Provenance Analysis of Detrital Zircons from Long Island Using Cathodoluminescence Imaging, Geochemistry, and Machine Learning

Davies, R. M.[1,2], Wuerz, M.[1]
1. Department of Biological Sciences and Geology, Queensborough Community College of the City University of New York, Bayside, NY 11364
2. Department of Earth and Planetary Sciences, American Museum of Natural History, New York, NY 10024

Introduction Detrital zircons preserved in glacial sediments across Long Island provide the opportunity to reconstruct the provenance of sedimentary material deposited by the Laurentide Ice sheet and to infer the direction and distance of glacial transport. Analysis of the ages of detrital zircon helps us to identify the bedrock sources from which these grains were eroded. U-Pb dating of zircons in moraines stretching from western Long Island at Huntington to Greenport and Hither Hills to the east reveals varying proportions of populations corresponding to contributions from New York and New England bedrock (e.g., Permian and Triassic magmas and intrusives, Acadian, Taconian, Avalonian, and Grenvillian), highlighting shifts in dominant source regions from west to east along the moraine system (Davies et al., 2024).

In this exploratory study, we focus on 216 detrital zircon grains from Greenport by integrating U-Pb age data, trace element geochemistry, and cathodoluminescence (CL) image analysis using computer vision techniques. The objective is to evaluate whether image-based zoning patterns and geochemical features correlate with zircon age and to test machine learning methods as a tool for pattern recognition and classification. This work serves as a pilot for future undergraduate-led research projects expanding to additional locations.

Methods. We combined zircon U-Pb ages, trace element compositions, and CL images using Python's Pandas library. Each image was processed using Pillow (for file handling and format conversion) and OpenCV (for grayscale conversion, contour detection, thresholding, and segmentation). From these images, we extracted metrics including average CL intensity, zone count, and mean zone area for each grain. These image-based features were then merged with age and geochemical data for statistical and machine learning analysis.

We used Matplotlib and Seaborn to visualize relationships through scatter plots and correlation matrixes. Principal Component Analysis (PCA) was applied via the python package scikit-learn to reduce the number of variables for pattern detection, and KMeans clustering was used to find patterns or groups that share similar characteristics without pre-labeling them to identify possible groupings based on multivariate features. We also tested a preliminary classification model to distinguish fractured versus intact grains based on their crystal shapes, though the results were inconclusive due to the zircons complex crystal shapes, leading us to classify the crystals manually.

Results. The zircon grains span a broad age range (183-2524 Ma), encompassing major orogenic and tectonic events (Fig. 1).

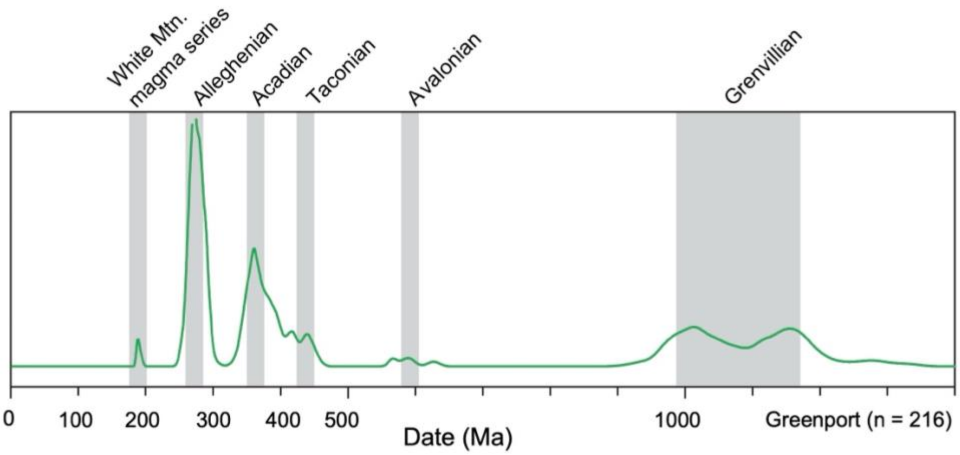

Figure 1. U-Pb age dates of 216 zircons collected from Greenport, Long Island.

Age and CL intensity showed little correlation, although the ~280Ma Alleghenian population displayed a narrower brightness range (Fig. 2), suggesting relatively uniform luminescence characteristics and potentially similar formation conditions compared to older populations with more variability. Two of the youngest grains have similar high intensities, while three of the oldest grains have lowest intensities possibly due to being more metamict.

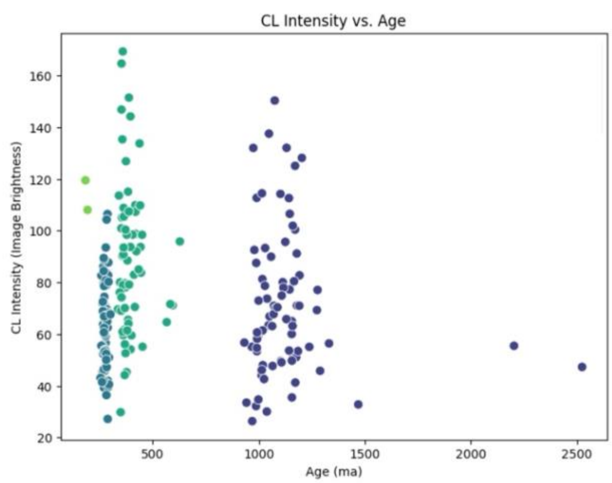

Figure 2. CL intensity versus age for 216 zircons collected from Greenport, Long Island.

All grain exhibit a wide range of internal structures and reveal no relationship between age and zoning complexity. Correlation analysis revealed a weak negative correlation (r ≈ –0.17) for CL intensity vs. zone count, a weak positive correlation (r ≈ +0.15) between CL intensity and mean zone area, and a moderate negative correlation (r ≈ –0.64) for zone count vs. mean zone area (Fig. 3). These trends suggest that crystals with numerous growth zones tend to have smaller average zone sizes.

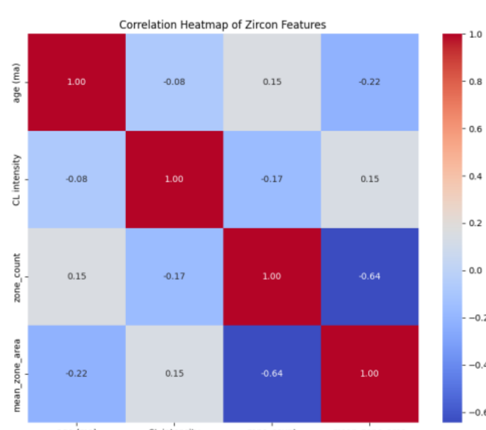

Figure 3. A comparison of zircon ages with CL characteristics including intensity, zone count or number of layered features, and mean zone area or the mean area for a layer or zone.

Fractured crystal forms were was observed in more than half the zircon grains, with both fractured and unfractured grains spanning the full age range. Anecdotally, older zircons show more rounding rather than fractured faces (Fig. 4).

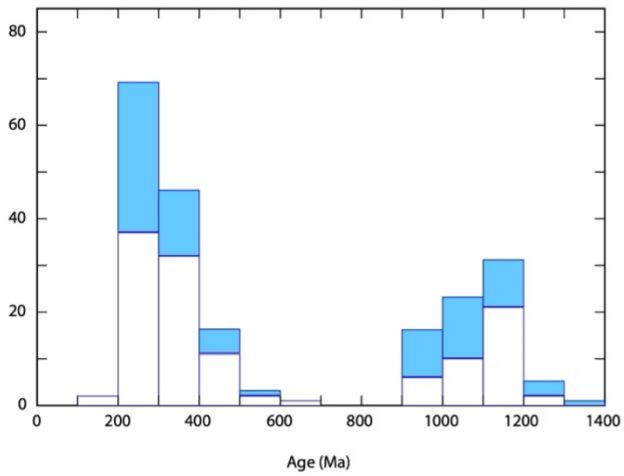

Figure 4. More than half the zircons in this study show breakage (white) verus intact (blue).

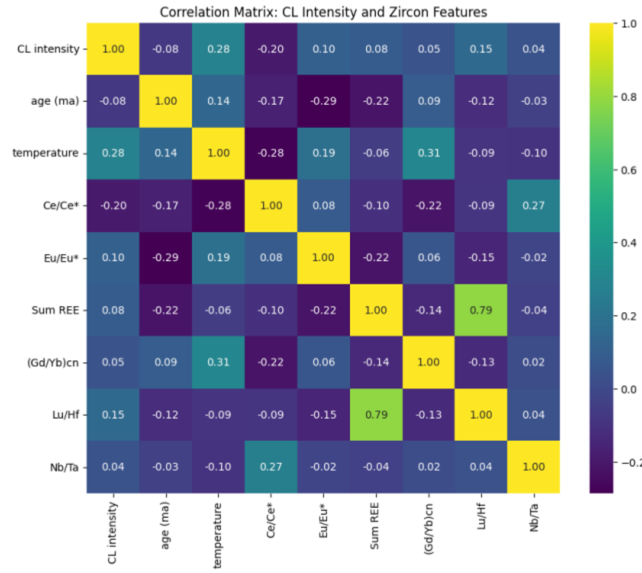

Figure 5. CL intensity plotted with ages, temperature (from Ti), and other geochemical features showing modest to little correlation.

Geochemical analysis showed limited correlations. A small negative correlation between zircon age and Ce/Ce* ($r \approx -0.29$) suggests that older grains may reflect different source conditions. Trace elements such as Ce/Ce*, ΣREE, and Lu/Hf showed some co-variation, reflecting potential similarities during zircon crystallization (Fig. 5).

Conclusion. Our integrated approach demonstrates the potential of combining zircon age, CL image analysis, and geochemistry for provenance studies of zircon populations. While strong correlations were limited, patterns in zoning and trace element data suggest subtle relationships. The heterogeneity of features likely reflects complex sediment mixing and long-distance transport from diverse bedrock sources. This work lays the groundwork for a scalable methodology suitable for expanded regional studies and for engaging undergraduate students in interdisciplinary geoscience research involving mineral properties and machine learning.

Davies, R.M., Jaret, S., Nicholas, J., Crowley, J., Kinney, S. (2024) The Provenance and Ages of Glacial Sediments from Long Island, NY. Long Island Geology Conference.