The iPlant Collaborative

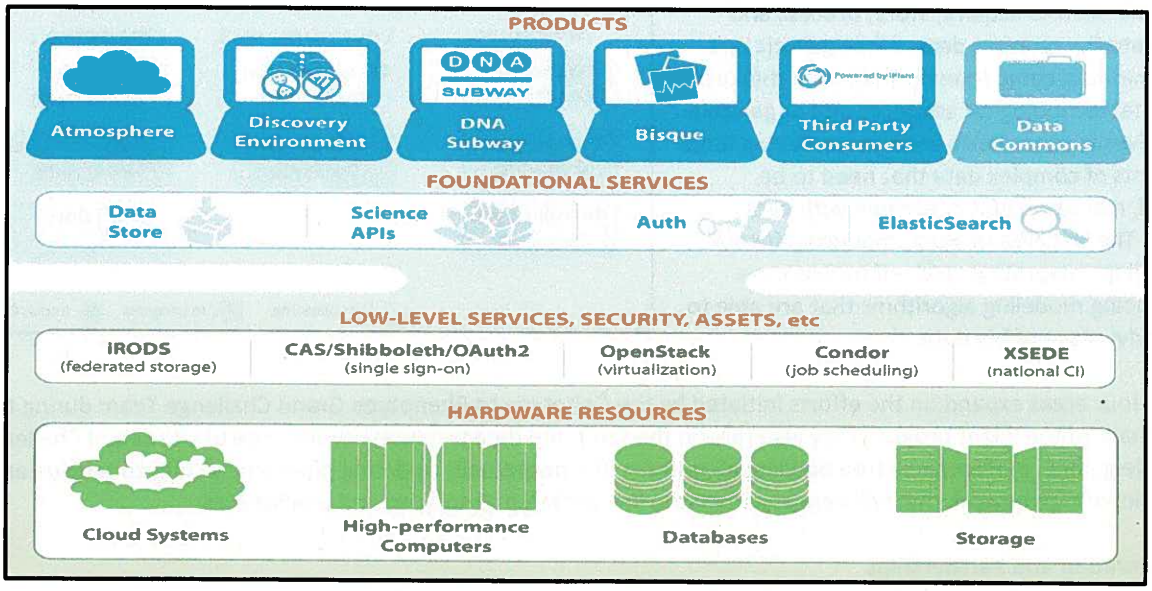
http://www.iplantcollaborative.org/

The iPlant Collaborative is a virtual organization of biologists and computer scientists who have developed a large-scale cyberinfrastructure (CI) of bioinformatics and computational resources to support modern biology research. The iPlant Collaborative was funded initially as an investment to enable plant sciences, but the ubiquity of iPlant's tools and CI has made it a large part of NSF/BIO's long-term CI investment portfolio, which includes other data-driven investments such as the Protein Data Bank and Dryad.

The iPlant CI

iPlant provides an online collaborative working environment that includes numerous user-focused software tools and workflows, data sets and data storage capacity, direct access to high-performance computing services, and educational and training resources.

The iPlant CI Architecture and Components

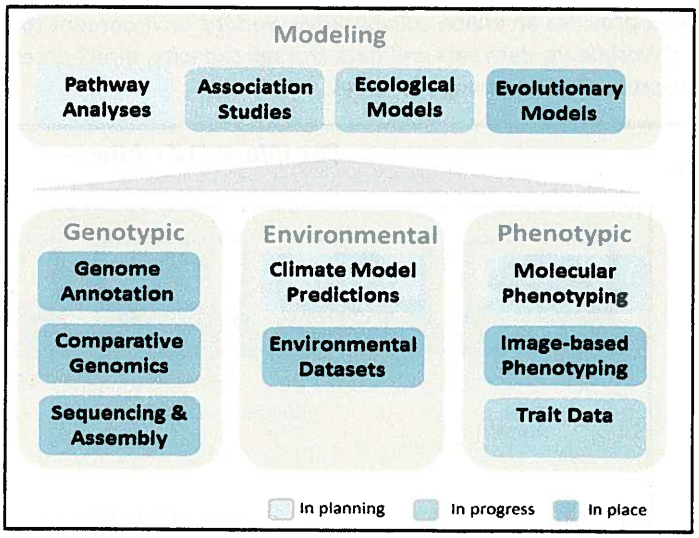


Component	Description	Target Audience
Atmosphere	User-friendly, on-demand cloud computing and persistent services	Users with desktop use cases or complex software environments
Discovery Environment	User-friendly, petascale graphical Workbench	Command-line naive users who need to do scalable bioinformatics
Bisque	Platform to facilitate cloud-based exchange and exploration of biological images	Command-line naive users and savvy users alike who need to do image analysis
DNA Subway	Educational interface to genomics topics	Beginning users and educators
Science APIs	RESTful interface to all iPlant services	Advanced users, developers, 3rd party infrastructure or service providers
Data Store	Capacious, scalable, shareable storage	Shared and used by all iPlant users
Data Commons	Makes value-added data available for discovery and reuse	All levels of iPlant users

The iPlant Science Enablement Vision

The initial scientific focus of iPlant was established by an 18-month, community driven grand challenge process (beginning at the start of the project in February, 2008), which led to the identification of two grand challenge focal areas: mapping the Tree of Life and understanding Genotype-to-Phenotype relationships. These two areas drove iPlant's initial CI developments during their first five years of funding. Conceptual design of the CI was completed in 2009 and the first set of CI deliverables (Discovery Environment and DNA Subway) was released in 2010. Atmosphere and the foundation Application Programming Interfaces were released in 2011, along with Powered by iPlant Program and several stand-alone applications.

Today, iPlant continues to focus resource development activities on deliverables identified by the research community as critical for addressing complex biological questions in four general areas of work: genotypes, phenotypes, environment and modeling . The first three deal with the need to acquire, store, process and integrate the relevant data. While genetic and genomic data come to mind when one thinks of big data, technological advances and large-scale initiatives in other fields are also producing large amounts of complex data that need to be stored, managed, and integrated with each other. The last area of work, modeling, focuses on scaling, integrating, and -- if necessary -- developing modeling algorithms that are able to take advantage of big data.



These four areas expand on the efforts initiated by the Genotype to Phenotype Grand Challenge Team during the first phase of the iPlant project. They also rely on the tools and data produced by the Tree of Life Grand Challenge Team (e.g., high performance tree building algorithms, the perpetually updating plant tree of life) to provide an evolutionary framework that will help elucidate the link between genotypes and phenotypes.

Collaboration and Partnerships

Federation and collaboration are critical components of the iPlant strategy, both to leverage external CI and to increase scientific impact. iPlant partners with three kinds of CI projects:

- Resource providers (like XSEDE) from whom iPlant consumes services;
- Peer CI development projects (like KBase, the Department of Energy Systems Biology Knowledgebase), who provide similar services; and
- Projects at various scales, which consume iPlant CI via the "Powered by iPlant" initiative. For example, the highly popular Galaxy biological analysis environment actually runs on top of the iPlant CI, as does the CoGe suite of software for comparative genomics analysis.

By partnering to provide user-focused tools with seamless access to high performance computing resources, iPlant is accelerating the rate of scientific discovery and enabling more complex questions to be addressed.