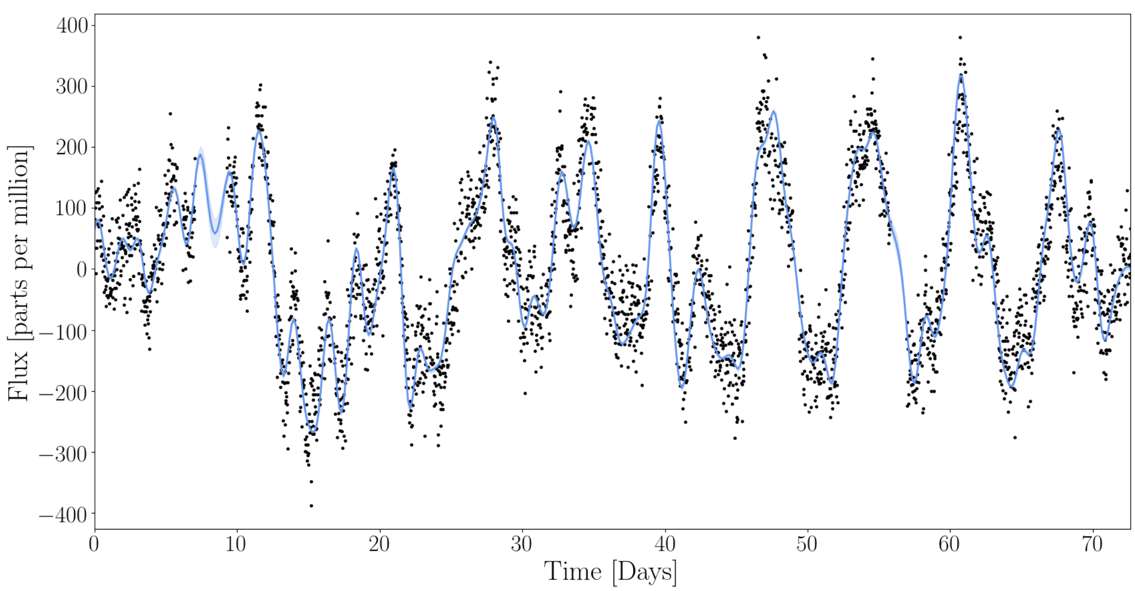
Title: Measuring the Ages of Stars

Scientific Mentor: Dr. Ruth Angus, Assistant Curator and Professor, Department of Astrophysics

The ages of stars are difficult to measure using standard techniques, however the *rotation rates* of stars can reveal their ages extremely precisely. The rotation rates of stars can be measured from slight changes in their overall brightness, because dark 'star spots' rotate into and out of view once every time the star completes a full rotation. These rotation signals can be seen by space-based telescopes like NASA's Kepler and TESS spacecrafts, which measure the brightness of stars once every 30 minutes to produce measurements of brightness (flux) over time, as shown in the plot below.

In this project, the Helen Fellow will measure the rotation rates and ages of stars from their brightness fluctuations using time-series analysis techniques and/or machine learning methods. After this, the project could take many different directions. For example, the ages measured could be used to study the evolution of planetary systems, stars, or even the Milky Way Galaxy itself. In addition to learning about stellar astrophysics, exoplanet science and/or Milky Way astronomy, the Helen Fellow will have the opportunity to learn a variety of data analysis and machine learning skills, including random forest regression, neural networks, and Gaussian processes. The Fellow will also have an opportunity to learn how to package, document and release code as an open-source software package.



Title: Co-Moving and Accelerating Stars in the Milky Way

Scientific Mentor: Dr. Jacqueline Faherty, Senior Scientist, Department of Astrophysics

Since 2013, The European Space Agencies' (ESA) Gaia mission has been imaging the entire sky with the purpose of measuring distances and motions for over a billion stars in our Milky Way Galaxy. In April 2018, ESA released the second data catalog to the community beginning a revolution in our understanding of the structure of our home galaxy. A dataset of this size is unprecedented and a challenge to our computational abilities. However, astronomers at the American Museum of Natural History have been diligently preparing for this data release by developing software and code that takes full advantage of the scientific progress possible with the catalog.

In this project, the Helen Fellow will work with Dr. Jackie Faherty and will be instructed on how to use machine learning techniques to mine the new dataset for co-moving, as well as accelerating stars in the Milky Way. A pre-cursor dataset of Gaia released in 2017 with 2 million stars showed several new substructures within the galaxy uncovered through this technique. With a factor of 1000 increase in stars, the yield of this project will vastly improve our understanding of the architecture of the galaxy.

Title: Atmosphere Characterization of Exoplanets and Brown Dwarfs

Scientific Mentor: Dr. Jacqueline Faherty, Senior Scientist, Department of Astrophysics and BDNYC Group

In the last few years, advancements in instrumentation have enabled the field of direct exoplanet characterization studies. Astronomers can now use imagers and/or spectrographs and study the light from a planet outside of our solar system directly. The planets that have been imaged so far are unlike the planets in the solar system—they are most similar to Jupiter, but considerably warmer and younger. So-called "warm-Jupiters" appear most similar to objects called brown dwarfs, failed stars that lack enough mass to establish stable Hydrogen burning.

In order to understand current exoplanet data and to place it in context with sibling brown dwarfs, scientists need to model the complex chemistry ongoing in the atmosphere of each object. At the American Museum of Natural History, the Brown Dwarfs in New York City research group (BDNYC) co-led by Dr. Jacqueline Faherty, has begun a collaboration with leading theorists to host atmospheric retrieval code on the museum's high powered computer cluster.

In this project, the Helen Fellow will learn how to implement the code, applying it to the BDNYC database of observed data on giant exoplanets and brown dwarfs—the largest and most complete repository of spectra and photometry for these objects that currently exists in the world. This method of computational remote sensing for exoplanet and brown dwarf data is largely regarded as the pathway to characterizing extrasolar planets. A clear and important component of the research will be deciphering the weather on other worlds including whether they are made of iron, silicate or even ruby dust.

Title: Black Hole Migration Leading to Mergers in Active Galactic Nucleus Disks

Scientific Mentors: Dr. Mordecai-Mark Mac Low, Curator and Professor, Department of Astrophysics, Dr. Saavik Ford and Dr. Barry McKernan, Research Associates, Department of Astrophysics, and Professors, Department of Physics, Graduate Center, City University of New York

The Laser Interferometer Gravitational-wave Observatory detected mergers of black holes more massive than expected from the evolution of massive stars, and occurring towards the upper end of the predicted range of frequency. Standard channels such as mergers at the centers of globular clusters appear insufficient to explain these observations.

McKernan et al. (2014) proposes an alternative. Stellar-mass black holes form at the end of the lives of stars in massive galactic nucleus clusters. When gas disks form around central supermassive black holes during active galactic nucleus episodes, the disks can capture the stellar-mass black holes. Once in the disk, the smaller black holes behave similarly to planets in protostellar disks, and migrate to trapping orbits. In these orbits, mergers occur efficiently, helping to explain the gravitational-wave observations.

Making concrete predictions from this channel requires study of the migration of stellar-mass black holes through such disks, along with accompanying improvement of models of the disks and interactions of stellar mass black holes with the disk gas.

In this project, the Helen Fellow will run magnetized gas dynamics simulations on parallel cluster computers using the Pencil Code and the Athena++ adaptive mesh refinement code to model the speed and direction of migration.

Title: Structure of Extreme Star-Forming Clumps in High Redshift Galaxies

Scientific Mentor: Mordecai-Mark Mac Low, Curator and Professor, Department of Astrophysics

Observations of high redshift star-forming galaxies during the era around 10 Gyr before present show that they look very different from modern disk galaxies. Rather than smooth disks with star formation primarily in spiral arms, the high redshift galaxies show giant clumps of intense star formation as much as a kiloparsec across (Elmegreen et al. 2009). These appear to form in gas-rich young galaxies (Daddi et al. 2010). Preliminary models of the structure of these objects using simplifying assumptions (isothermal gas) show that enormous amounts of substructure must form in these objects (Behrendt et al. 2016).

Our group has developed a method to combine magnetized gas dynamics simulations including supernovae, ionizing radiation and other stellar feedback with N-body gravitational models of stellar dynamics and stellar evolution (Wall et al. 2019). We have developed the capability to use this technique to model the formation of a giant star-forming clump in a gas-rich stratified disk.

In this project, the Helen Fellow will run and analyze supercomputer simulations using this model in order to examine the detailed structure of these clumps. In particular, we want to understand whether they generate bound globular clusters such as those that can still be observed in the present day.

Behrendt, M., et al. 2016, Astrophysical Journal Letters, 819, L2 (5 pp.)
Daddi, E., et al. 2010, Astrophysical Journal, 713, 686-707
Elmegreen, D.M., et al. 2009, Astrophysical Journal, 701, 306-329
Wall, J., et al. 2019, Astrophysical Journal, in press (ArXiv:1901.01132)

Title: Stellar Collisions in the Milky Way Galaxy
Scientific Mentor: Dr. Michael Shara, Curator and Professor, Department of Astrophysics

Long thought to be so rare that they could be ignored, direct collisions between, and violent mergers of stars are now known to be common events in rich clusters of stars. Such events produce some of the most "exotic" stars known, with apparent ages, partners and locations in clusters impossible to explain by any mechanism other than collisions. This happens because binary, triple and quadruple stars have collisional cross-sections thousands to millions of times larger than those of single stars.

Much less explored is the question: how often do stars collide OUTSIDE of star clusters? The question isn't just of academic interest. Just because the Sun and planets of our Solar System have escaped a catastrophic encounter with another star for the past 4.6 Billion years does not mean that it cannot happen in the future. More generally, the discovery of "red transient" stars, intermediate in brightness between novae and supernovae, suggests that roughly 10% of all the stars in any galaxy undergo a collision and/or merger during their lifetimes.

In this project, the Helen Fellow will collaborate with Dr. Michael Shara and be mentored in the creation of a realistic, three-dimensional star map of the environs of the Sun, using data from the Yale Bright Star Catalog, the RECONS survey and the 2nd data release of the GAIA satellite. This map will be used as a "target'; the Fellow will simulate sending binary and triple stars through the "target", and study the effects of close encounters with the computer code FEWBODY. The outcome of this project will be the best-yet estimate of the rate of direct stellar collisions in the Milky Way galaxy.

Title: Hunting for the first Exoplanet outside our galaxy

Scientific Mentor: Dr. Michael Shara, Curator and Professor, Department of Astrophysics and Dr. Ruth Angus, Assistant Curator and Professor, Department of Astrophysics

The discovery that exoplanets must be orbiting almost every star in the Milky Way has profoundly changed astrophysics, and our understanding of the cosmos, in the past 20 years. Remarkably, though, we still know nothing about planets in any galaxy other than our own. To remedy this important shortcoming, we have collected 2 million continuous light curves, with 300 million data points, to search for planets eclipsing their host stars in each of the two galaxies nearest to our own: the Large and Small Magellanic Clouds. The same dataset will also yield the largest samples ever of extragalactic eclipsing contact binary stars and cataclysmic binary stars. Each of these data sets will be critically important in testing theories of the formation and evolution of exoplanets and binary stars.

In this project, the Helen Fellow will work closely with Drs. Michael Shara and Ruth Angus to locate and characterize the expected ~1000 periodic transiting planets and eclipsing binaries expected in the dataset. The Helen Fellow will use, and compare the results of:

1) box-least squares (BLS) methodology applied to all 2 million light curves, and

2) automated search algorithms using classifiers based on light curve amplitudes, colors, skews, kurtosis and previous histories, as well as other statistical tests.

Title: Mineral Phase Identification in Stacks of Element Maps of Meteorites

Scientific Mentor: Dr. Denton S. Ebel, Curator and Division Chair, Department of Earth and Planetary Sciences

Meteorites are the most valuable resource we have for understanding the formation of larger bodies in the Solar System as they represent the ground truth that all models must satisfy. Certain meteorites, called chondrites, contain unaltered products that formed in the early stages of the Solar System's evolution. Analysis of these components is critical if we are to answer the question of how the planets formed. One way to analyze these components is to make maps of them using the electron microprobe. These maps represent the abundance of an individual element in every pixel of a flat taken from the meteorite. They are essentially stacked elemental "photographs" of the meteorite, like a hyperspectral cube. Intensities for each element in each pixel can be used to identify the mineral phase of that pixel.

The goal of this project is to produce code that can independently identify the mineral phase of each pixel in such sets of elemental maps of a meteorite and report mineral abundances across many thousands of pixels. While there are numerous mineral species, we can restrict the number of possible minerals to a fixed set that is most commonly found in meteorites. More advanced iterations of the code could potentially query a database of all minerals (terrestrial and meteoritic) and give probability distribution of a specific pixel matching a pre-identified set of minerals input by the users. Below is a sample of elemental maps that could be part of a target data source. The raw individual element maps are 32 bit grayscale .tif files. Existing code combines three maps on the R-G-B color channels to provide a visual map to aid in eyeball identification of the mineral phase based on the color of the pixel. However, this is not a precise method since minerals can be pure phases (e.g., spinel Mg_2AlO_4) or be solid solutions (e.g., olivine $(Mg,Fe)_2SiO_4$). Minerals with similar compositions (in this example a pyroxene $MgSiO_3$) can deceive algorithms and eyeballs. Our goal is to enable counting the amount of each mineral phase present, with high accuracy.

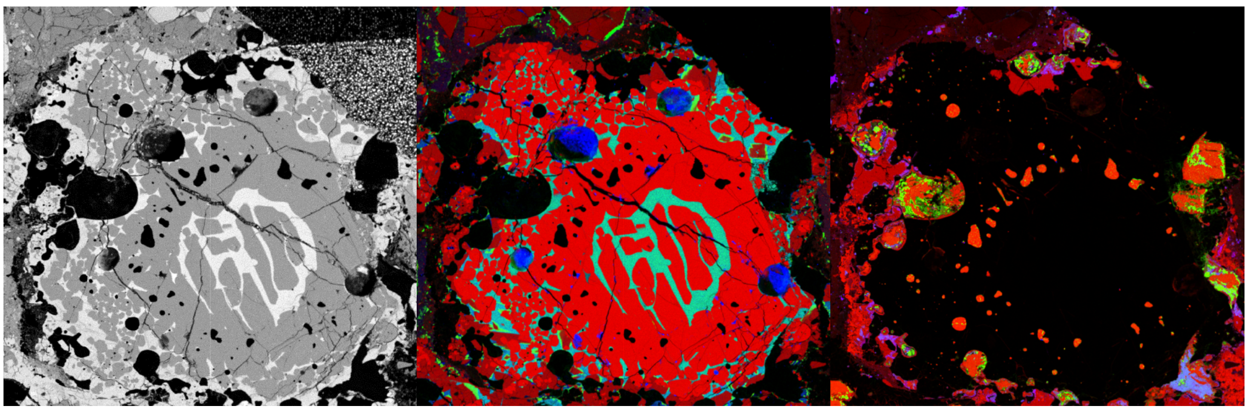


Figure. Maps of an individual object (a chondrule with metal nodules imbedded in it) in the meteorite Semarkona (LL3.00). Left: Silicon grayscale element map (example of raw output from electron microprobe). Center: R-G-B combined mosaic of Mg-Ca-Al grayscale maps. Right: R-G-B combined mosaic of Fe-Ni-S grayscale maps. From these maps we can see that there are combinations of grayscale maps which can be used to identify specific mineral phases but not others. Generally map stacks contain Si, Ca, Mg, Al, Ti, Fe, Ni, and S.

As a jumping-off point for this project, our group has written code (in IDL) that applies user-defined thresholds for each element in each mineral and compares every pixel in linear combinations of maps to those thresholds. Minerals are identified in a decision tree in order of their chemical exclusivity, so an easy to identify mineral such as quartz (SiO_2) is checked before a more complex phase such as olivine or calcium silicates. In such a tree, a pixel is considered identified when a mineral match is found, after which that pixel is disregarded for the remaining phases. While this approach has been shown to be acceptable (see Ebel et al. 2016, *Geochimica et Cosmochimica Acta*) it is highly inefficient and is only as precise as the user-defined thresholds which generally do not take into account all the necessary phases.

In this project, the Helen Fellow will help develop an algorithm that should be able to correct the data for backgrounds associated with the electron microprobe, and adaptable to different conditions of collection of element x- ray intensity.

Title: The North Atlantic Oscillation—Synthesizing the Data

Scientific Mentor: Dr. Nathalie Goodkin, Assistant Curator and Professor, Department of Earth and Planetary Sciences

The 45-year record of CO_2 at Mauna Loa initiated by Dr. Charles Keeling, and continued by NOAA, clearly indicates that greenhouse gases are increasing in the atmosphere. It has been proposed that this increase may not lead to a gradual change in climate over time, but could trigger abrupt changes with rapid and possibly devastating effects to society and the environment (NRC, 2002). Yet, we currently lack the ability to attribute specific climate changes exclusively to human activities, and thus to predict the future impact of these influences. Increased knowledge about long-term patterns of change, particularly with respect to ocean-atmosphere coupling such as the North Atlantic Oscillation (NAO), may allow us to identify anthropogenic perturbations within the context of long-term natural background climate variability.

The NAO, a meridional oscillation in atmospheric mass over the North Atlantic Ocean, is measured by pressure anomalies between Iceland (65°N and 23°W) and the Azores (38°N and 26°W) and recorded as the NAO Index (NAOI) by Hurrell (1995). In a positive NAO phase, both the low (Iceland) and high (Azores) pressure zones are intensified. This leads to an increased frequency of strong winter storms crossing the Atlantic Ocean along a northeast transect toward the Icelandic low. These conditions

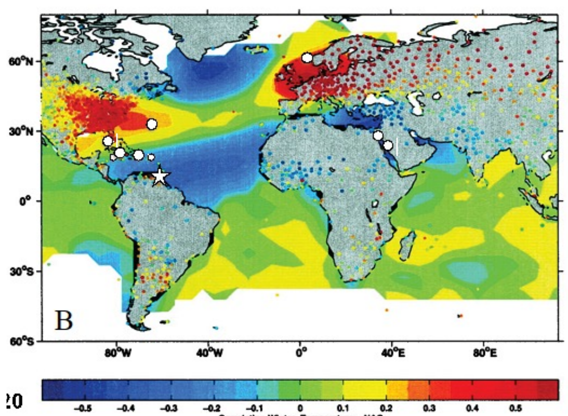


Figure 1: Spatial correlation (r) of December, January, February and March sea surface temperature and surface temperature to the NAO index (DJFM) (figure from Visbeck et al. 2001). Large circles indicate existing and star indicates forthcoming NAO reconstructions from Bermuda and the North and Norwegian Seas, as well as existing SST records longer than 150 years to be analyzed in this proposal, from Dry Tortugas, Cuba and Puerto Rico and the Red Sea. Small circles indicate SST records <150 years.

cause colder temperatures over the Northwestern Atlantic, and warmer and stormier conditions in the eastern United States. The deflection of the winter storm track north toward the intensified low-pressure zone leads to colder and drier conditions in Southern Europe and the Middle East (Figure 1). In the negative phase of the NAO, both pressure anomalies are reduced and the gradient is weakened. Over the past ~150 years, positive and negative phases of the NAO influence spatial patterns of surface temperature, showing clear 'centers of action' with maximum and minimum temperatures, as well as null zones of no variability (Figure 1).

Currently, two NAO reconstructions exist from networks of tree rings from across Canada and Europe, whereas multiple individual corals and mollusks have been used to reconstruct either the NAO or the sea surface temperature (SST) at NAO sensitive locations (Figure 1).

In this project, the Helen Fellow will synthesize the marine SST and NAO records from around the globe to statistically analyze and model changes to the NAO at different time frequencies to investigate whether 1) the footprint of the NAO changes with time or 2) the variability of the NAO changes with time.

Title: A Flow Visualization Tool to Identify, Characterize, and Track Biologically Generated Eddies in the Ocean

Scientific Mentor: Dr. David Lindo-Atichati, Research Associate, Department of Earth and Planetary Sciences, and Assistant Professor, Department of Engineering Science and Physics, College of Staten Island, City University of New York

This project is aimed at developing a tool to study how collective behavior of marine animals affect the motion of the ocean at different scales. Two questions inspire this project. 1) Why is the motion of the ocean important? The ocean produces 50% of the planet's oxygen, meaning that every other breath we take comes from the sea. Also, the oceans store and transport the vast majority of the carbon dioxide on Earth. 2) Why is collective behavior of marine animals important? Their small individual size notwithstanding, biological communities form dense aggregations of tens of meters that move vertically hundreds of meters during vertical migrations.

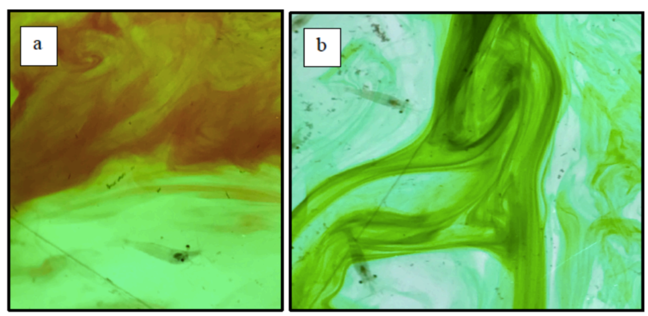


Figure 1. Side (a) and top (b) views of vortices generated by baby shrimp in Dr. Lindo's MIT rotating tank.

These vertical migrations are the greatest migrations in the world. At the length scale of the community aggregation, the collective behavior and motion of marine animal can be relevant to the planetary driven motion of the surrounding water column. Theory of fluid dynamics and limited empirical evidence suggest that groups of animals generate eddies that cannot be generated by isolated individuals.

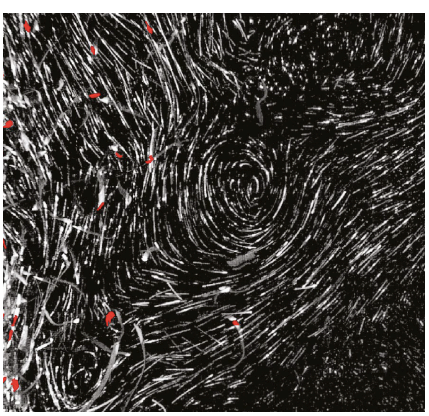


Figure 2. (a) Pathlines of 10-µm neutrally buoyant particles illustrate the fluid motion to the right of swimmers' migration. Individual swimmers are overlaid in red. Adapted from Houghton et al. (2018).

To address the above-mentioned need, the Helen Fellow will collaborate with Dr. Lindo-Atichati to write an algorithm and develop a flow visualization tool using data obtained from Dr.

Lindo's MIT weather tank (Figure 1). Also, the Helen Fellow with use data from experiments that Dr. Lindo is conducting in a large seawater tank located at Lamont Doherty Earth Observatory. These new experiments that Dr. Lindo is conducting at Columbia University are expanding the revolutionary work started by Houghton et al. (2018) (Figure 2).

In addition to using laboratory data obtained embedding baby fish and baby crustaceans in the turntables and tanks, the Helen Fellow will build upon a simple new tool for visualization of coherent structures in biological fluid flows tool: Flowtrace (Gilpin et al., 2016) (Figure 3).

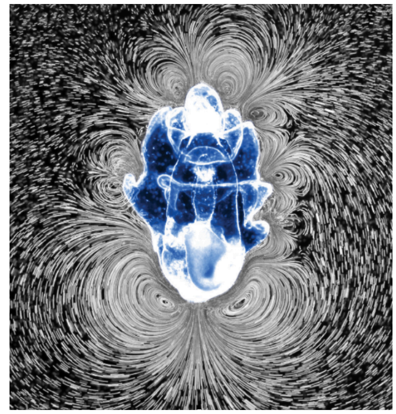


Figure 3 Nine counter-rotating vortices generated by an eight-week-old starfish larva using 6-μm tracer particles. Adapted from Gilpin et al. (2016).

The developed code will be publicly available on GitHub. The visuals developed by the Fellow will be publicly available on a website. One article as first author and one article as co-author will be submitted to peer-reviewed journals by the end of the residency. This project will contribute to build a paradigm-shift in marine biophysics, with novel implications for a planetary scale biological locomotion of (not in) the ocean.

Title: Computational Identification of Units of Evolution within Spider Silk Genes

Scientific Mentor: Dr. Cheryl Hayashi, Curator, Professor and Leon Hess Director of Comparative Biology Research and the Director of the Sackler Institute for Comparative Genomics

This project will develop bioinformatics tools to reconstruct complex, highly repetitive DNA and protein sequences of silk genes based on long- and short-read transcript data. Spider silk has remarkable mechanical properties. It is one of the toughest known materials, is self-assembling and self-healing. With these attributes, spider silk is one of the most intriguing biological materials on the planet. While silk production has evolved independently several times among arthropods, it is best known for its critical role in spider ecology and evolution. Many spider species, particularly orb-web weavers, produce numerous different types of silk that originate from a series of specialized abdominal glands and that serve different specialized functions such as orb web construction, prey capture, and egg encasement.

Despite the evolutionary and applied significance of silk, little is still known about the molecular structure and evolutionary diversity of silk proteins. Silk genes are extremely difficult to fully characterize at the DNA sequence level because of their large size, complex repeat structure and a homogenization process that minimizes variation and erases evolutionary signal within and between gene copies. Despite these difficulties, long-read next-generation sequencing methodologies are now providing full-length reconstruction of these problematic genes.

Given the accumulation of full-length silk genes across various spider species, we now have the opportunity to investigate the fine-scaled structure and evolution of the repetitive organization in silk proteins. This repetitive structure includes canonical sequence motif of various sizes organized in a hierarchical fashion. What is needed is a customized pipeline for the automated identification of these repeated motifs in order to understand how they change over evolutionary time and correspond to silk mechanical properties.

In this project, the Helen Fellow will develop this pipeline while being immersed in a lab that studies spider silk production from evolutionary, biomechanic, and biotechnology perspectives. This pipeline has the potential to transform our understanding of the molecular structure and function of spider silk. This research will also have applicability beyond silk genes to other genomic regions across the tree of life that contain similarly complex and enigmatic sequence structure.

Title: Elucidating Mitogenome Arrangements in Deep-Sea Hexacorals (Cnidaria: Actiniaria and Antipatharia)

Scientific Mentors: Dr. Estefanía Rodríguez, Curator and Professor, Division of Invertebrate Zoology and Dr. Mercer Brugler, Research Associate, Division of Invertebrate Zoology and Assistant Professor, Department of Biological Sciences, New York City College of Technology

Are you interested in analyzing the genetic blueprint of bizarre marine creatures that inhabit the largest environment on Earth? If so, Dr. Estefania Rodriguez's deep-sea evolutionary biology lab wants you! We recently sequenced the complete mitochondrial genome of several species of sea anemone and black coral (both of which are relatives of jellyfish and corals). The project entails assembling and annotating these new mitogenomes, and also re-annotating anthozoan mitogenomes that are publicly available on GenBank.

The Helen Fellow will combine traditional approaches that search for start and stop codons with a host of newly-released software that takes into account gene length and 'TA' and 'T' bases that would serve as a stop codon if 3' A residues were added. Given the diversity of taxa analyzed, it is possible that new gene orders (and gene content) will be revealed. We also anticipate having a complete transcriptome available from a highly venomous sea anemone (*Alicia* sp.) that will need to be bioinformatically scanned for potentially medically-important venom genes. Upon successful completion of the project, the Helen Fellow will be eligible to participate in a future research cruise to collect additional deep-sea anthozoans.

Title: Comparative Genomic Analysis of Two Phenotypically Divergent Spiny Eel Species from the Lower Congo River

Scientific Mentor: Dr. Melanie Stiassny, Curator and Professor, Department of Ichthyology

This project aims to create and analyze high-quality genome assemblies for two spiny eel species (*Mastacembelus brichardi* and *M. brachyrhinus*) to investigate the evolutionary relationship between the two strikingly phenotypically divergent sister taxa. Both species are found in central Africa's lower Congo River (LCR), an area with pronounced species diversity and endemism, including four mastacembelid eels. The two species represent recently-diverged sister species based on previous phylogenetic analyses, yet *M. brichardi* exhibits a cryptophthalmic phenotype (deeply embedded, reduced optic globes and reduced melanin pigmentation, among other features) while *M. brachyrhinus* has functional eyes and is fully pigmented. While work has been done on molecular explanations for similar phenotypes in other systems such as cave fish, the environment of the LCR differs markedly from that of subterranean fishes in that the LCR is highly dynamic hydrologically with numerous rapid systems. A comparison of these congeners could reveal potential molecular pathways that contribute towards cryptophthalmic phenotypes in non-subterranean environments.

In this project, the Helen fellow will create a pipeline for the assembly of these two species' genomes, as well as annotate the genomes for functional regions (genomic data have already been generated). They will then identify candidate genes and specific variants that might contribute towards *M. brichardi*'s cryptophthalmic phenotype by implementing comparative genomic and variant effect predictor tools, as well as leveraging extant genomic data for previously sequenced cryptophthalmic fish species. The results of this project will bring insight into molecular underpinnings of evolution and speciation in a highly diverse and species-rich environment.

Figure. The divergent phenotypes of *M. brichardi* and *M. brachyrhinus*. *M. brichardi*'s cryptophthalmic phenotype is seen by its reduced pigmentation and the reduced and recessed optical globes seen in the computed tomography (CT) skull scan. Also shown is *M. brachyrhinus*' pigmented phenotype with functional eyes and larger optical globes.

Title: Extracting Land Cover Information from Ultra-High Resolution Aerial Imagery

Scientific Mentors: Dr. Mary Blair, Director of Biodiversity Informatics Research and Mr. Ned Horning, Director of Applied Biodiversity Informatics, Center for Biodiversity and Conservation

In recent years, the use of drones to capture aerial imagery has become commonplace among researchers and practitioners in biodiversity conservation. Often the goal is to create land cover maps of an area that was imaged to inform conservation activities. Unfortunately, extracting land cover information and creating accurate maps from these ultra-high resolution images is difficult. This is due to many factors including high intra-object variability and variable illumination and reflection geometry within objects. A factor complicating an inherently difficult situation is that, in many cases, drone imagery for land cover mapping is acquired under less than ideal conditions using consumer-grade cameras.

Recent improvements in machine learning-based image recognition methods and the hardware required to run powerful algorithms can be applied to advance the extraction of land cover information from ultra-high resolution images. Machine learning algorithms can learn to identify patterns in data without being programmed with specific preconceived rules. A subset of machine learning, deep learning, sets itself apart by its effective use of large amounts of data to train models and make predictions. The AMNH's Center for Biodiversity and Conservation is working to leverage deep learning to extract land cover information from aerial and satellite imagery and inform biodiversity conservation efforts around the globe. The goal is to improve the state of the art and the state of the practice by developing new open source software and image processing workflows.

In this project, the Helen Fellow will apply algorithms and develop new workflows to improve our ability to extract information from images. This will involve contributing to our Neural Network Image Classifier (Nenetic) software code base, as well as developing and testing new approaches for accuracy assessment and classification model training.