# Ionomics Atlas - A Tool To Explore Interconnected Ionomic, Genomic and Environmental Data

Eduard C. Dragut[1] Mourad Ouzzani[2] Amgad Madkour[1] Nabeel Mohamed[1]
Peter Baker[1] David E. Salt[3]
[1]Purdue University, Computer Science Department
[2]Qatar Computing Research Institute
[3]University of Aberdeen, Institute of Biological and Environmental Sciences
[1]{edragut, amgad, nabeel, pnbaker}@purdue.edu [2]mouzzani@qf.org.qa [3]david.salt@abdn.ac.uk

## 1. INTRODUCTION

Ionomics Atlas facilitates access, analysis and interpretation of an existing large-scale heterogeneous dataset consisting of *ionomic* (elemental composition of an organism), *genetic* (heritable changes in the DNA of an organism) and *geographic information* (geographic location, altitude, climate, soil properties, etc). Ionomics Atlas allows connections to be made between the genetic regulation of the ionome of plant populations and their landscape distribution, allowing scientists to investigate the role of natural ionomic variation in adaptation of populations to varied environmental conditions in the landscape.

The goal of the Ionomics Atlas is twofold: (1) to allow both novice and expert users to easily access and explore layers of interconnected ionomic, genomic and environmental data; and (2) to facilitate hypothesis generation and testing by proving direct querying and browsing of the data as well as different display modes of the results.

*Contribution.* Ionomic Atlas is an example of a cyber-infrastructure collaboratively developed by database, knowledge management, visualization and biology communities for scientists in areas such as biology, agriculture and medicine. Its main contribution is that it creates a platform for hypothesis generation and testing by offering complex ways to explore the interconnection between heterogeneous yet interlinked data layers: i.e., ionomic, genomic and environmental data. It is constructed to help scientists identify genes of critical importance to plant mineral nutrition, and understand their role in adaptation to varied *edaphic conditions* (e.g., drainage, texture, or chemical properties such as pH) in the landscape.

Meaningful visualisation and organization of genetic association data is in its infancy—we are aware of only one other tool related to ours [5]. This tool does not handle ionomic, soil and climate data as ours does and it has a different set of features than ours.

Numerous researchers from the database, information retrieval and knowledge management are actively involved in biology, climate, agricultural and/or GIS projects and we believe that our tool is of interest to them.



## 2. SYSTEM ARCHITECTURE

Ionomics Atlas has a classic *three tier* architecture: data, application and presentation tiers. The data tier component manages the ionomics, GIS and genetic datasets. The application tier controls the functionality of the application by performing detailed processing: e.g., query processing, process data for graph plotting, compute statistics (e.g., histograms). The presentation tier displays information related to ionomic, genetic and environmental data. We present them in turn now.

### 2.1 Data Tier

The data tier consists of a database server that hosts a relational database. The project utilizes an existing large data set of **ionomic information** consisting of quantification of P, Ca, K, Mg, S (the macronutrients in fertilizer), Cu, Fe, Zn, Mn, Co, Ni, Se, Mo (micronutrients of significance to plant and human health), Na, As, and Cd (metals causing agricultural or environmental problems) on 361 different populations of the plant *Arabidopsis thaliana*. This plant is ideal for such research as it has extensive molecular genetics and ecological research tools available. Coupled with this ionomic information we also have **genetic information** at 250,000 locations in the genome of all 361 populations, and multiple **GIS datasets** (e.g. precipitation, temperature, soil properties, humidity, pressure, etc.) linked to the geographical location where each population was collected in the plants native habitat. We are also actively cooperating with other research group to collect, integrate or link additional data into our system. The data is publicly available and can be downloaded from the Ionomics Atlas Web site[1].

The ionomics data from 93 Arabidopsis thaliana accessions in the Atlas are used in [1] and the entire ionomic data in the Atlas is used in [2]. The methods used to generate the association data are presented in [1].

### 2.2 Presentation Tier

The presentation tier consists of several components: query component, browsing component, and table and charts com-

[1]`http://www.ionomicshub.org/ionomicsatlas`

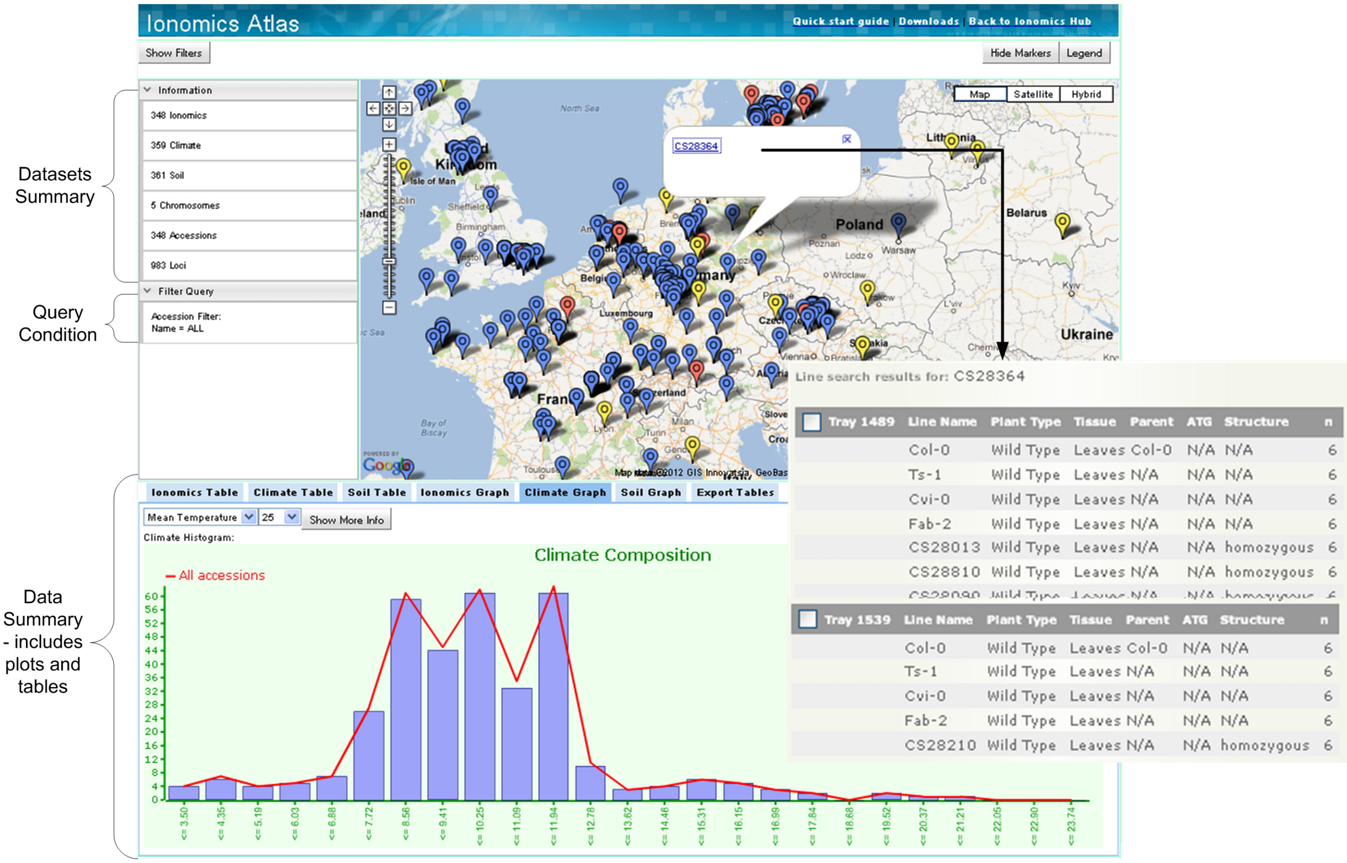


**Figure 1: Human classification of (in)consistent words.**

ponent. Figure 1 shows the main Web page of the Ionomics Atlas. We describe them below.

*Query Component.* This allows users with diverse backgrounds to retrieve data based on familiar parameters. For example, a biologist can query the data by genetic associations and/or soil properties, while an environmentalist can query by climate parameters. The parameters are currently grouped into so called *filters*: genetic, phenotype, climate, accession, soil and location. This component is accessed by clicking on the button Show Filters (Figure 1, left upper corner).

*Browsing Component.* We emphasize the visual browsing of the data in Ionomics Atlas, because, as the saying goes, "A picture is worth a thousand words". Browsing is a key mode of interaction for scientists, particularly biologists. Browsing supports the kind of "what if" scenarios that often arise in biological contexts, such as which geographical regions have *accessions* with the capacity to accumulate high concentration of Ca (Calcium), or which accessions has the highest concentration of Mg (Magnesium) across all 361 accessions. We develop a rich browsing component that allows a user to investigate the data layers individually as well as by meaningfully overlaying them on the graphical user interface. We give more details in Sections 4.3.1 and 4.3.2, respectively, when we describe the demo plan for our application.

*Table and Charts Component.* Ionomics Atlas is equipped with access to plotting and descriptive statistical analysis tools: e.g., z-scores (number of standard deviations from the mean) and percentage difference (from the mean). The items of this component are shown in the lower part of Figure 1. For example, for each query, Ionomics Atlas displays the tabular and chart views of the retrieved data along with the map of accessions. The table displayed is quite flexible. One can sort by any of the fields. If one clicks on a row, it shows where the accession is located on the map. The charts are mostly design to contrast visually two sets of data, e.g., the data retrieved in response to a query versus the entire data (In Figure 1 we show the distribution graph of the mean temperature of our entire dataset.)

### 2.3 Application Tier

The application tier component implements the logic of Ionomics Atlas and it is responsible for making the access to the data transparent to the users. It is responsible for query processing, supporting the browsing, perform the various statical computations and format the data for output.

## 3. SYSTEM IMPLEMENTATION

The project is implemented using open source software and free Web APIs. The back end is a PostgreSQL[2] database. The implementation of the front end of Ionomics Atlas utilizes a number of open source APIs. The geographic map browser is developed with the Google Earth API[3], which allows displaying a Google Earth interface in a web browser while enabling powerful rendering capabilities. We develop the plotting tools with Open Flash Chart[4]. Most of the code

[2] http://www.postgresql.org/
[3] developers.google.com/earth/
[4] http://teethgrinder.co.uk/open-flash-chart-2

to access the database is developed using a combination of Java and scripting languages.

Ionomics Atlas is Web accessible and is part of our existing www.ionomicshub.org. **ionomicshub** is a large scale, multidisciplinary cyber-infrastructure project [3], whose chief goal is to improve our understanding of how plants take up, transport and store their nutrient and toxic elements, collectively known as the **ionome**. Here scientist can find curated ionomic data on many thousands of plant samples freely available to the public.

# 4. DEMO PLAN

The goal of our demo plan is to illustrate all components of Ionomics Atlas to visitors and demonstrate to them that indeed Ionomics Atlas is a step forward in the ongoing endeavour of the scientific community to build tools that allow scientific discovery through data exploration [4].

## 4.1 Datasets

We will first explain to our audience the kind of datasets Ionomics Atlas is currently hosting: accessions, ionomics, climate and soil. For that part of our audience which is not familiar with this kind of data, we will give them details about the datasets. Our main goal however is to convey to our audience the combined scientific potential of these datasets when they can be studied together under a coherent cyber-infrastructure, which would otherwise be missed.

## 4.2 Interactive Query Mode

This mode allows users to explore the datasets in the Ionomic Atlas using a complex user query interface. The user can specify one or a combination of various criteria including: Accession, Gene, Ionomic profile, Geographic location, Growth conditions, Edaphic conditions, Climatic conditions. We give here two examples of the kind of complex queries that can be formulate on our tool:

**Query 1:** Return the ionomic data for the pollen in the accession Br-O (Bromine monoxide is important in the catalytic destruction of stratospheric ozone, especially in the lower stratosphere) under neutral soil conditions linked to the geographic location of the accession and overlaid with soil composition, mean temperature, and rain fall.

**Query 2:** Return accessions that show high levels of As (Arsenic) in leaves when plants are grown in acidic soil; show each returned accession in its geographic location overlaid with soil composition, mean temperature, and rain, as well as ionomic data.

We will prepare a number of (complex) query scenarios for our audience that will best emphasize the capabilities of our system. This will allow us to appropriately delve into the details of the system based on the audience background and interest. We will however encourage users to freely explore our system and suggest their own queries.

## 4.3 Browsing Mode

Our tool supports two main browsing modes that allow natural exploration of the underlying datasets of the Atlas:

### *4.3.1 Geographic map browser*

This browsing mode fits best with the idea of Atlas that we are proposing. We will illustrate to our audience how users can interact with a world map (much like Google map for local search, which most users are familiar with) and zoom-in/zoom-out and explore the available data. For example, users are able to display different kinds of data using different filtering mechanisms based on the criteria used in the query mode. They can also zoom in to a specific region and request either summary or detailed data about that specific region, e.g., average concentration of Li (Lithium) in leaves for all accessions in this region (summary), or all ionomic data for each accession in that region (details).

### *4.3.2 Ionomic Interactome Browsing*

This data exploration mode is particularly important for users coming from a genomic perspective and who are looking for significant genetic associations. Most of the results from either the query mode or the browsing mode can be displayed in one of the browsers depending on the choice of the users and the context of the search. Results can also be returned in the form of tables (list of accessions, list of elements, etc.) along with a number of plotting tools. For example, ionomics data for specific accessions is returned along with all associated metadata (e.g., growth conditions, edaphic conditions) with the ability to plot the ionomic data using histograms. Users can also easily switch from the query mode to any of the browsing modes, or between the different browsing modes, with the ability of keeping all the filtering they did in the previous step. For example, if they found some interesting accessions with high concentrations of P (Phosphorus) in leaves through the query mode, they can easily plot them on the geographic map. While querying or browsing, the Atlas users are given the option to link to our existing ionomics data [3] using accession name and similar elemental profile. We use Figure 1 to illustrate such a scenario. By clicking the one of the markers on the map, the user is presented with a brief description of the accession residing at those coordinates on the map. If the user desires more details about this particular accession then he/she can obtain them by simply clicking on the name of this accession. This will take the user to a different web page showing the set of results relevant to his/her search.

We will encourage an interactive and inquisitive demo session for each visitor. We believe that the amount of features and the reach graphical interface of the Ionomics Atlas will results in a lively demo session.

**Acknowledgements**: This work was supported in part by the US NSF grants DBI-0735191 (iPlant Collaborative, University of Arizona), DBI-0953433 and DBI-0701119, and by Cyber Center, Purdue University.

# 5. REFERENCES

[1] S. Atwell and et al. Genome-wide association study of 107 phenotypes in arabidopsis thaliana inbred lines. *Nature*, 465, June 2010.

[2] I. Baxter and et al. A coastal cline in sodium accumulation in arabidopsis thaliana is driven by natural variation of the sodium transporter athkt1;1. *PLoS Genet*, 6, Nov. 2010.

[3] I. Baxter, M. Ouzzani, S. Orcun, B. Kennedy, S. S. Jandhyala, and D. E. Salt. Purdue ionomics information management system. an integrated functional genomics platform. *Plant Physiology*, 143, Feb. 2007.

[4] T. Hey, S. Tansley, and K. Tolle, editors. *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Research, 2009.

[5] Y. S. Huang and et al. Analysis and visualization of arabidopsis thaliana gwas using web 2.0 technologies. *Database (Oxford)*, 2011.