

Galaxy / CoralSNP: Baums Lab, Department of Biology

Greg consulted with Dr. Iliana Baums in the Department of Biology to develop the Galaxy / CoralSNP analysis environment. Reliable and standardized identification of genotypes is a critical need for basic research and restoration planning, especially in plants and animals that reproduce asexually and form large clonal populations like aspens, seagrasses, waterfleas and corals. Figure 4 provides the general process.

- The scientist collects the coral, extracts the DNA and submits it to the processing facility
- The raw sequenced samples and sample metadata are uploaded to the Galaxy / CoralSNP environment for analysis
- The analysis produces new sample MLGs and genotype information that can be accessed via the Galaxy / CoralSNP interface

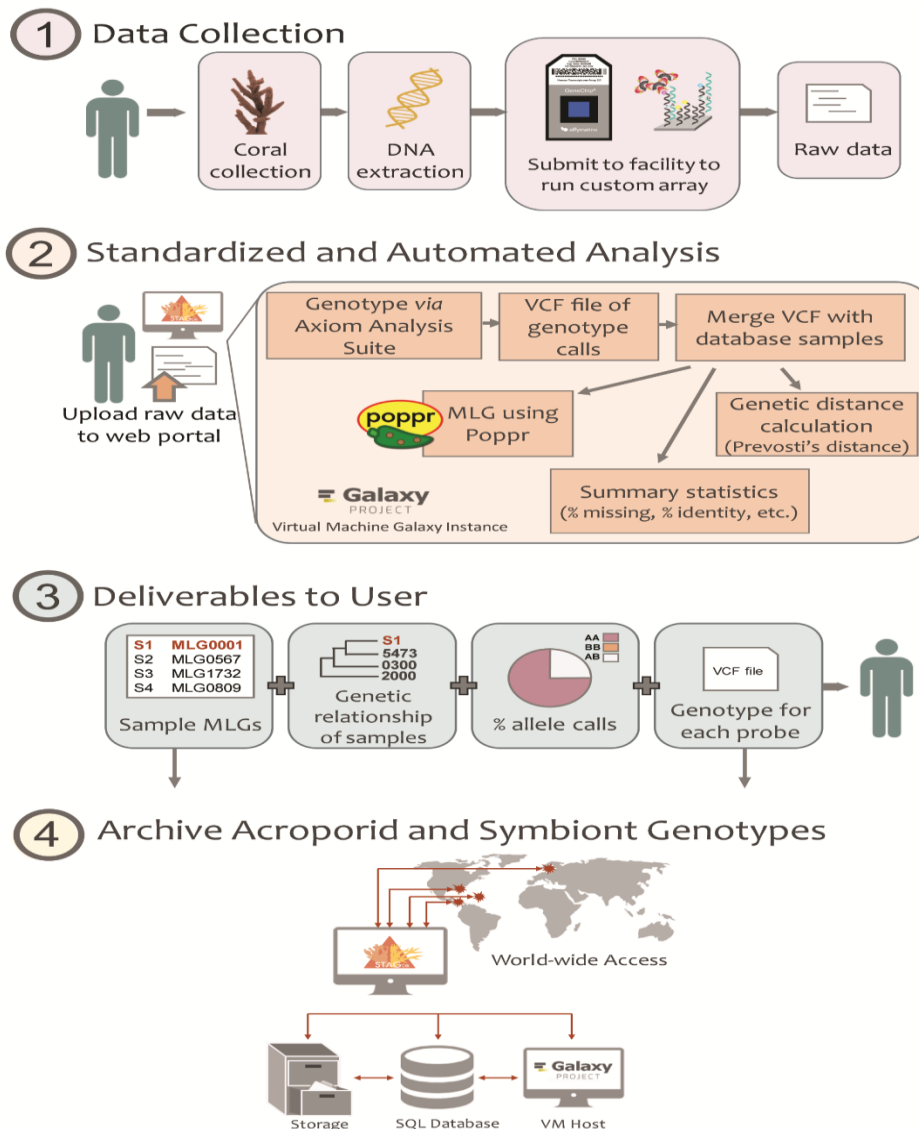


Figure 4: General process for standardized identification of genotypes

Galaxy / CoralSNP provides a high-resolution hybridization-based genotype array coupled with a standardized analysis workflow and database for the most speciose genus of coral, *Acropora*, and their symbionts (see figure 5).

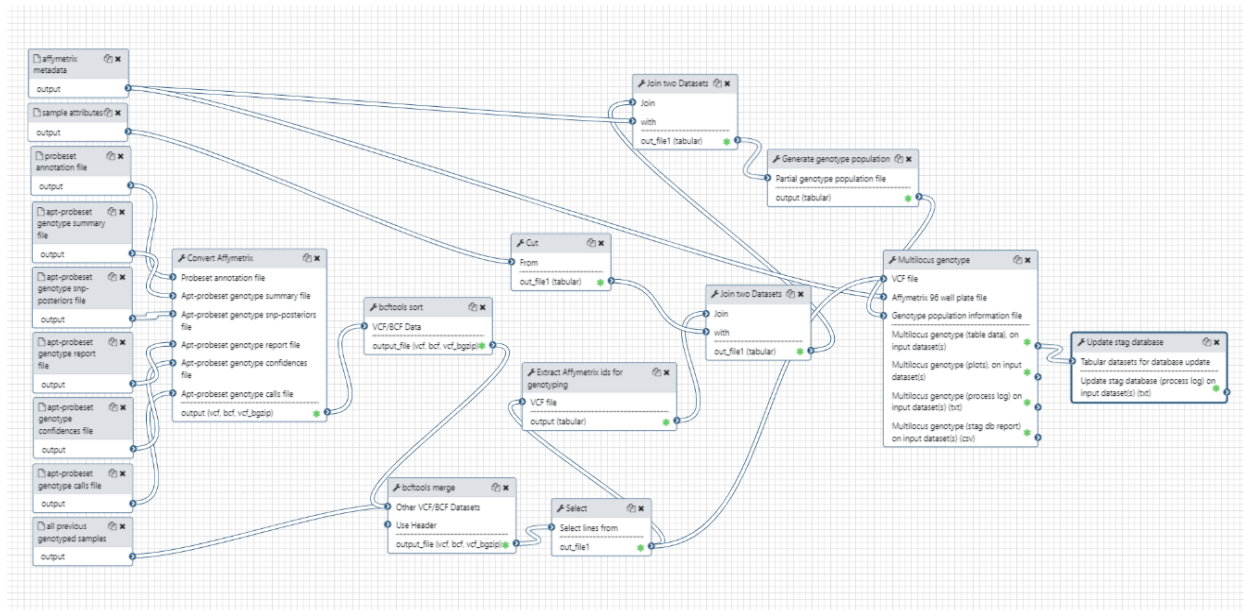


Figure 5: Galaxy / CoralSNP Workflow

The process is simple. A sample metadata file is created by the user from a template and uploaded along with their raw Affymetrix data files into the Galaxy CoralSNP environment. The appropriate files are selected as inputs to the *Queue Genotype Workflow* tool (Figure 6) which validates the metadata, executes the CoralSNP workflow (Figure 5) and updates a dataset that contains all previously genotyped samples as well as the STAG database with the samples in the current run. From the user's perspective, the entire analysis is as simple as uploading data and specifying it as the input to execute a tool.

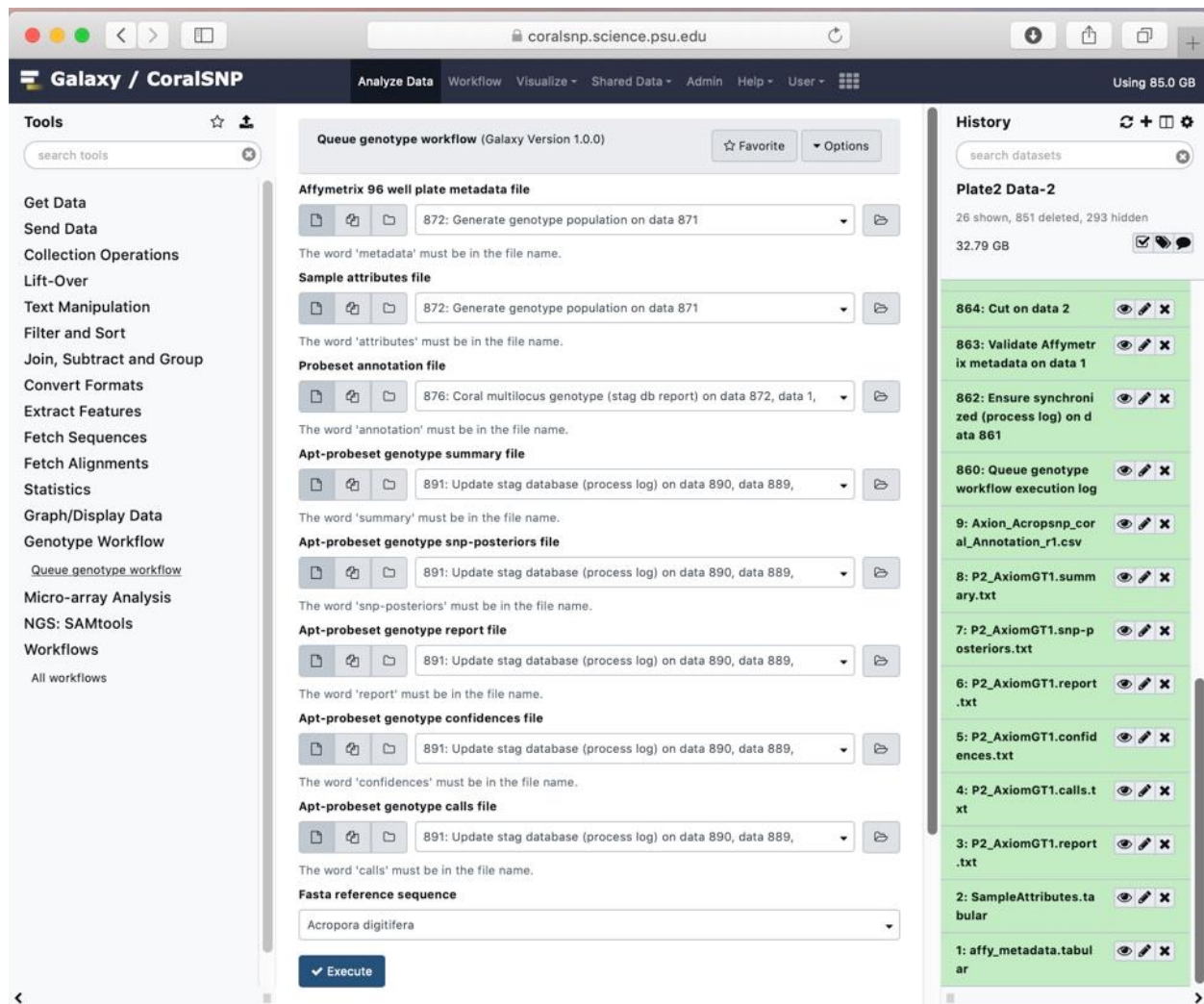


Figure 6: The Galaxy / CoralSNP Queue Genotype Workflow Tool

The *Queue Genotype Workflow* tool shields the complexity of the analysis from the user, and performs its functions via the Galaxy REST API to produce multi-locus genotypes for Caribbean or Pacific acroporids. The SNP array combined with the analysis allows for reliable, standardized identification of host genet and symbiont strains and serves as a template for the development of arrays for additional coral genera. This data can be used for downstream genomic analyses as well as restoration planning.

In the workflow, new user-supplied samples are compared to previously genotyped samples and the results are integrated into the CoralSNP STAG database. This archive of coral genets and symbiont strains can be used to identify reefs with high host and/or symbiont genetic diversity, temporal and spatial changes, and shuffling in host-symbiont pairings. In addition, a subset of the Caribbean genotyping probes can be used to genotype Pacific acroporids, expanding the utility of the STAG workflow to hundreds of species. The workflow outputs include a map of collection locations (Figure 7), neighbor-joining trees based on Prevosti's

genetic distance (Figure 8) and pie charts identifying generation hybrids (Figure 9).

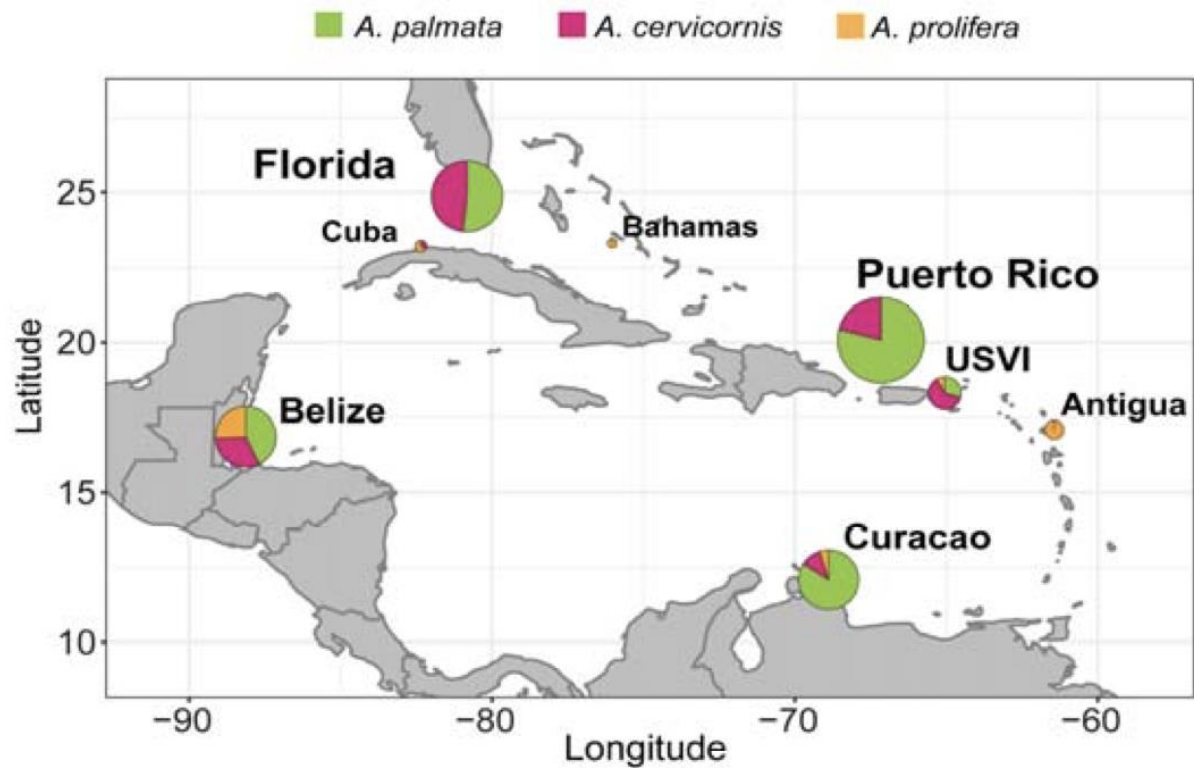


Figure 7: Example map of sample collection locations

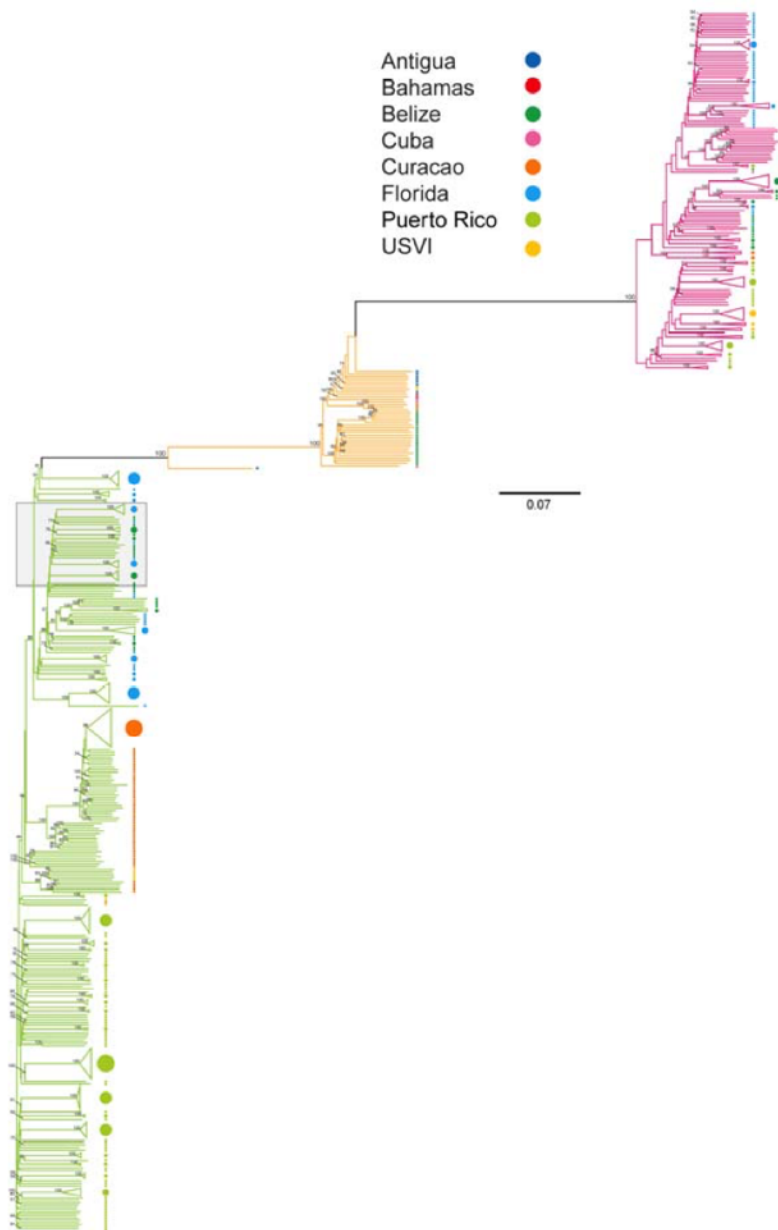
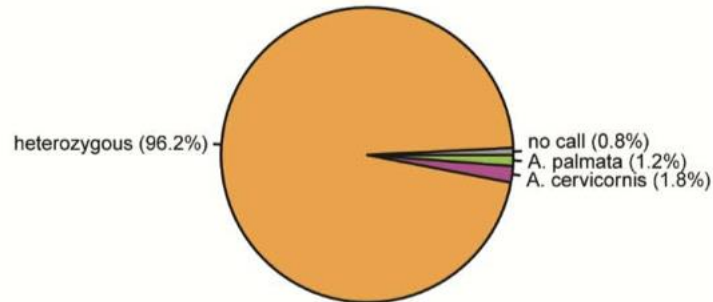
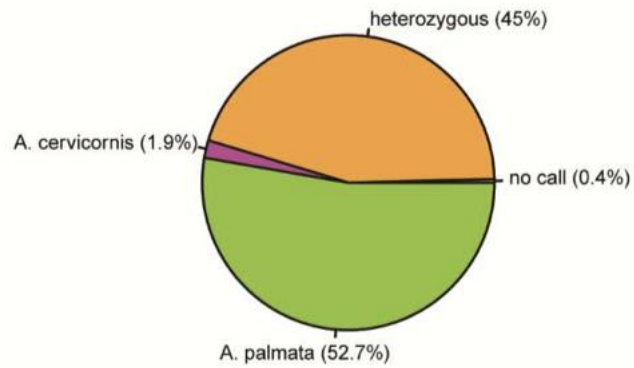


Figure 8: Prevotti's genetic distance of 19,694 SNPs is used to construct a neighbor-joining tree

F1 Hybrid: 13692



Later Generation Hybrid: 15727



A. cervicornis: 6299

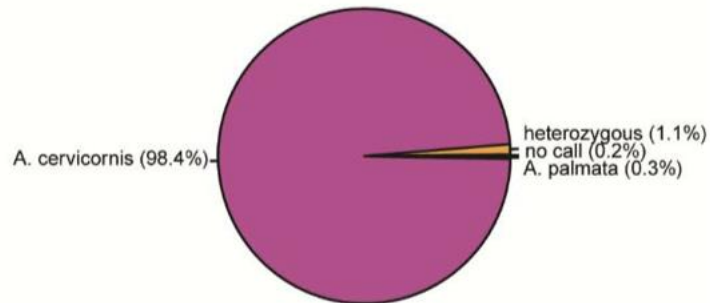


Figure 9: Species-specific SNPs identify hybrids