loading vectors are used to determine which chemical elements are important in discriminating between two groups of samples. Loading vectors are the contribution of each variable (wavelength) to the linear regression. If a linear regression trends in the direction of increasing value for a certain wavelength variable, for instance increasing intensity of a Ca wavelength, that wavelength, and thus the concentration of the emitting element, will have a strong positive influence on the linear regression and a large positive loading vector. In contrast, if the regression trends in a direction of decreasing value for a wavelength, that wavelength (and element) will have a negative loading vector. Loading vectors for the models were qualitatively inspected to discern which elements have high positive loading vectors for the deposit being considered in the model (Table 2). Elements with very high loading vectors were considered to have strong influence; those with high loading vectors were considered to have lesser but significant influence. Elements with moderate and lower loading vectors are not listed in Table 2; they exert lesser influence on the model and are thus less significant for discrimination purposes. These elements are important in discriminating between corundum deposits, in approximate order of decreasing importance: Ca, Zr, Fe, Ba, Mg, Ti, Sr, Na, V, Si, Cr, H, C, and Li. Some of these elements may reside in inclusions in the analyzed samples; however, the inclusions must have been sufficiently abundant for them to contribute significantly to the loading vectors. The chemical differences between corundum deposits are not simple (Table 2); it is the ratios of many elements that comprise the unique chemical signature of each deposit. The signature is too complex to be discerned by traditional geochemical techniques (Abduriyim and Kitawaki, 2006a, 2006b; Calligaro et al., 1999; Muhlmeister et al., 1998; Peucat et al, 2007; Pornwilard et al., 2011; Rankin et al., 2003; Saminpanya et al., 2003) because it is inherited from subtle compositional differences in the host rock during metamorphism.

It is interesting that most of the elements listed in Table 2 are not chromophores. This study indicates that provenance determination using chemometric analysis of LIBS spectra is independent of the color of the stones. Some of the sample sets analyzed in this paper have consistent colors; however, three of the sapphire sets have a wide variety of colors (Table 1).

The Australian, Umba Valley, and Montana sample sets all exhibit multiple colors. However, the Australian and Montana (USA) models are 100% successful, indicating that the models are independent of color.

The ideal method for provenance determination would compare spectra of unknown samples to a database that contains spectra from every known ruby or sapphire deposit. Obviously, such a sample collection is prohibited by time and funding. The finite size of our database opens the question of how samples from localities other than those present in the calibration are handled in the algorithms. There are two possibilities. First, a sample may be sufficiently similar to samples from one of the localities in the database that it is classified incorrectly. In some cases, such samples can be recognized by using chemometrics to compare the LIBS spectrum of the unknown sample to those of the known samples. If the unknown sample is an outlier, it is most likely from a different location. However, the possibility exists that the unknown deposit may actually be similar in composition to one of the localities in the calibration. In this case, the sample will be mis-identified; the best method to avoid this is to build a sample base that contains as many deposits as possible.

Alternatively, an unknown sample may be unlike any samples from any of the localities in the calibration. In this case, the sample will consistently be classified as a "0" and will be identified as one of the localities at the end of the flow charts (Montana for sapphires; Mysore or Tanzania for rubies; Figures 6 & 8). In this case, chemometrics (i.e., Principal Component Analysis) can be used to compare all of the spectra that are classified in the final localities; if the sample is an outlier, its provenance is most likely one not found in the calibration sample base.

If this method were to be used commercially, building the sample base would be a high priority. As new localities are added, the algorithms would need to be revised. The order in which the samples are considered is of primary concern, although there is more than one order that produces satisfactory results. The addition of any new locality requires development of an entirely new algorithm, because the most chemically unique samples are separated out first, leaving the more chemically similar samples at the end of the process. While the process of determining the order of the algorithm is time-consuming, it is not intrinsically difficult.

Implications

Provenance determination is an important aspect of gem values as well as a political and national security issue. Traditional methods for determining country of origin or deposit of origin include comparing gemological observations and/or trace concentrations using XRF, PIXE, and ICP-MS data, of unknown samples to those from known localities. Results have been promising but no unique solution has emerged for several reasons. First, the variety of color, inclusions, and unique mineralogical features is fairly large in each deposit and not all specimens from a deposit contain all of the identifying features. Second, trace element studies have used a relatively small number of variables (the number of elements analyzed), and in some cases, a relatively small number of samples.

Chemometric treatment of Laser-Induced Breakdown Spectroscopy (LIBS) spectra from sapphires and rubies is a new and successful method for provenance determination. The main advantages of LIBS are 1) ease of analysis; 2) rapidity of analysis; and 3) little or no required sample preparation. Because of this, it is possible to analyze many (i.e., hundreds) samples rapidly, creating a data base that closely represents the chemical variation in each locality. LIBS instruments could be operated at strategic points such as border crossings, international port facilities, and even military outposts to determine the provenance of gems and conflict minerals in real time. Instruments could also be used in gem galleries and studios to insure that the gem provenance is accurately reported.

The ability to obtain spectra from large sample sets and to analyze the large resulting data set provides a new direction for mineral analysis and research. For instance, the process described in this paper could be applied to correlation of strata using mineral compositions. Examples include correlating ash-flow tuffs by LIBS analysis of phenocryst minerals, or correlating clastic sedimentary rocks through LIBS analysis of the heavy minerals zircon, tourmaline, and rutile. In this case, one would expect to see several populations of each phase; however, analysis of hundreds or thousands of grains would provide an adequate statistical base for estimation of the population distribution of each group. Such a project is daunting in scale using traditional technology.

559

560 This study demonstrates that quantifiable, high-precision provenance determination is possible
561 using technology that is durable, portable, relatively inexpensive, and operable by a range of
562 end-users. Overall, the sapphire algorithm correctly identified the provenance in 432 of 437
563 cases (98.9%) for country of origin and in 465 of 475 cases for deposit of origin (97.9%).
564 Success rates for rubies are similar. Overall, 759 of 791 ruby spectra in the algorithm were
565 correctly identified (96.0%) at the country of origin level; the overall deposit of origin success
566 rate is 95.4% (835 of 875). The application of chemometric analysis of LIBS spectra could bring
567 new transparency to the gem industry.

568

569 **Acknowledgements**

570 The authors gratefully acknowledge Tristan Likes and Karly Baughn for logistical support and
571 Materialytics, LLC, for funding. We also thank M. Darby Dyar and an anonymous reviewer for
572 insightful comments that improved the quality of this contribution to mineral science.

573

574

575 **References Cited**

576 Abduriyim, A., and Kitawaki, H. (2006a) Applications of laser ablation-inductively coupled
577 plasma-mass spectrometry (LA-ICP-MS) to gemology. Gems & Gemology, 42, 98-118.
578
579 Abduriyim, A., and Kitawaki, H. (2006b) Determination of the origin of blue sapphire using
580 laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). Journal of
581 Gemmology, 30, 23-36.
582
583 Calligaro, T. Piorot, J.-P., and Querré, G. (1999) Trace element fingerprinting of jewelry rubies
584 by external beam PIXE. Nuclear Instruments and Methods in Physics Research B, 150, 628-634.
585
586 Esbensen, K.H. (2004) Mutlivariate Data Analysis-in practice: An introduction to multivariate
587 data analysis and experimental design, 598 p. CAMO process AS, Oslo.
588
589 Cremers, D.A., and Radziemski, L.J. (2006) Handbook of Laser-Induced Breakdown
590 Spectroscopy, 283 p. John Wiley & Sons. Ltd., Chichester.
591
592 Devouard, B., and Notari, F. (2009) The identification of faceted gemstones; from the naked eye
593 to laboratory. Elements, 5, 163-168.
594

Giuliani, G., Fallick, A.E., Garnier, V., France-Lanord, C., Ohnenstetter, D., and Schwarz, D. (2005) Oxygen isotope composition as a tracer for the origins of rubies and sapphires. Geology, 33, 249-252.

Giuliani, G., Fallick, Al, Ohnenstetter, D., and Pegere, G. (2009) Oxygen isotopes composition of sapphires from the French Massif Central: implications for the origin of gem corundum in basaltic fields. Miner Deposita, 44, 221-231.

Gray, A.L. (1985) Solid sample introduction by laser ablation for inductively coupled plasma source mass spectrometry. Analyst, 110, 551-556.

Gübelin, E.J. (1953) Inclusion as a Means of Gemstone Identification, 220 p. Gemological Institute of America, Los Angeles.

Gübelin, E.J., and Koivula, J.I. (1986) Photoatlas of Inclusions in Gemstones, 531 p. ABC Edition, Zurich, Switzerland.

Guillong, M., and Günther, D. (2001) Quasi 'non-destructive' laser ablation-inductively coupled plasma-mass spectrometry fingerprinting of sapphires. Spectrochimica Acta Part B, 56, 1219-1231.

Günther, D., and Kane, R.E. (1999) Laser ablation-inductively coupled plasma-mass spectrometry: A new way of analyzing gemstones. Gems & Gemology, 35, 160-161.

Hughes, R.W. (1997) Ruby & Sapphire, 511 p. RWH Publishing, Boulder, CO.

Liddicoat, R.T. (1990) The country of origin question. Gems & Gemmology, 26, 247.

McMillan, N.J., Montoya, C., and Chesner, W.H. (2012) Correlation of limestone beds using laser-induced breakdown spectroscopy and chemometric analysis. Applied Optics, 51, B213-B222.

Muhlmeister, S., Fritsch, E., Shigley, J.E., Devouard, B., and Laurs, B.M. (1998) Separating natural and synthetic rubies on the basis of trace-element chemistry. Gems and Gemology, 43, 80-101.

Multari, R.A., Cremers, D.A., Dupre, J.M., and Gustafson, J.E. (2010) The use of laser-induced breakdown spectroscopy (LIBS) for distinguishing between bacterial pathogen species and strains. Applied Spectroscopy, 64, 750-759.

Peucat, J.J., Ruffault, P., Fritsch, E., Bouhnik-Le Coz, M., Simonet, C,. and Lasnier, B. (2007) Ga/Mg ratio as a new geochemical tool to differentiate magmatic from metamorphic blue sapphires. Lithos, 98, 261-274.

Pornwilard, M.-M., Hansawek, R., Shiowatana, J., and Siripinyanond, A. (2011) Geographical origin classification of gem corundum using elemental fingerprint analysis by laser ablation inductively coupled plasma mass spectrometry. International Journal of Mass Spectrometry, 306, 57-62.

Rankin, A.H., Greenwood, J., and Hargreaves, D. (2003) Chemical fingerprinting of some East African gem rubies by laser ablation ICP-MS. Journal of Gemmology, 28, 473-482.

Rossman, G.R. (2009) The geochemistry of gems and its relevance to gemology: Different traces, different prices. Elements, 5, 159-162.

Saminpanya, S., Manning, D.A.C., Droop, G.T.R., and Henderson, C.M.B. (2003) Trace elements in Thai gem corundums. Journal of Gemmology, 28, 399-415.

Sansonetti, J.E., Martin, W.C., and Young, S.L., (2005), Handbook of Basic Atomic Spectroscopic Data (version 1.1.2) (Online) Available: http://physics.nist.gov/Handbook (accessed March 23, 2003). National Institute of Standards and Technology, Gaithersburg, Maryland.

Schwarz, D., and Schmetzer, K (2001) Rubies from the Vatomandry area, eastern Madagascar. Journal of Gemmology, 27, 409-146.

Sharp, Z.D. (1990) A laser-based microanalytical method for the in situ determination of oxygen isotope ratios of silicate and oxides. Geochimica et Cosmochimica Acta, 54, 1353-1357.

Shor, R., and Weldon, R. (2009) Ruby and sapphire production and distribution: A quarter century of change. Gems & Gemology, 45, 236-259.

Silverman, B. W. (1986) Density Estimation for Statistics and Data Analysis, 175 p. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

Sutherland F.L., and Abduriyim A. (2009) Geographic typing of gem corundum: a test case from Australia. Journal of Gemmology, 31, 203-210.

673 Sutherland, F.L., Zaw, K., Mefre, S., Giuliani, G., Fallick, A.E., Graham, I.T., and Webb, G.B.
674 (2009) Gem-corundum megacrysts from east Australian basalt fields: trace elements, oxygen
675 isotopes, and origins. Australian Journal of Earth Sciences, 56, 1003-1022.
676
677 U.S. Department of State (2008a) 2008 Human Rights Report: Burma. (accessed January 2015)
678 http://www.state.gov/g/drl/rls/hrrpt/2008/eap/119035.htm.
679 U.S. Department of State (2008b) H.R. 3890 (110th): Tom Lantos Block Burmese JADE
680 (Junta's Anti-Democratic Efforts) Act of 2008. (accessed January 2015)
681 http://www.govtrack.us/congress/bills/110/hr3890.
682
683 Vysotsky, S.V., Yakovenko, V.V., Ignat'ev, A.V., and Karabstov, A.A. (2009) The oxygen
684 isotopic composition as an indicator of the genesis of "basaltic" corundum. Russian Journal of
685 Pacific Geology, 3, 64-48.
686
687 Ward, F. (1995) Rubies and Sapphires, 64 p. Gem Book Publishers, Bethesda, Maryland.
688
689 Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics.
690 Chemometrics and Intelligent Laboratory Systems, 58, 109-130.
691
692 Yui, T-F., Zaw, K., and Limtrakun, P. (2003) Oxygen isotope composition of the Denchai
693 sapphire, Thailand: a clue to its enigmatic origin. Lithos, 67, 153-161.
694
695 Yui, T-F., Wu, C-M., Limtrakun, P., Sricharn, W., and Boonsoong, A. (2006) Oxygen isotope
696 studies on placer sapphire and ruby in the Chanthaburi-Trat alkali basaltic gemfield, Thailand.
697 Lithos, 86, 197-211.
698
699 Yui, T-F., Zaw, K, and Wu, C-M. (2008) A preliminary stable isotope study on Mogok Ruby,
700 Myanmar. Ore Geology Reviews, 34, 192-199.
701

Figure Captions

Figure 1. LIBS analysis starts with a pulsed laser (B, with control box A), in this case a Nd-YAG laser. A digital camera (C) allows one to target and photograph specific areas on samples. A burst of laser light is focused in the sample by lens D, causing atoms to ablate and burn in a short-lived plasma, or burning gas. Electrons in this high-temperature plasma are excited into higher-energy orbitals, and give off energy in the form of light as they decay into lower-energy orbitals during plasma cooling. This spectrum of light (different wavelengths for each element) is collected by optic fiber and diffracted by a spectrometer (E) and recorded by a computer (F).

Figure 2. Representative ruby and sapphire LIBS spectra (averages of 30 spectra). Major elemental emission lines are identified.

Figure 3. Partial Least Squares Regression model process. Half of the spectra from every locality are used for calibration; the other half are used for validation. The spectra are divided into two groups. One group contains spectra from the locality of interest; the other group contains spectra from all other localities. A. Calibration. In calibration, the spectra are regressed against the integers "1" (locality of interest) or "0" (all other localities) B. Determination of the VAD (Value of Apparent Distinction), the value of the provenance variable used to distinguish between the two groups, using provenance variables for each sample calculated during the calibration stage. C. Prediction of provenance variables for the validation samples. D. Calculation of success rate.

Figure 4. Kernel density estimates of the population of predicted values for rubies from Kenya (on the right) and of the population of predicted values for rubies from other mines (on the left). The sample predicted values are indicated by the vertical lines below the density curves – dashed vertical lines are for predicted values for rubies from the Kenya mine; solid lines are for rubies from other mines. The VAD (Value of Apparent Distinction) that minimizes the total crossover areas of the two densities is indicated by the dashed vertical line.

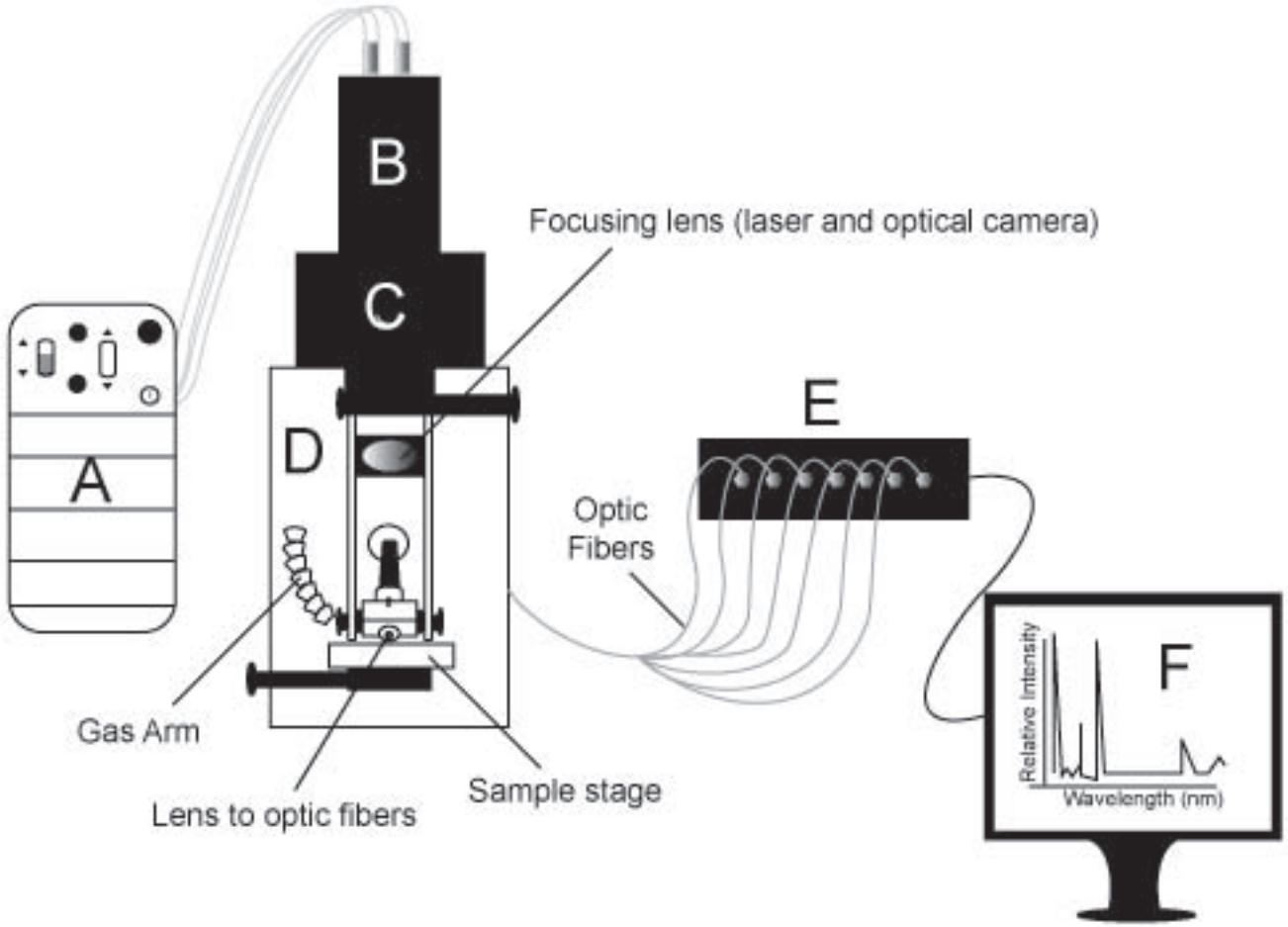
Figure 5. The results of the validation step are shown for the Kenya ruby model. The predicted provenance value (Y-axis) is shown for each sample in the model. Rubies from Kenya have provenance variables greater than the Value of Apparent Distinction (0.29 for this model) and thus are classified as Kenyan. The other samples, from India, Greenland, Madagascar, Myanmar, Pakistan, Tanzania, and USA, yield provenance variables less than 0.29, and are classified as belonging to this group of countries. One sample is incorrectly classified, yielding a success rate of 99.4%.

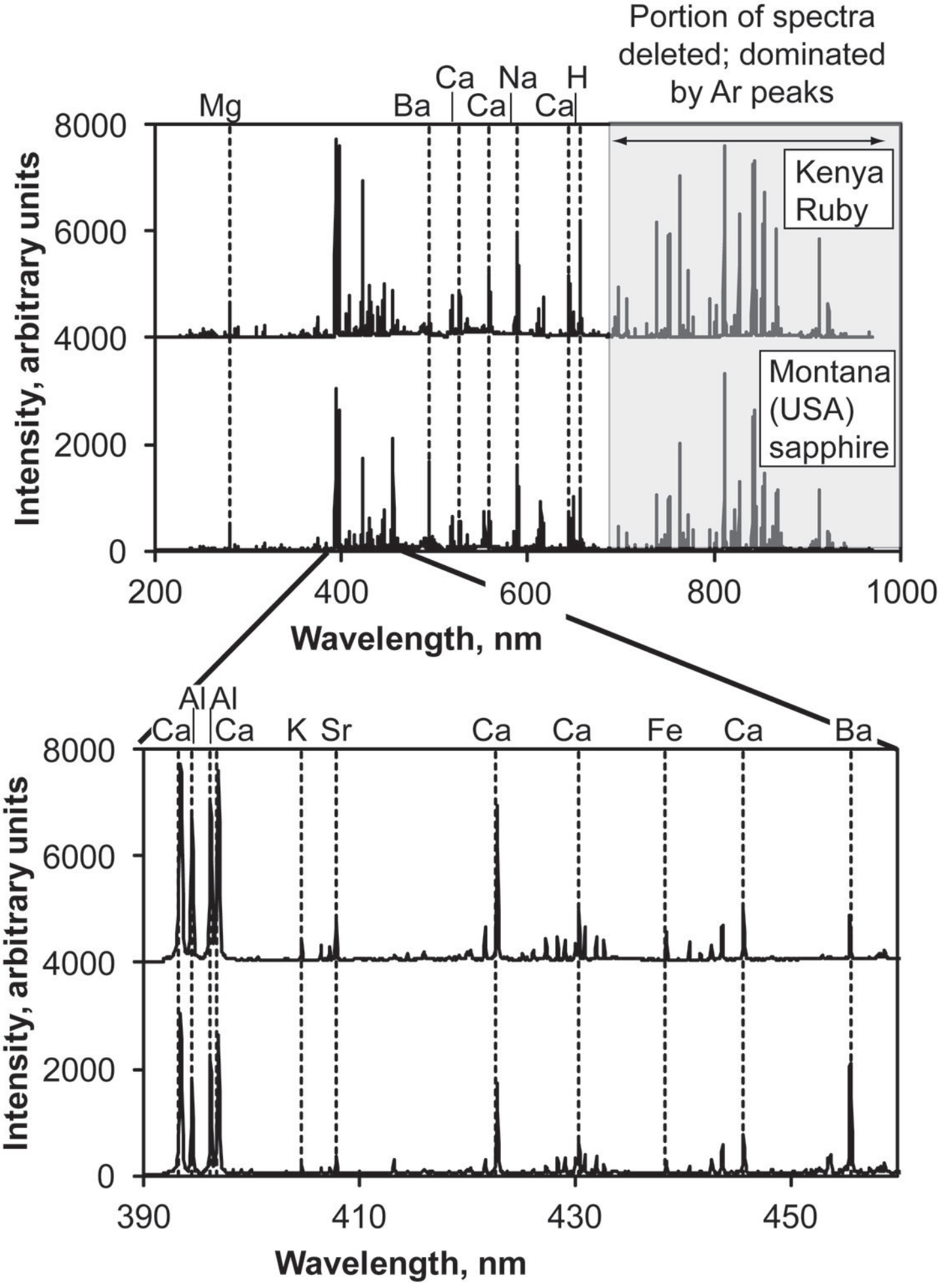
Figure 6. The matching algorithm for sapphires consists of 8 models. The model correctly predicts the country of origin for 98.9% of the validation spectra and the deposit of origin for 97.9% of the validation spectra.

Figure 7. Score plots and validation plots for PLSR (Partial Least Squares Regression) models in the sapphire matching algorithm.

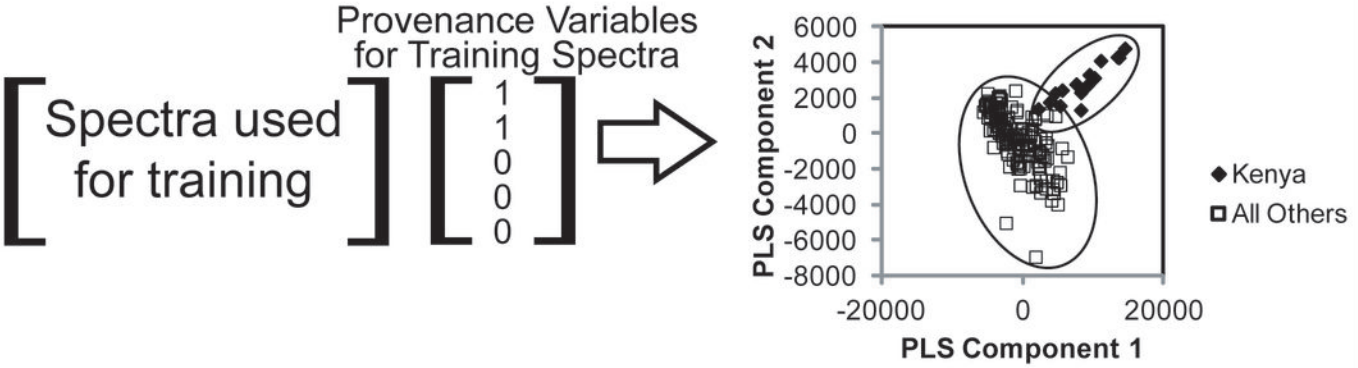
Figure 8. The matching algorithm for rubies consists of 12 models. The model correctly predicts the country of origin for 96.0% of the validation spectra and the deposit of origin for 95.4% of the validation spectra.

Figure 9. Score plots and validation plots for PLSR (Partial Least Squares Regression) models in the ruby matching algorithm.

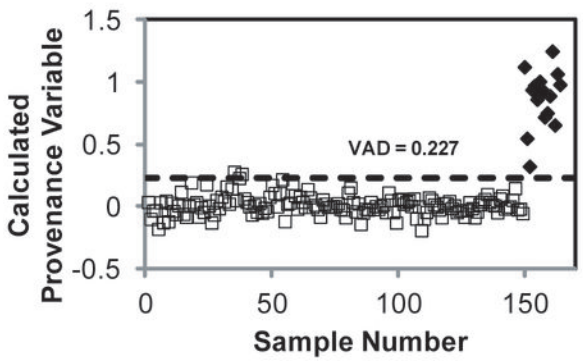

B
C
A
D
E
F
Focusing lens (laser and optical camera)
Optic
Fibers
Gas Arm
Lens to optic fibers
Sample stage
Relative Intensity
Wavelength (nm)


Portion of spectra deleted; dominated by Ar peaks
Mg
Ba
Ca
Ca
Na
Ca
H
Kenya Ruby
Montana (USA) sapphire
Intensity, arbitrary units
8000
6000
4000
2000
0
200
400
600
800
1000
Wavelength, nm
Al
Al
Ca
Ca
K
Sr
Ca
Ca
Fe
Ca
Ba
Intensity, arbitrary units
8000
6000
4000
2000
0
390
410
430
450
Wavelength, nm

A. Step One: Training the PLSR model

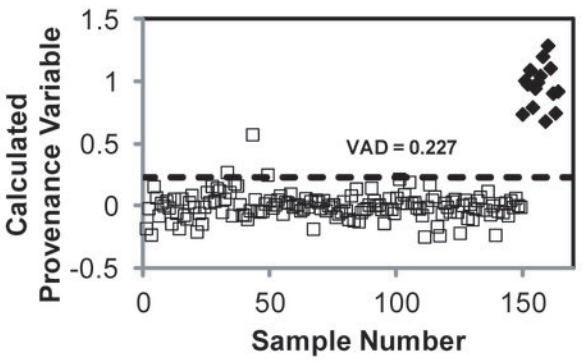


B. Step Two: Determine the Value of Apparent Distinction

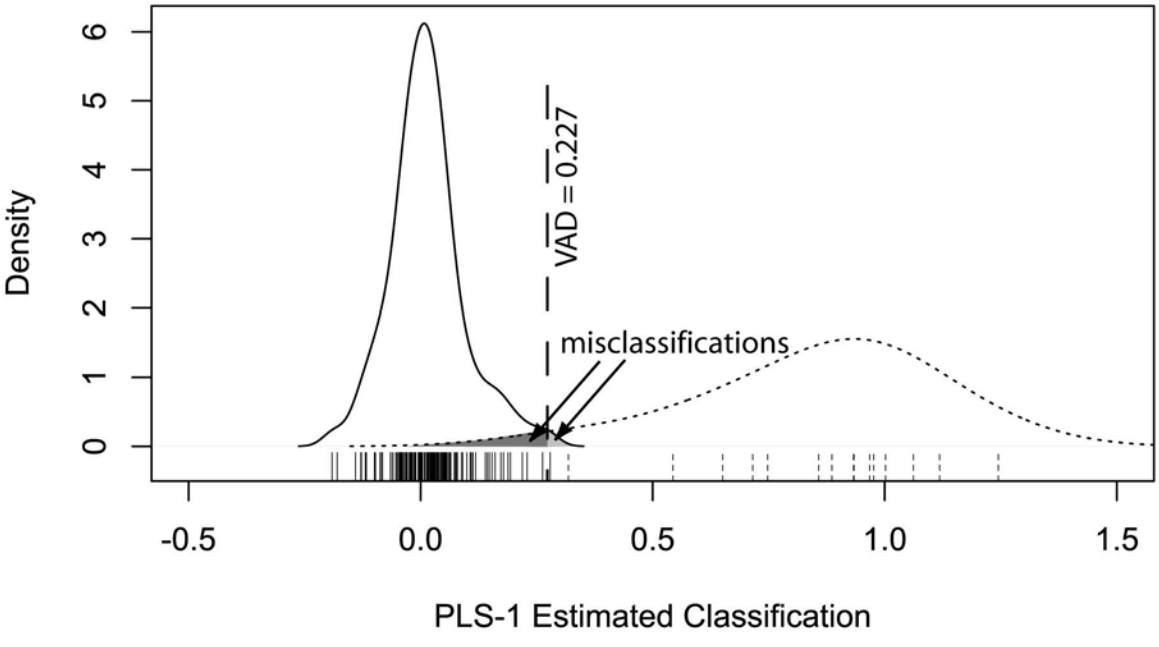


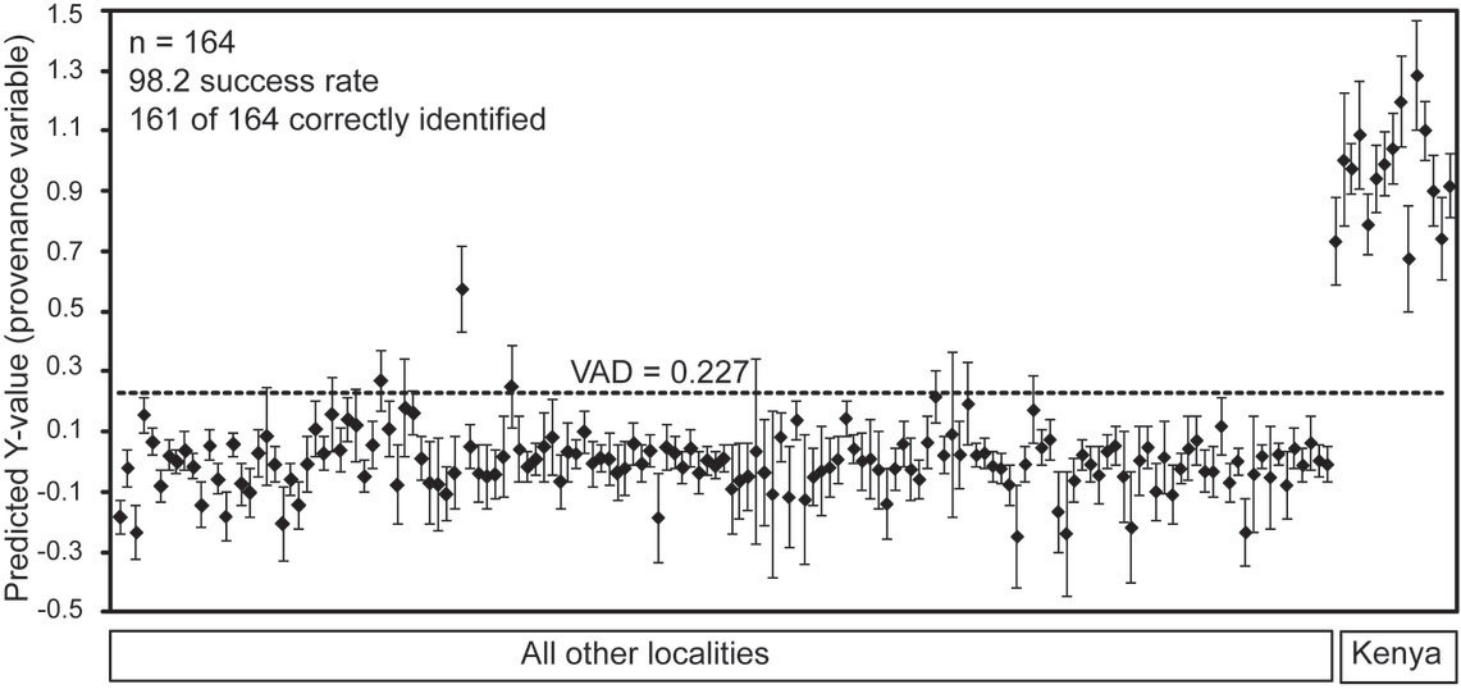
C. Step Three: Prediction of Provenance Variables for Validation Samples

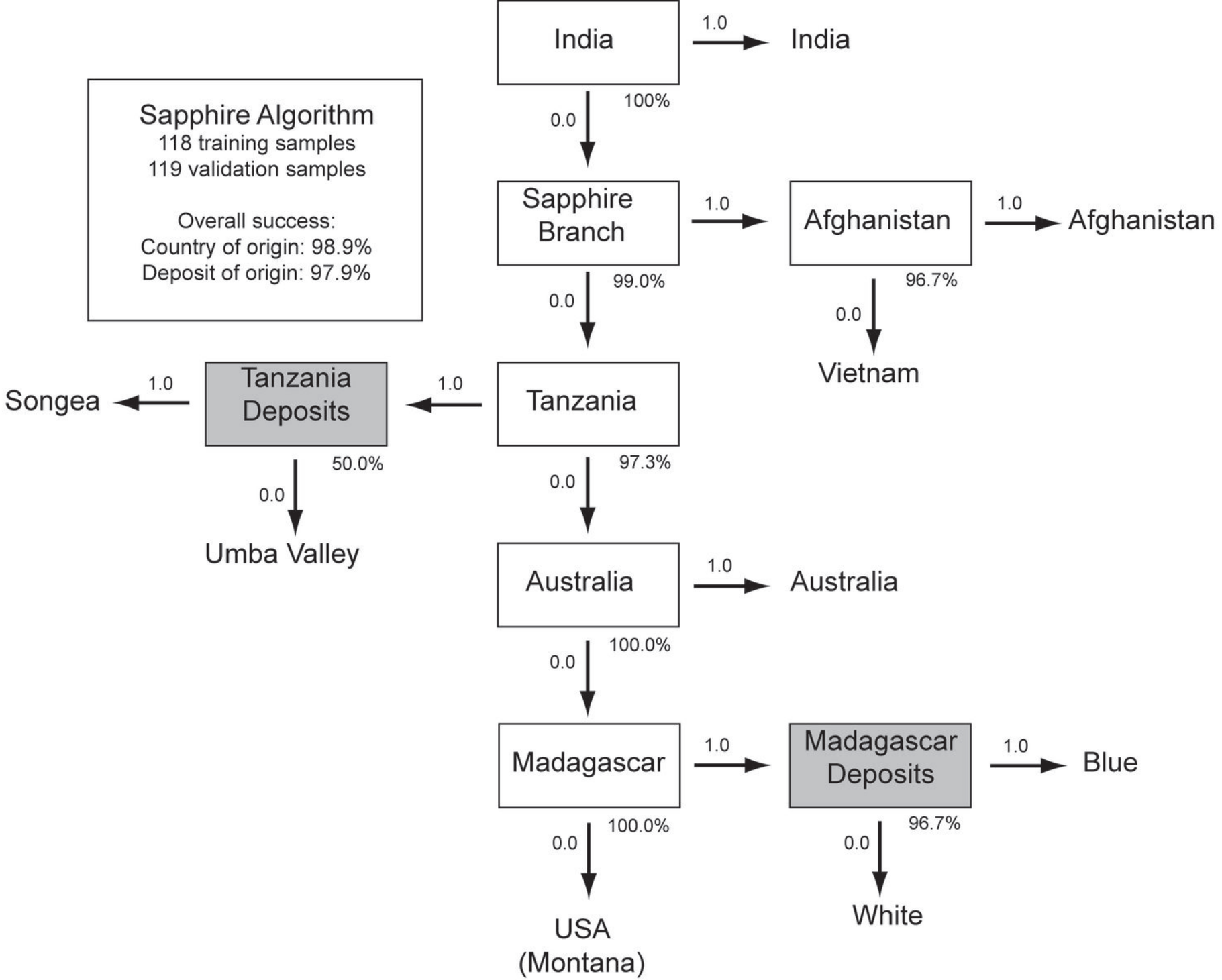
D. Step Four: Calculation of Success Rate

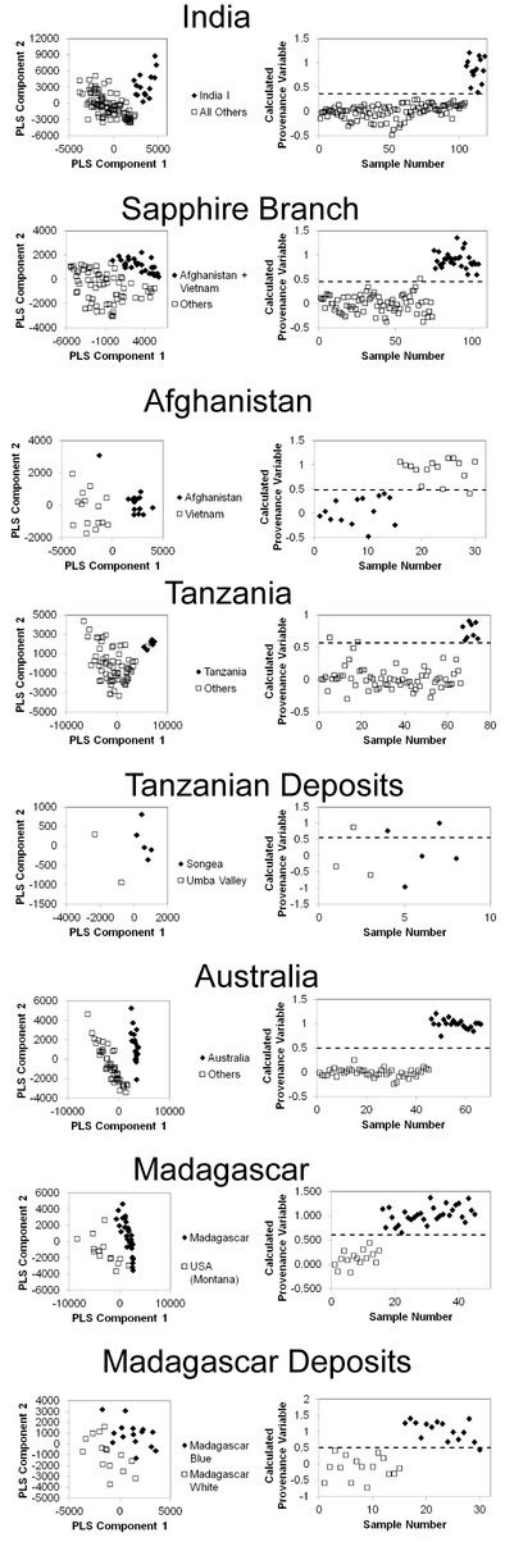


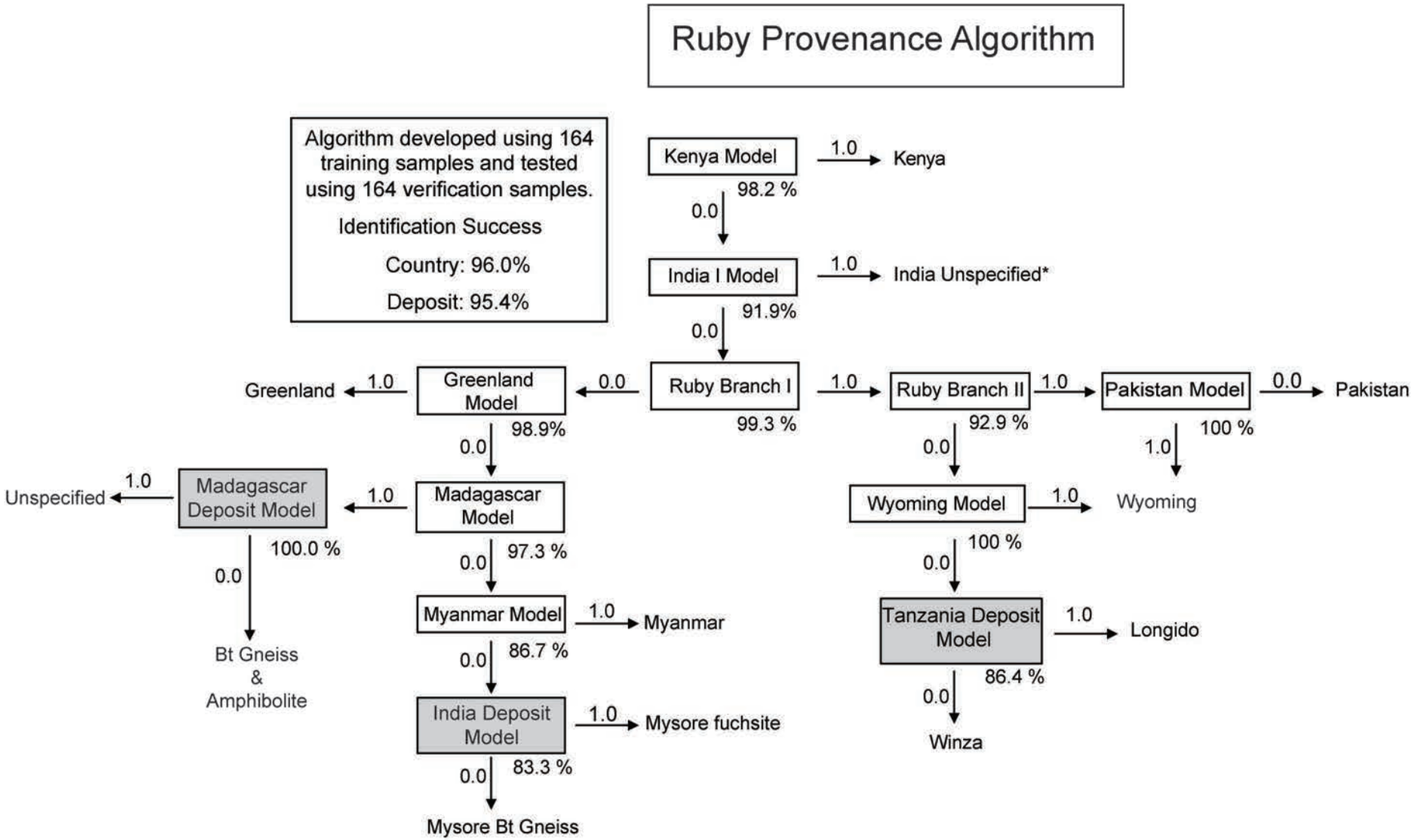
Correct Prediction
161 of 164 samples
98.2% success rate


Density
0
1
2
3
4
5
6
VAD = 0.227
misclassifications
-0.5
0.0
0.5
1.0
1.5
PLS-1 Estimated Classification

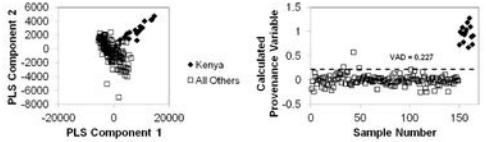

n = 164
98.2 success rate
161 of 164 correctly identified
Predicted Y-value (provenance variable)
1.5
1.3
1.1
0.9
0.7
0.5
0.3
0.1
-0.1
-0.3
-0.5
VAD = 0.227
All other localities
Kenya


Sapphire Algorithm
118 training samples
119 validation samples
Overall success:
Country of origin: 98.9%
Deposit of origin: 97.9%
India
1.0
India
100%
0.0
Sapphire Branch
1.0
Afghanistan
1.0
Afghanistan
99.0%
96.7%
0.0
0.0
Vietnam
Songea
1.0
Tanzania Deposits
1.0
Tanzania
50.0%
97.3%
0.0
0.0
Umba Valley
Australia
1.0
Australia
100.0%
0.0
Madagascar
1.0
Madagascar Deposits
1.0
Blue
100.0%
96.7%
0.0
0.0
USA (Montana)
White

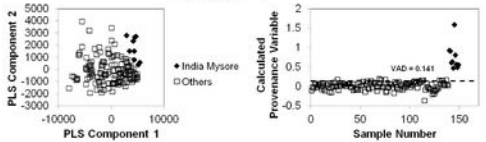

India
PLS Component 2
PLS Component 1
India I
All Others
Calculated Provenance Variable
Sample Number
Sapphire Branch
Afghanistan + Vietnam
Others
Afghanistan
Vietnam
Tanzania
Tanzanian Deposits
Songea
Umba Valley
Australia
Madagascar
USA (Montana)
Madagascar Deposits
Madagascar Blue
Madagascar White


Ruby Provenance Algorithm
Algorithm developed using 164 training samples and tested using 164 verification samples.
Identification Success
Country: 96.0%
Deposit: 95.4%
Kenya Model
1.0
Kenya
98.2 %
0.0
India I Model
1.0
India Unspecified*
91.9%
0.0
Ruby Branch I
99.3 %
1.0
Ruby Branch II
92.9 %
1.0
Pakistan Model
100 %
0.0
Pakistan
1.0
Wyoming Model
1.0
Wyoming
100 %
0.0
Tanzania Deposit Model
86.4 %
1.0
Longido
0.0
Winza
0.0
Greenland Model
98.9%
1.0
Greenland
0.0
Madagascar Model
97.3 %
1.0
Madagascar Deposit Model
100.0 %
1.0
Unspecified
0.0
Bt Gneiss & Amphibolite
Myanmar Model
86.7 %
1.0
Myanmar
India Deposit Model
83.3 %
1.0
Mysore fuchsite
Mysore Bt Gneiss

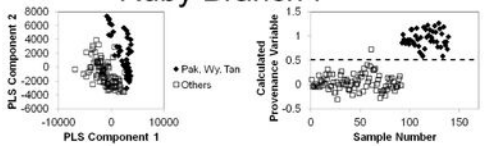
Kenya



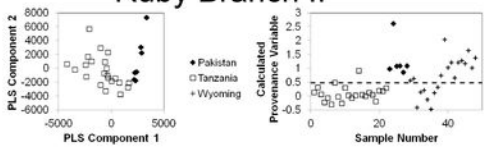
India 1



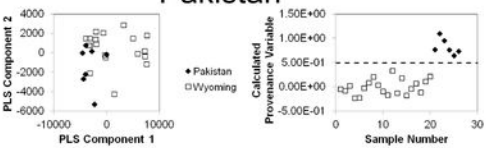
Ruby Branch I



Ruby Branch II

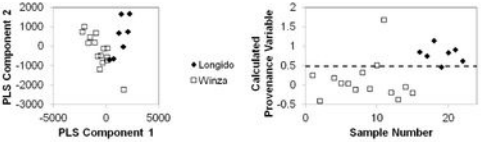


Pakistan

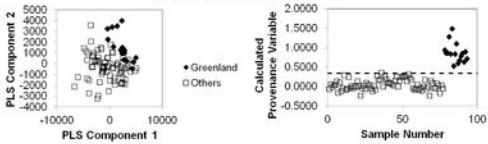


Wyoming

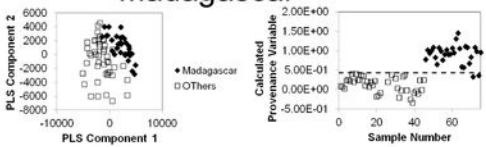
Tanzania Deposits



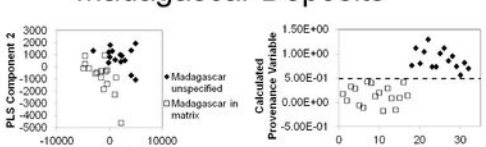
Greenland



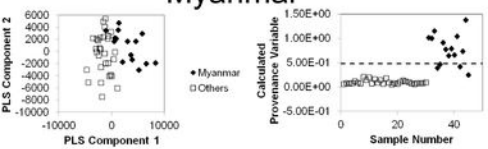
Madagascar



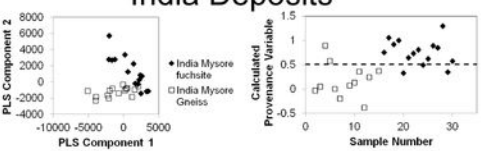
Madagascar Deposits



Myanmar



India Deposits



Sapphire Samples

Country	Deposit	Host Lithology	Color(s)	No. of Sapphires
Afghanistan	Badakashan	Phologpite Schist	blue	30
Australia	Queensland	Unknown	yellow, dark blue	43
India	Mysore	Unknown	black	29
Madagascar	Southern	Unknown	white, beige	30
Madagascar	Southern	Unknown	bluish purple	30
Tanzania	Songea	Pegmatite	grey and black	10
Tanzania	Umba Valley	Unknown	brown & tan with blue tint	5
USA	Rock Creek, MT	Unknown	blue, green, yellow	30
Vietnam	Unknown	Unknown	black	30

Ruby[a] Samples

Country	Deposit	Host Lithology		No. of Rubies
India	Mysore	Unknown	maroon	19
India	Mysore	Fuchsite	fuchsia	30; 8 samples
India	Mysore	Biotite Gneiss	fuchsia	30; 9 samples
Greenland	Fiskenaesset	Amphibolite	maroon	32
Kenya	Rift Valley	Unknown	fuchsia	30
Madagascar	Southern	Unknown	fuchsia	30
Madagascar	Southern	Amphibolite & Biotite Schist	fuchsia	35; 18 samples
Myanmar	Unknown	Unknown	maroon	30
Pakistan	Hunza Valley	Marble	magenta	13; 11 samples
Tanzania	Winza	Unknown	maroon	30
Tanzania	Longido	Fuchsite/Zoisite	fuchsia	14; 6 samples
USA	Wyoming	Fuchsite	purple	39

[a] All pink to purple-red samples were classified as rubies in this study.

Country	Deposit of Origin	Success Rate, %	Correct Predictions	Elements Enriched in Deposit being Modeled[1]
India	Psudipada	100	118 of 118	Ca (Zr, V, Na, H)
Afghanistan	Badakashan	96.7	29 of 30	Ba (Mg)
Vietnam	Unknown	96.7	29 of 30	Sr, Na
Australia	Queensland	100	66 of 66	Zr (Ba, Mg, Na)
USA	Rock Creek, MT	100	45 of 45	Zr (Mg)
Madagascar[2]	country of origin	100	45 of 45	Li (Si, Sr)
Madagascar[3]	Southern (White)	96.7	29 of 30	Mg
Madagascar[3]	Southern (Blue)	96.7	29 of 30	Ba (Li)
Tanzania[2]	Country of origin	97.3	72 of 74	(Fe, Ba)
Tanzania[3]	Songea	50	4 of 8	Ba (Na, Li)
Tanzania[3]	Umba Valley	50	4 of 8	Ca, Zr

Ruby Algorithm, listed in model order

Country	Deposit	Success I	Correct Predictions	
Kenya	Rift Valley	98.2	161 of 164	Zr (Ti, Ba, Ca)
India	Mysore, unspecified	91.9	137 of 149	Zr, Fe (Ti, Na, Ca, C)
Pakistan	Hunza Valley	100	26 of 26	Ca, Na
USA	Wyoming	100	42 of 42	(Ba, Sr, Na, Li)
Tanzania[2]	country of origin	100	42 of 42	Mg
Tanzania[3]	Winza	86.4	19 of 22	Ca (Na, Mg, Fe, Li)
Tanzania[3]	Longido	86.4	19 of 22	Zr (Mg, Ti)
Greenland	Fiskenaesset	98.9	92 of 93	Ca, Mg (Fe, Ba, Zr, Cr, Na)
Madagascar[2]	country of origin	97.3	73 of 75	Mg, Ca, Sr, Zr, Ba (Na, H)
Madagascar[3]	Southern, unspecified	100	32 of 32	Zr (Mg, Sr, Ba)
Madagascar[3]	Southern, gneiss and amphibolite	100	32 of 32	Ca, Na (Fe)
Myanmar	Unknown	86.7	39 of 45	(Mg, Ti, Cr, Ba, C)
India[2]	country of origin	86.7	39 of 45	Si, Sr, H, Hf
India[3]	Mysore, fuschite	83.3	25 of 30	Ba (Ca, Cr, H)

India[3]	Mysore, gniess	83.3	25 of 30	(Mg, Na)

[1]Elements with strong influence; those with lesser but significant influence listed in parentheses.

[2]Model identifies country of origin only.

[3]Model identifies deposit of origin.