2009 TAIR Community Survey Results

Robert J. Muller

December 2, 2009

Department of Plant Biology/TAIR
260 Panama Street
Stanford, CA 94305

Introduction

The 2009 TAIR Community survey polled the non-commercial TAIR user community to find out how well the TAIR web site is serving the needs of that community of users. It also surveyed the opinions of the community on approaches related to the sustaining of TAIR access and data over time.

The survey itself is an online survey that TAIR built on the surveymonkey.com web site and presented to the community through an email mailing. The sample was a random sample of 838 non-commercial TAIR users (users that did not designate themselves as commercial) with valid email addresses. The response rate of 40% yielded 336 responses as the result of two reminder emails over a 16-day period.

The report first presents tables and charts of the data along with summary statistics that describe the basic characteristics of the response. Each section interprets the results in a summary sentence at the beginning of the section which also appears in the executive summary. The section then gives the question asked and the response summary with an accompanying graphic representation.

The analysis section presents some additional analysis of the results using statistical techniques.

Appendix A discusses some of the issues with data quality. Appendix B contains a complete list of the organisms on which respondents work.

Executive Summary

The overwhelming majority of respondents feel TAIR is essential to their work.

Most respondent labs have between 1 and 10 people. Most respondents are professors or postdocs. Most respondents work in academic institutions. Respondents are global, equally distributed between the USA, Europe, and Asia. Most respondents carry out or supervise lab or field experiments as their main daily activity. Only a small minority (9%) work exclusively on Arabidopsis. The majority (62%) work mainly on Arabidopsis but also on other species. A significant fraction (29%) work primarily on another species but also use Arabidopsis.

Most respondents believe that TAIR should focus on extracting functional annotations and phenotypes from the literature and on updating gene structure of the genome. Lesser priorities included correcting genome sequencing errors and integrating community submissions of gene function, gene structure, and other data into TAIR. A third cluster combines providing sequence data for ecotypes and links to other resources. A fourth cluster includes access to data for other species, providing access to tools developed by other groups (Biomart, Genevestigator, etc.), and building novel tools. Finally, providing a helpdesk is the lowest priority activity.

A majority of respondents feel that it is essential to preserve most aspects of the current TAIR free-access model, with the exception of keeping access free for companies and keeping the site free of advertisements.

Two-thirds of the respondents have never provided data to TAIR; the other third is equally divided between those who have provided data once and those who have provided data more than once. A majority felt that they would be either much or somewhat less likely to submit data if TAIR did not share data freely with the research community.

A detailed analysis of differences between Arabidopsis-only, Arabidopsis-and-other-species, and primarily-other-species researchers highlighted some differences:

- Arabidopsis researchers feel TAIR is essential or very useful; most researchers working primarily on other species also chose these options but a significant minority found TAIR useful or somewhat useful.
- Arabidopsis researchers of both types used TAIR as a data source with great frequency (most weekdays or once a week or more). They also used SIGnal/Salk with somewhat less frequency. Researchers working primarily on other species used TAIR with less frequency but still used it often. Many of the other data sources show little difference between the two groups of Arabidopsis researchers, though there is a tendency for multiple-organism researchers to use the other data sources more frequently. Non-Arabidopsis researchers tend toward the "Never" response more than the other groups (except for the TAIR data source, which they use frequently).
- Researchers who worked only on Arabidopsis tend to save more time by using TAIR than those who work on Arabidopsis and other organisms. Researchers who don't work with Arabidopsis still save some time but not as much by working with TAIR.

Summary of Results

Demand for TAIR

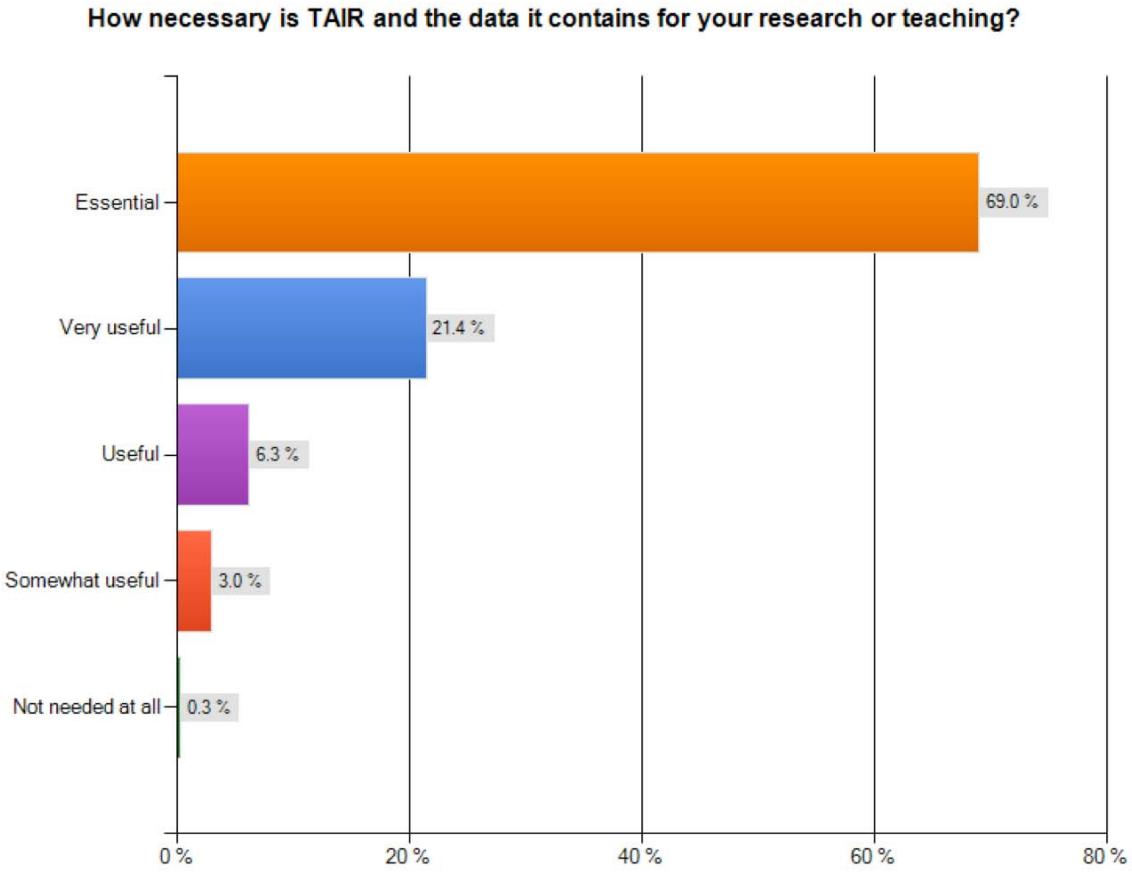
The overwhelming majority of respondents feel TAIR is essential to their work.

Question: How necessary is TAIR and the data it contains for your research or teaching?

Responses:

Essential	69.0%	232
Very useful	21.4%	72
Useful	6.3%	21
Somewhat useful	3.0%	10
Not needed at all	0.3%	1

The median value is Essential at the 99.9% level of confidence.



Lab Size

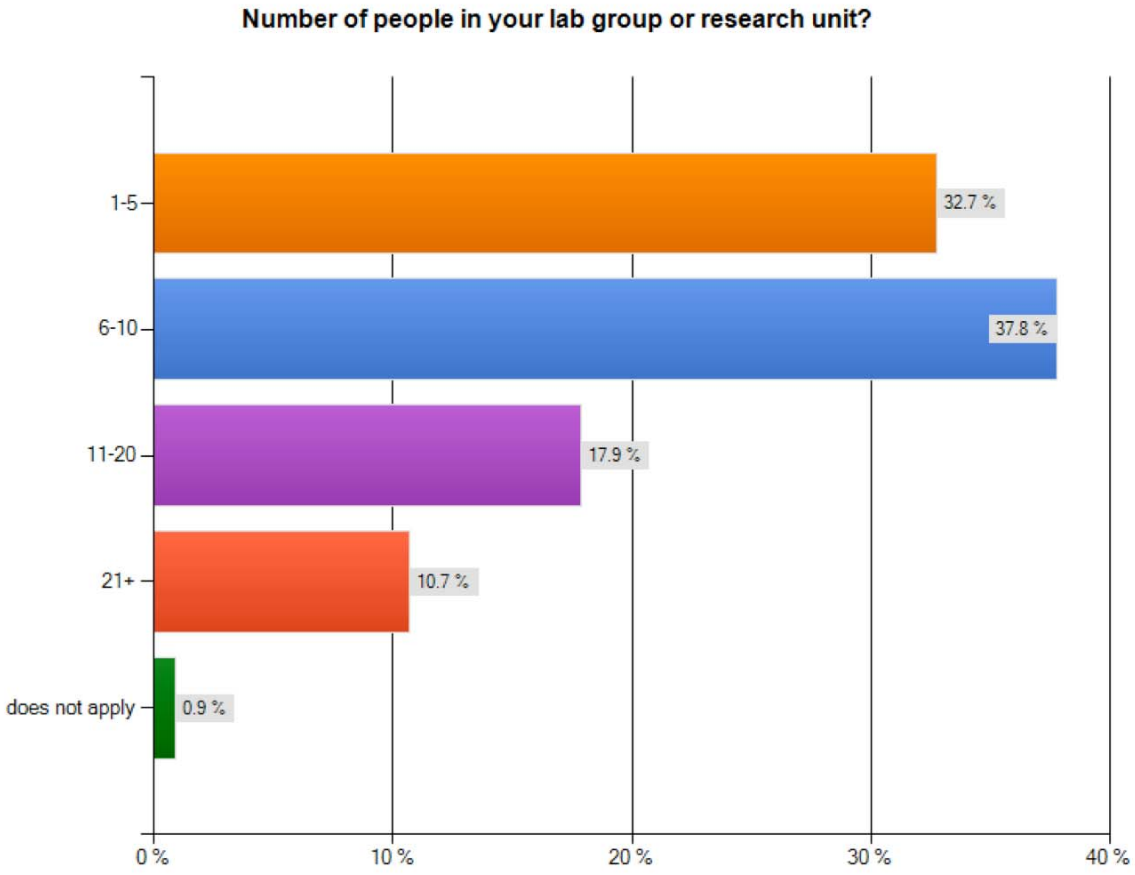
Most respondent labs have between 1 and 10 people.

Question: Number of people in your lab group or research unit?

Responses:

1-5	32.7%	110
6-10	37.8%	127
11-20	17.9%	60
21+	10.7%	36
does not apply	0.9%	3

The lab size median is 6-10, which is significant at the 99.9% level.



Positions

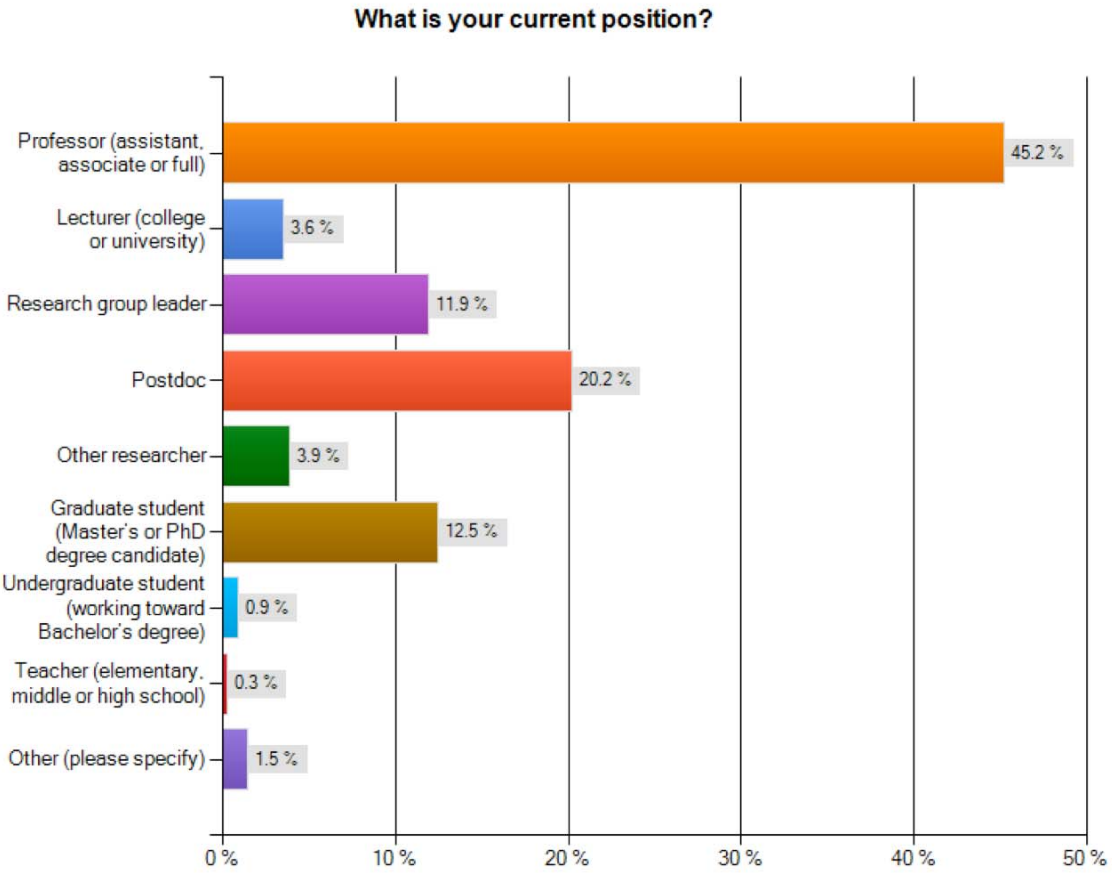
Most respondents are professors or postdocs.

Question: What is your current position?

Responses:

Professor (assistant, associate or full)	45.2%	152
Lecturer (college or university)	3.6%	12
Research group leader	11.9%	40
Postdoc	20.2%	68
Other researcher	3.9%	13
Graduate student (Master's or PhD degree candidate)	12.5%	42
Undergraduate student (working toward Bachelor's degree)	0.9%	3
Teacher (elementary, middle or high school)	0.3%	1
Other (please specify)	1.5%	5

- Research scientist, PI
- senior researcher (National Institute)
- Research associate/Lab manager
- Lab leader
- research assistant



Workplace

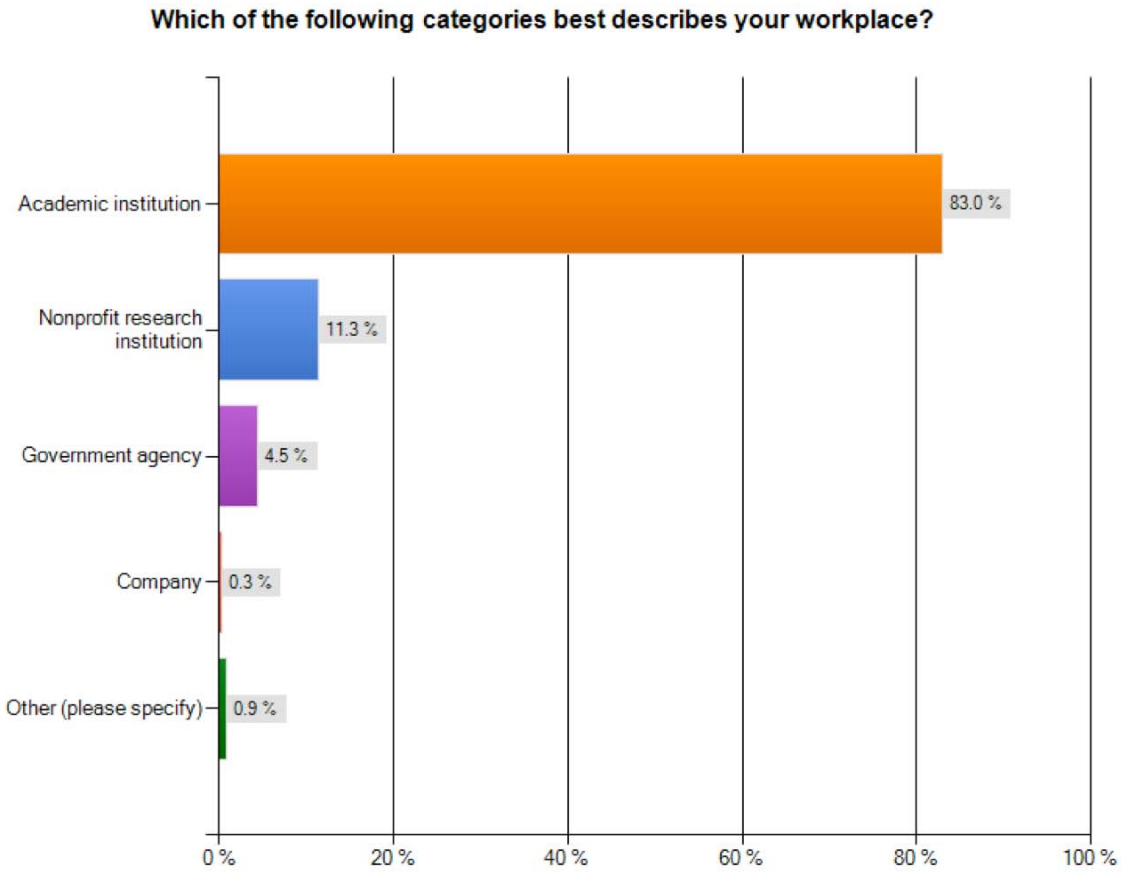
Most respondents work in academic institutions.

Question: Which of the following categories best describes your workplace?

Responses:

Academic institution	83.0%	279
Nonprofit research institution	11.3%	38
Government agency	4.5%	15
Company	0.3%	1
Other (please specify)	0.9%	3

- Nonprofit government-funded research institution
- Government governed university
- university



Geographic Location

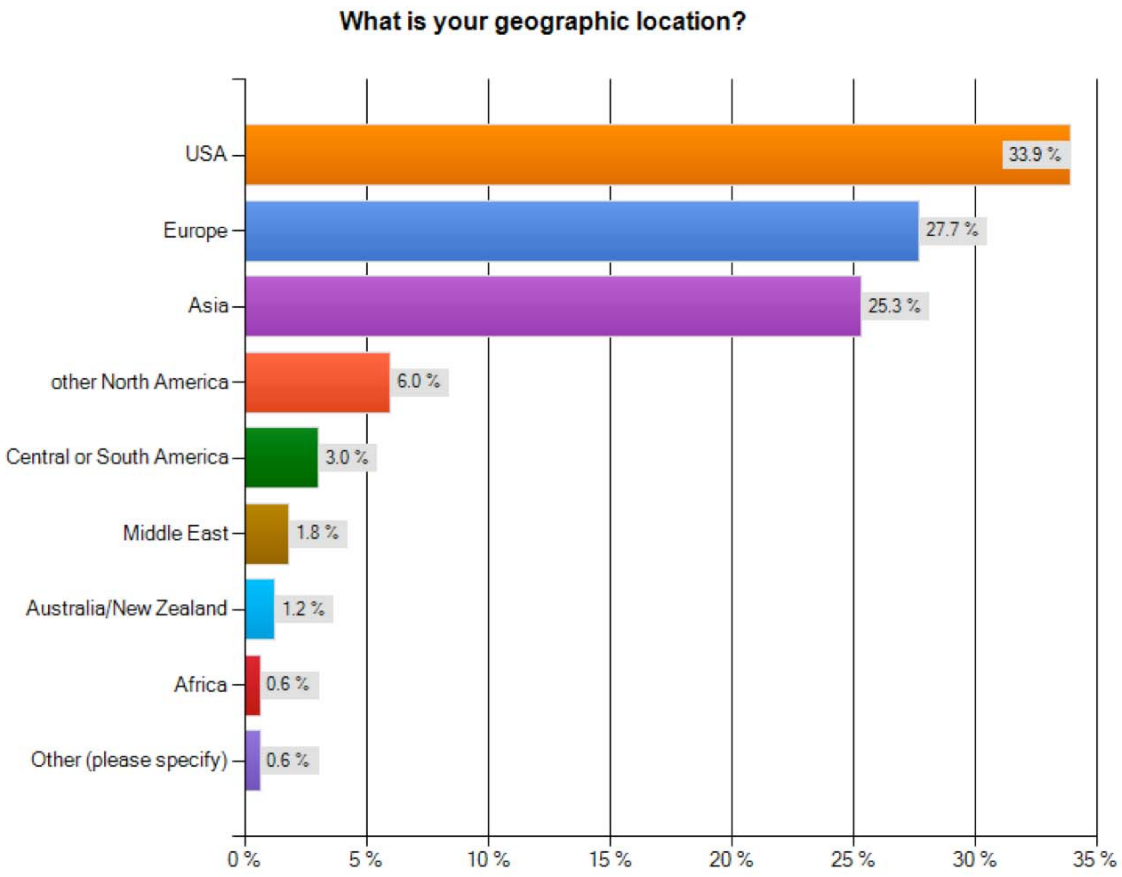
Respondents are global, equally distributed between the USA, Europe, and Asia.

Question: What is your geographic location?

Responses:

USA	33.9%	114
other North America	6.0%	20
Central or South America	3.0%	10
Europe	27.7%	93
Asia	25.3%	85
Africa	0.6%	2
Middle East	1.8%	6
Australia/NewZealand	1.2%	4
Other (please specify)	0.6%	2

- Israel, in Asia. but in some cultures reffered to be located at the middle east
- China



Daily Activity

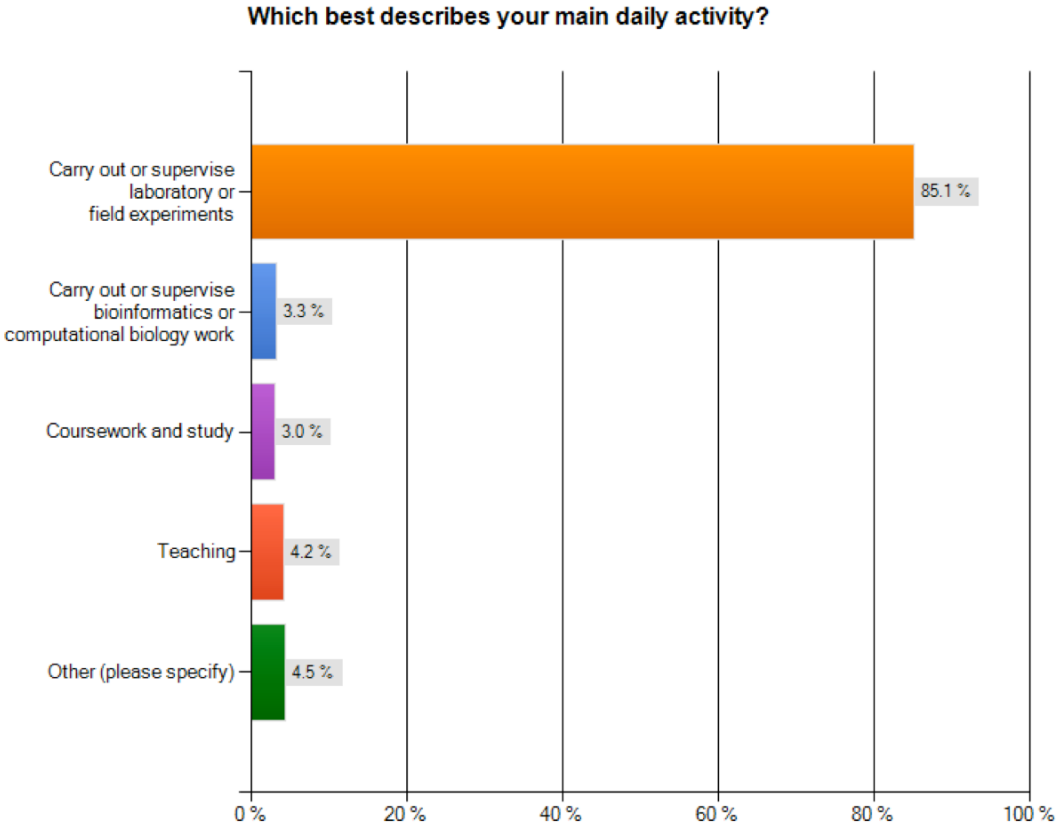
Most respondents carry out or supervise lab or field experiments as their main daily activity.

Question: Which best describes your main daily activity?

Responses:

Carry out or supervise laboratory or field experiments	85.1%	286
Carry out or supervise bioinformatics or computational biology work	3.3%	11
Coursework and study	3.0%	10
Teaching	4.2%	14
Other (please specify)	4.5%	15

- Teaching and lab work
- Both lab experiments and bioinformatic work
- plant biotechnology
- teaching and supervise laboratory and field experiments
- combination of supervising laboratory experiments and computational work and bioinformatics teaching
- both wet lab and bioinformatics
- Administrate core facility
- Coursework and supervision of research laboratory
- head of a research team and lecturer at my university



Research Organisms

Most respondents focus their research on Arabidopsis thaliana as their first or second most important organism. Nicotiana (tobacco) and Oryza (rice) fill out much of the remaining priorities. Respondents do, however, work on a large variety of organisms; 100 respondents entered "Other" species as one of their five priorities.

Question: Which organisms are the focus of your research or teaching? Please choose up to five species in order of importance to your work (excluding any species used only as tools, e.g. to propagate DNA or express proteins).

Responses:

Priority	A. thaliana		Rice		Tobacco		Yeast		Other	
	%	#	%	#	%	#	%	#	%	#
1	71.6	239	2.1	7	0.3	1	0.6	2	25.4	85
2	18.7	56	9.7	29	11.7	35	5.0	15	54.9	165
3	10.1	23	11.5	26	15.0	34	6.2	14	57.2	130
4	3.9	5	11.7	15	14.1	18	10.2	13	60.1	77
5	4.5	4	6.7	6	5.6	5	11.2	10	72.0	64

The percentage is the ratio of those choosing the organism to the total number of responses for the priority; hence, the percentages sum to 100% across the rows. See Appendix B for a complete list of the organisms entered.

The table's Pearson chi-squared goodness-of-fit statistic p-value is 0.000, indicating excellent fit of the table data to the expected distribution.

Use of Data Resources

TAIR, GenBank, and EntrezGene (NCBI) are the most-often-used Arabidopsis data resources, with TAIR and NCBI being used most weekdays and GenBank being used slightly less often. SwissProt, SIGnAL, and JCVI are distant runners up. Most other data resources are used infrequently.

Question: How often do you use the following resources to find Arabidopsis data?

Responses:

Resource	Most weekdays		Once a week or more		Once a month or more		Once a year or more		Never	
	%	#	%	#	%	#	%	#	%	#
TAIR	40.1	132	34.7	114	18.5	61	5.8	19	0.9	3
GenBank/EMBL/DDBJ	25.1	79	39.4	124	24.4	77	7.6	24	3.5	11
Entrez Gene (NCBI)	39.5	122	33.7	104	18.1	56	5.5	17	3.2	10
Swiss-Prot/UniProt	6.6	19	20.3	58	32.2	92	27.6	79	13.3	38
AmiGO/GOOSE	0.4	1	4.6	12	10.4	27	18.5	48	66.2	172
SIGnAL/Salk	11.5	34	26.0	77	29.1	86	18.6	55	14.9	44
JCVI/TIGR	5.5	15	13.6	37	27.6	75	27.6	75	25.7	70
NASC/AtEnsembl	5.4	15	14.1	39	20.2	56	25.3	70	35.0	97
MIPS/MAtDB	4.4	12	11.1	30	19.3	52	27.8	75	37.4	101
AtGDB	3.7	10	14.6	39	20.2	54	20.6	55	40.8	109
KEGG	0.8	2	8.3	22	19.2	51	27.8	74	44.0	117
RIKEN/RARGE	0.7	2	7.3	20	22.0	60	32.2	88	37.7	103
MPSS Plus	2.3	6	5.4	14	13.8	36	21.5	56	57.1	149
ATIDB	2.0	5	6.0	15	14.9	37	16.9	42	60.2	150
Internal database not publicly available	8.3	21	11.5	29	9.5	24	10.7	27	59.9	151
AraCyc	1.5	4	7.5	20	17.2	46	28.1	75	45.7	122

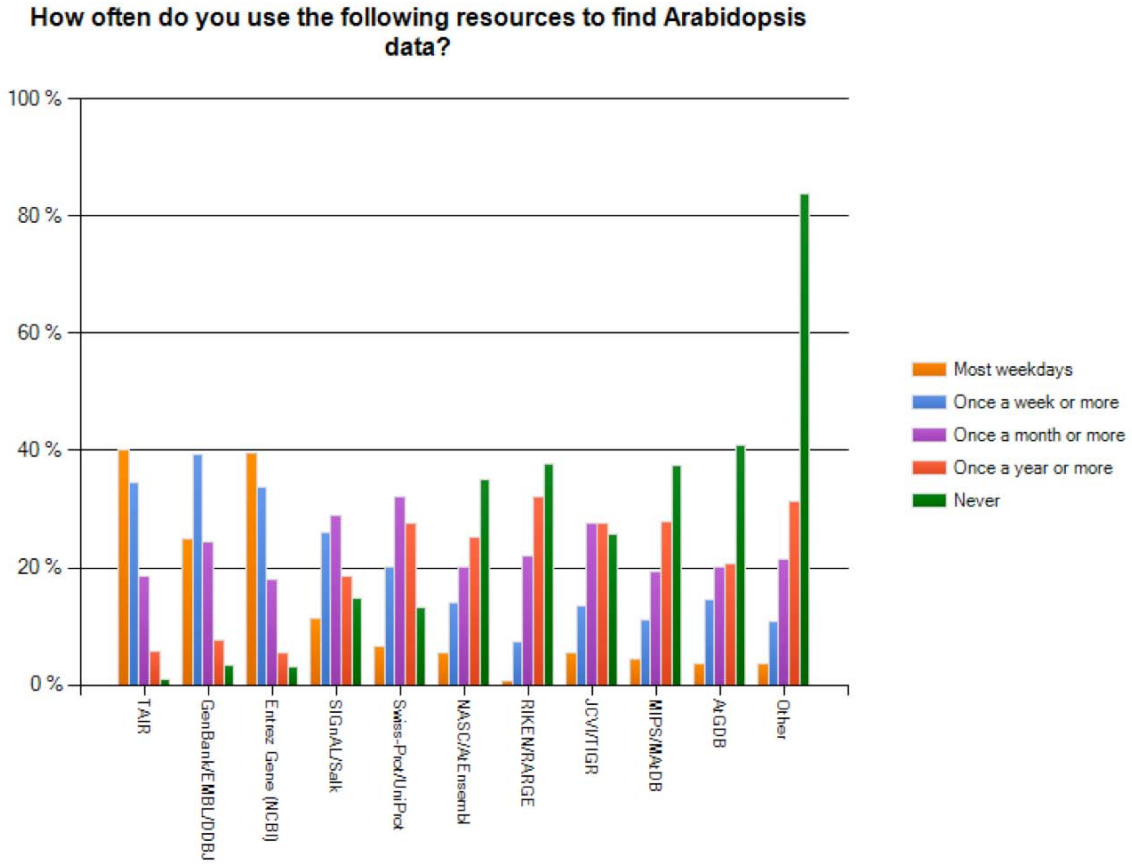
Other resources:

- Cyanobase
- Chloroplast Function Database
- Gnevestigator
- genevestigator
- BAR and Genevestigator, used several times a month
- ATTED, BAR
- Aramemnon (http://aramemnon.botanik.uni-koeln.de/) once a week or more
- When you say "you" I am estimating for my group, rather than me personally
- genevestigator
- DFCI
- ATTED
- EFP Browser at Univ. of Toronto-- I use this most weekdays
- Aramemnon
- SGN
- The Bio-Array Resource for Arabidopsis Functional Genomics http://bar.utoronto.ca/
- MSU-rice, MPSS-rice
- AGRIS

- phytozome
- I use aramemnon one per week
- sgn.cornell.edu every day
- Brachypodium database at Brachypodium.org
- Chromatin database Chromdb

Note that the respondent was not asked to specify how often he or she used the other resource.

The table's Pearson chi-squared goodness-of-fit statistic p-value is 0.000, indicating excellent fit of the table data to the expected distribution.



Time Saved

Most respondents saved between 1 and 8 hours per week by using TAIR, with 11% saving over 8 hours per week.

Question: Can you estimate the amount of time TAIR saves you in a typical week? (In other words, how much additional time would it take you if you needed to get information on Arabidopsis from other public resources?)

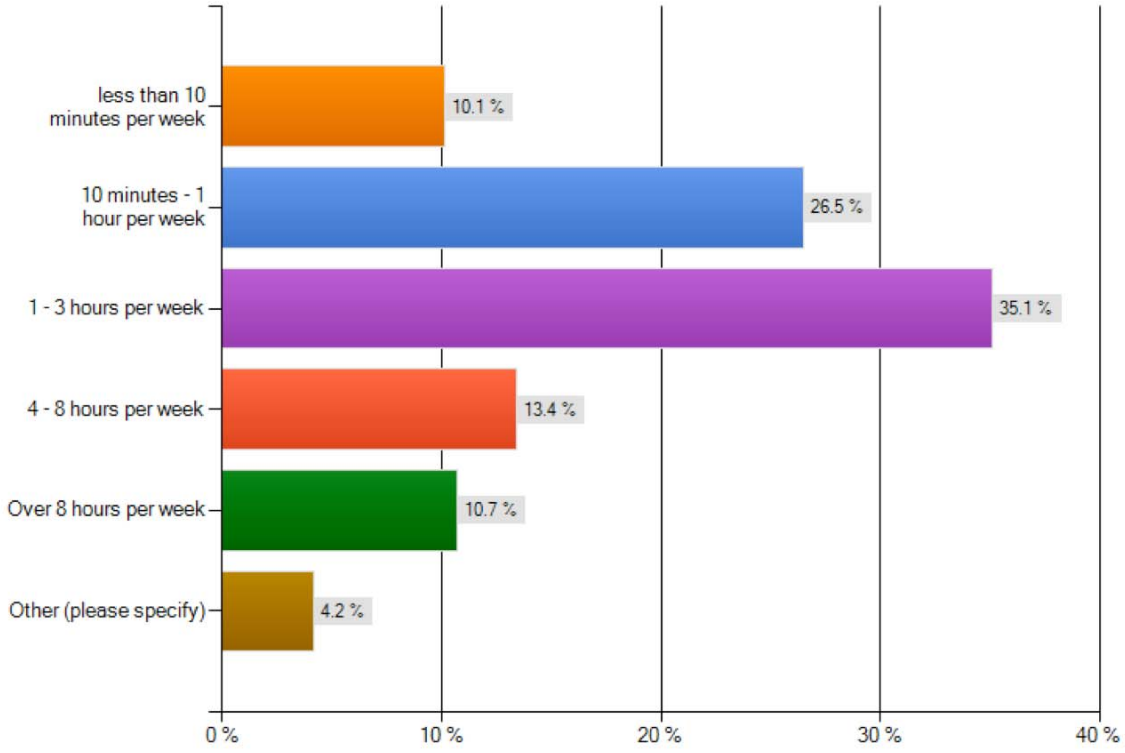
Responses:

less than 10 minutes per week	10.1%	34
10 minutes - 1 hour per week	26.5%	89
1 - 3 hours per week	35.1%	118
4 - 8 hours per week	13.4%	45
Over 8 hours per week	10.7%	36
Other (please specify)	4.2%	14

- I wouldn't even know where to start looking for the information I use TAIR for.
- for each gene search it will save at least 8 hours. usually I'm looking in few genes at once.
- Hard to evaluate. TAIR being the best resource for what I do O have hardly used any other ones. Moreover, I sometime use TAIR not because it is faster, but because it is more complete. For me, time is not the issue, but quality of the information.
- Hard to put a number on-"clone by phone" can save months in a single week
- I really have no idea- hard to know what you would do instead if it wasn't there
- This is too hard to measure-- but it clearly saves a lot of time.
- I can not estimate this for myself- but my students must save several hours per week
- cannot estimate
- not able to estimate
- I don't know
- On infrequent occasions, TAIR saves hours of time due to the available resources.
- Hard to evaluate because I hardly use any other database, as TAIR provides all the information I need and compiles it on the same WEB page. What I value the most in TAIR is no the amount of time that it saves me, but the quality of the information that it delivers.
- Only use periodically
- Not sure/can't say; I use mostly TAIR, because it is excellent and fast; I tend to get everything I need, and fast, from TAIR and its links.

The median time saved is 1-3 hours per week at a 99.9% level of confidence.

Can you estimate the amount of time TAIR saves you in a typical week? (In other words, how much additional time would it take you if you needed to get information on Arabidopsis from other public resources?)



Priorities

Most respondents believe that TAIR should focus on extracting functional annotations and phenotypes from the literature and on updating gene structure of the genome. Lesser priorities included correcting genome sequencing errors and integrating community submissions of gene function, gene structure, and other data into TAIR. A third cluster combines providing sequence data for ecotypes and links to other resources. A fourth cluster includes access to data for other species, providing access to tools developed by other groups (Biomart, Genevestigator, etc.), and building novel tools. Finally, providing a helpdesk is the lowest priority activity.

Question: Which of the following activities do you think should be TAIR's top five priorities?

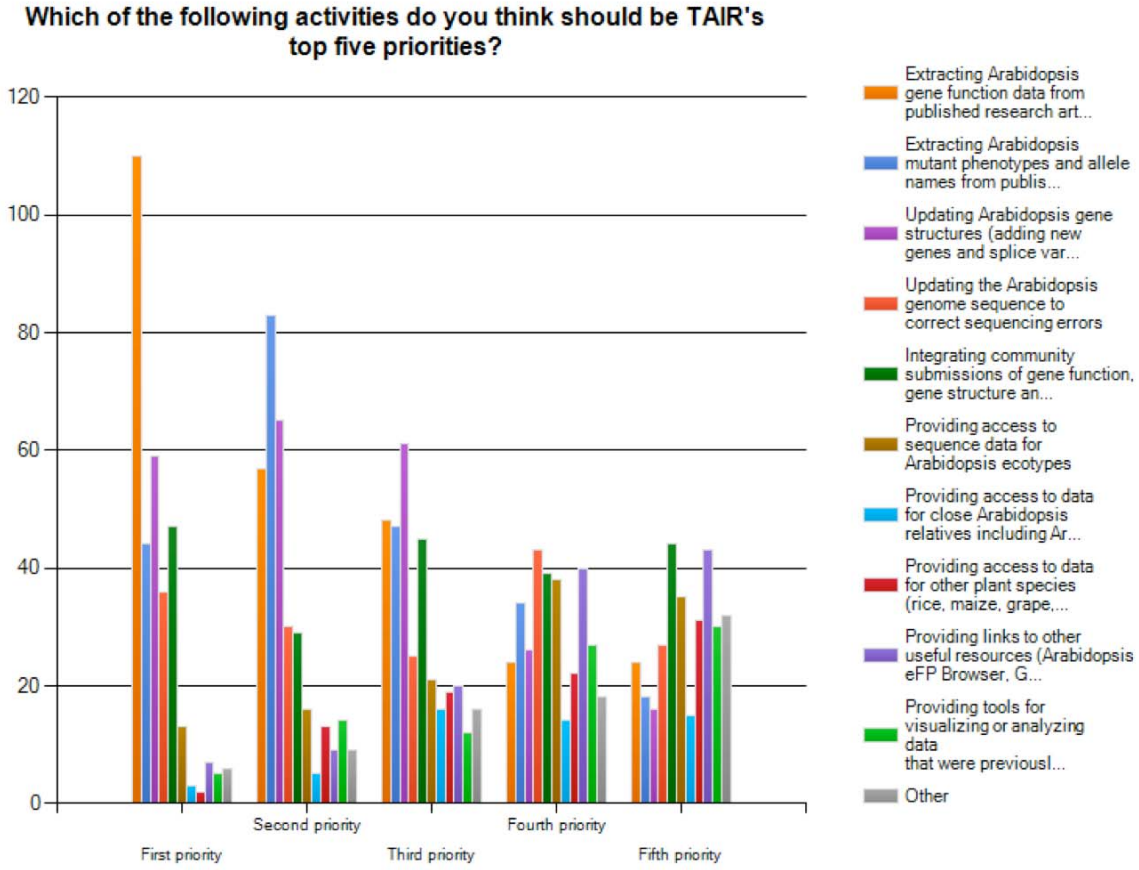
Responses:

Activity	First priority	Second priority	Third priority	Fourth priority	Fifth priority
Extracting Arabidopsis gene function data from published research articles	110	57	48	24	24
Extracting Arabidopsis mutant phenotypes and allele names from published research articles	44	83	47	34	18
Updating Arabidopsis gene structures (adding new genes and splice variants, correcting errors in gene structures)	59	65	61	26	16
Updating the Arabidopsis genome sequence to correct sequencing errors	36	30	25	43	27
Integrating community submissions of gene function, gene structure and other data (including gene families, markers, polymorphisms, gene symbols) into TAIR	47	29	45	39	44
Providing access to sequence data for Arabidopsis ecotypes	13	16	21	38	35
Providing access to data for close Arabidopsis relatives including Arabidopsis lyrata, Capsella rubella and Brassica species	3	5	16	14	15
Providing access to data for other plant species (rice, maize, grape, poplar, etc)	2	13	19	22	31
Providing links to other useful resources (Arabidopsis eFP Browser, Genevestigator, SIGnAL, etc)	7	9	20	40	43
Providing tools for visualizing or analyzing data that were previously developed by other groups (e.g. GBrowse, BioMart, Nbrowse protein interaction viewer)	5	14	12	27	30
Building novel tools to visualize or analyze data in new ways	6	7	13	15	19
Providing an email helpdesk service for locating and using Arabidopsis data and tools	0	2	3	3	13

Other responses:

- How about adding the comment of the general or various experimental methods with Arabidopsis?
- all of the above?
- I can't do this, there are more than 5 high priority activities
- Disregard my answers to this question. I do not feel I use the database enough at this point to address the topic.

The table's Pearson chi-squared goodness-of-fit statistic p-value is 0.000, indicating excellent fit of the table data to the expected distribution.



Access Model Preservation

A majority of respondents feel that it is essential to preserve most aspects of the current TAIR free-access model, with the exception of keeping access free for companies and keeping the site free of advertisements.

Question: Which of the following aspects of TAIR's current free access model do you feel are important to preserve?

Responses:

Aspect	Essential		Very important		Somewhat important		Unimportant		Don't know	
	%	#	%	#	%	#	%	#	%	#
No login required to access TAIR data, facilitating free exploration by researchers and students who don't yet know what data is available in TAIR	67.7	224	18.7	62	9.4	31	3.3	11	0.9	3
TAIR data freely available to companies	12.3	37	12.3	37	20.9	63	34.2	103	20.3	61
TAIR genome releases (new genes, new and updated splice forms) freely propagated to other resources such as GenBank, SIGnAL, UniProt, etc.	49.4	158	37.5	120	10.0	32	0.9	3	2.2	7
TAIR gene function information (gene description, GO annotations, expression pattern, gene symbols, phenotypes) freely propagated to other resources such as GenBank, SIGnAL, UniProt, etc.	54.5	177	32.0	104	9.5	31	2.2	7	1.8	6
Equal access to TAIR data for all users regardless of ability to pay, including researchers from developing countries, educators and students, and less frequent users of Arabidopsis data	59.8	193	24.8	80	8.0	26	3.7	12	3.7	12

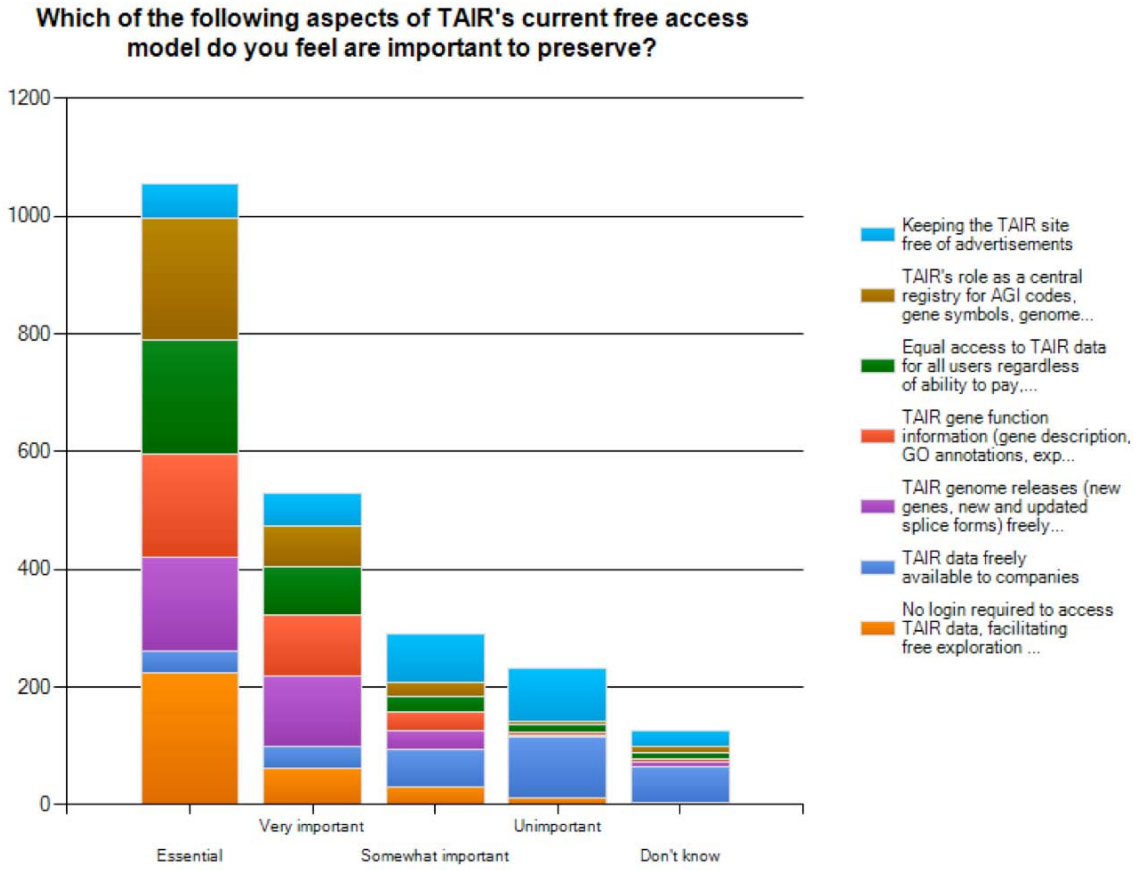
TAIR's role as a central registry for AGI codes, gene symbols, genome sequence, etc., allowing the research community to share and integrate data based on universally accepted identifiers and genome coordinates	65.4	208	22.3	71	7.5	24	1.9	6	2.8	9
Keeping the TAIR site free of advertisements	18.3	57	17.9	56	26.6	83	28.5	89	8.7	27

Comments:

- TAIR could possibly seek funding from resources other than NSF. Advertisements are fine with me.
- Advertisements are not bad. It is sometimes useful.
- I'd much rather have ads than have to pay to access the data. I never would have taken the time to figure out what was there if I had to pay first. I think it would be ok to require a login so that you can charge companies but provide the info free to academics. Though this would discourage new users
- Q2 - depending on size of company = small companies
- essentially TAIR is very good as it is. however, forcing users to log in during their work with TAIR services - even if the services are being kept for free - might be a good idea
- Keeping a constant updating of data and sequence information is Very Important
- TAIR gene function information: GO annotations are not based on demonstrated function, but on sequence similarity to an inferred function, and can therefore be misleading. Links to the other websites showing expression patterns would be better serve users.
- Keeping the Tair site free of advertisement is no important if it's one way to keep the site and the data on it available for the research community.
- I think that a small user fee, perhaps collected annually, would not be a problem, if it helps keep TAIR running. Or setting up a fee structure based on size of group, funding levels, academia vs. industry, etc.
- no response
- If advertisements would help to TAIR financially, it would be fine for me to see commercial information on its pages.
- Companies profit from the research we do in academy and have a bigger budget. They should use a login and pay a fee for using TAIR. It should be free for Academic institutions
- If advertisements are a way to support TAIR and if they are not to prominent (like pop-ups and such kind) it would be ok to implement them.

- Advertisement could contribute to funding. I personally would tolerate advertisement if it helps TAIR.
- I use TAIR to teach an upper division molecular biology class studying the ability of A. thaliana to respond to environmental stress. We use TAIR to identify gene sequences, map primers onto A thaliana genes, link to Salk T-DNA mutants and identify other available resources. We cannot afford to pay for this service and it is an integral part of our lab which is taught to ~50 students each quarter at University of Washington. We appreciate what you do!
- FREE ACCESS IS VERY VERY IMPORTANT, PLEASE DO NOT CHANGE IT!

The table's Pearson chi-squared goodness-of-fit statistic p-value is 0.000, indicating excellent fit of the table data to the expected distribution.



Data Provision

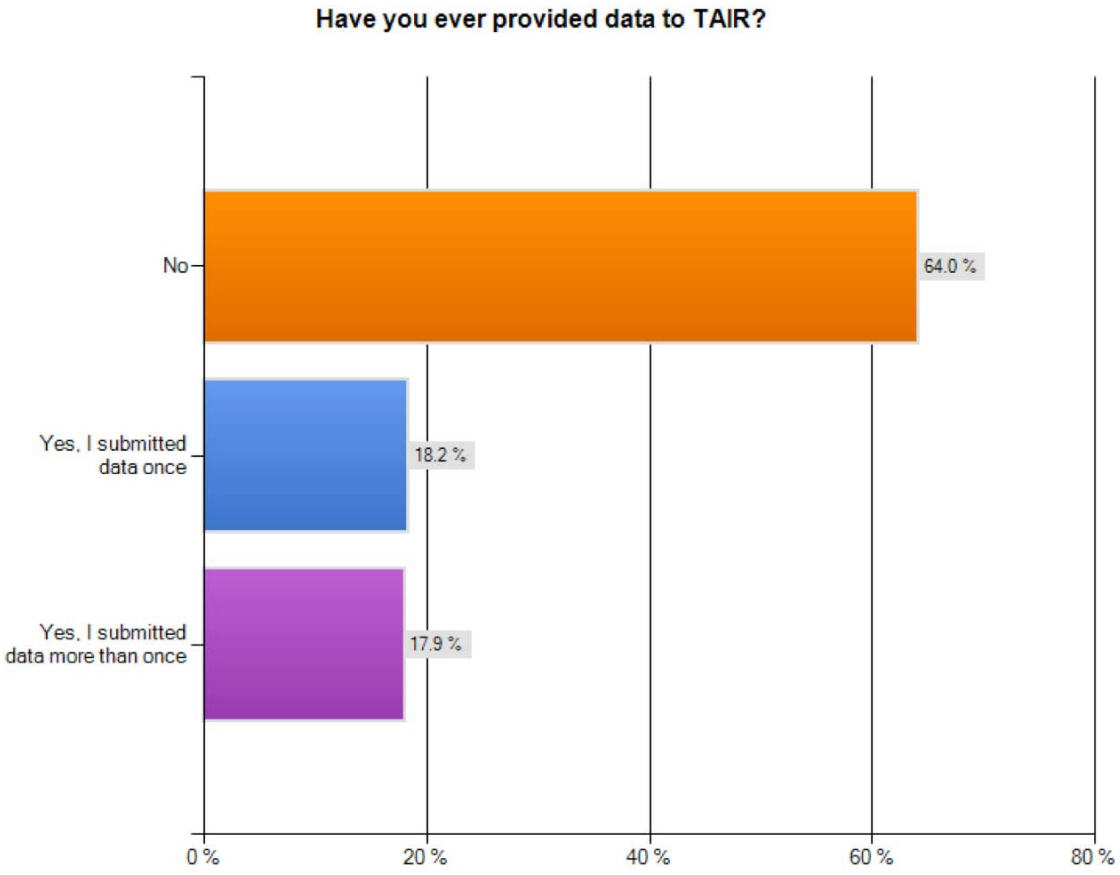
Two-thirds of the respondents have never provided data to TAIR; the other third is equally divided between those who have provided data once and those who have provided data more than once.

Question: Have you ever provided data to TAIR?

Responses:

No	64.0%	215
Yes, I submitted data once	18.2%	61
Yes I submitted data more than once	17.9%	60

The median response is No, which is significant at the 99.9% level.



Data Provision If Not Freely Shared

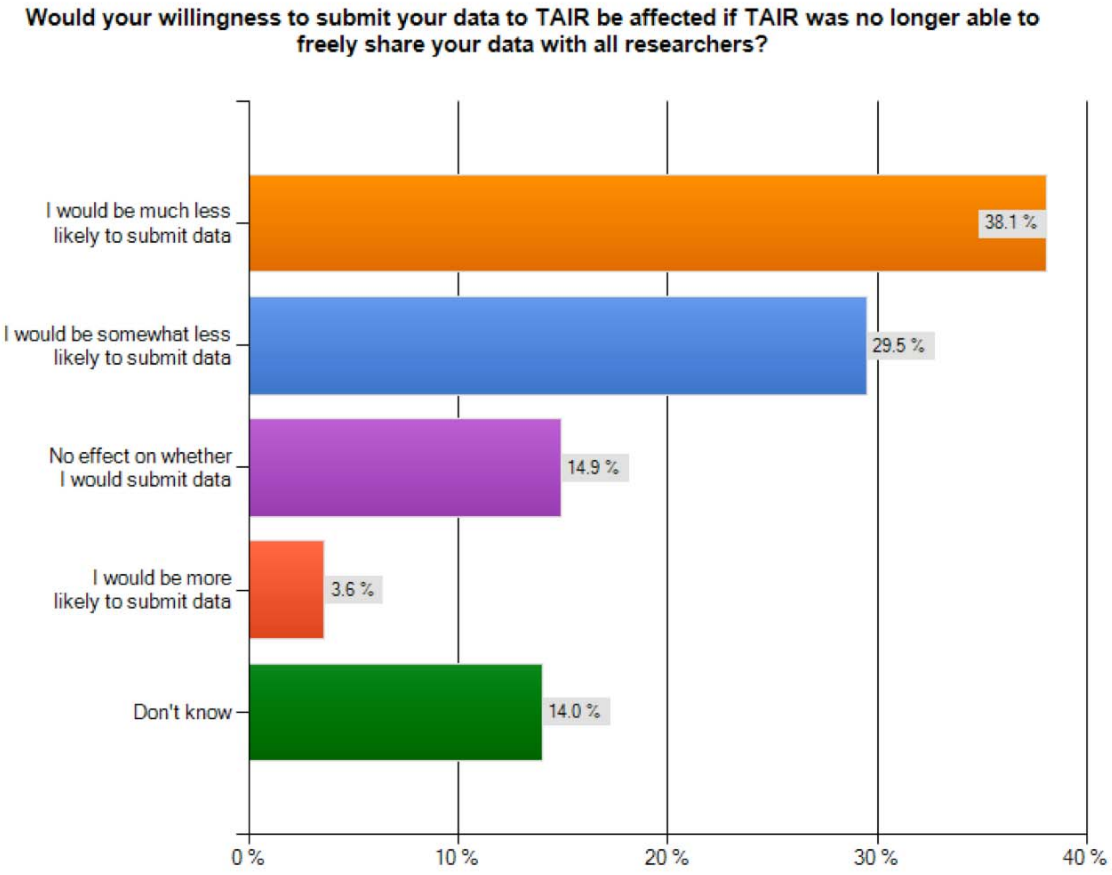
A majority felt that they would be either much or somewhat less likely to submit data if TAIR did not share data freely with the research community.

Question: Would your willingness to submit your data to TAIR be affected if TAIR was no longer able to freely share your data with all researchers?

Responses:

I would be much less likely to submit data	38.1%	128
I would be somewhat less likely to submit data	29.5%	99
No effect on whether I would submit data	14.9%	50
I would be more likely to submit data	3.6%	12
Don't know	14.0%	47

The median response is to be somewhat less likely to submit data, which is significant at the 99.9% level.



Suggestions and Feedback

There were a total of 65 comments giving suggestions and feedback on TAIR priorities and/or future bioinformatics needs of the plant biology community:

- keep up the good work.
- It hasn't been a long time since I use TAIR. It's need some time for me familiar with TAIR.
- I strongly suggest that TAIR get involve in the integration of data from other plants species. The experience is there and a great job have been done. There is many sequencing initiatives why not use TAIR model to integrate many of them. I believe that TAIR could be something greater by integrating many plant species together. Plant biology have been change in a very positive way forever. TAIR have a great responsibility in the very positives changes that we have seen.
- I hope I can have more seeds of one line in one seed stock order. And sometimes the vector's profiles are not enough.
- TAIR is very very very important in plant sciences. I hope it can keep on goining.
- Genome interactive data among the different organisms may be helpful with respect to a particular drug.
- at least. Updating of the existing database is important.
- TAIR has been a reference to me and to my post-grad students in almost all aspects of bioinformatics analysis. I think damage can be really bad if it becomes out date or not free acess anymore.
- No
- It would be fantastic if there were coverage of arabidopsis relatives such as Arabidopsis lyrata
- we would be more likely to submit our data...
- It needs to be maintained as free as possible - research funding is becoming increasingly difficult to obtain and small research groups do not have the funds to pay for access. Keep up the good job
- I think TAIR is a great tool for researchers. It definitely helped me a lot to build up my project on two genes with previously unknown functions/annotations etc.. Without the help of TAIR my task would have been much more difficult. In my opinion it is most important that published data continuous to be freely accessible and is efficiently interconnected and collected in TAIR. Otherwise access to data might be restricted to a privileged minority who can afford it. In my opinion such a scenario would be against the essence of science itself.
- High priority: Extracting Arabidopsis mutant phenotypes and allele names from published research articles
- I really hope that TAIR will be able to continue supporting researchers all over the world in the future. Thanks!

- I greatly appreciate the free access to the data that TAIR provides for everyone. It is an important tool for my work.
- TAIR is not only a wonderful resource for research. I use it extensively in teaching. I give a series of lectures on genomics, the lecture material fully integrates links to TAIR and I impress upon the students the importance of internet based resources and that TAIR and similar resources for other organisms represent a critical research tool
- In my opinionn, data integration (with the development of tools for such integration and for user friendly/fast/intuitive management) are the main tasks that TAIR should not abandon. Institutional fees (rated according to countries and absent for developing countries) is an alternative that I would accept in order to maintain current quality.
- TAIR is great and an excellent example on how generosity among scientists to share data and resources have provided tremendous progress in plant research. A well curated data base with many options like TAIR is also ideal for teaching purposes. It would be a great pity if TAIR could not continue to provide plant researchers with excellent up-to-date and free information. I think commercial advertising is acceptable on the web site, that companises should pay and user fees could be rised for seeds and clones as a way to improve economy before charging the use of the data base. With the plant community support we should be able to keep up the TAIR good work - Yes we can.
- Less bioinformatics that is proven only on statistics and not on biological data. 2) Less data but clear, solid (published would be the best) is much better that a lot of data sustained only by bioinformatics.
- When available, sequencing data from Arabidopsis ecotypes will be fundemental. 2) TAIR interface could be more user friendly (use JGI as a model).
- Nowadays it happens very often that I need to wait so long (several minutes or forever sometimes) to open the webpage. In the past, it would come out just in a second. If this is something you can fix, please help! Thanks! TAIR has been very helpful for my research.
- TAIR has to be kept operational by all means and must remain freely accessible!
- NSF is making a huge mistake is shutting down TAIR. It is clear that this rash decision had not been carefully considered.
- all free will be good.
- Doing an excellent Job and I have one thing to add...Keep going..
- TAIR should pair with NESCENT to develop and maintain data and analysis platforms for comparative (evolutionary) genomics for plants.
- TAIR is the central location for all the molecular genetics work we do in the lab. I also teach my students from this website. Our research completely depends on it.
- TAIR is essential for plant biology. It will be a shame to retard its growth.
- Please offer details for submitted data, such as vector information. I ordered a wrong vector recently due to insufficient details

- 1.Provide gene expression patterns by microarry or the tissue specificity. 2.Share the seeds with NASC or gene trap, provide a widely and quick service for the request of insertion seeds.
- I think that TAIR is a model for bioinformatic information. Today there are so many genomes available- but they are so poorly curated or annotated that many are close to impossible to use or worse they have incorrect annotations or gene structures. The scientific community needs TAIR more than we need more genomes!
- TAIR is essential for plant researchers!!!! If needed go for user-fees.....
- I especially appreciate the generation and distribution of collections of Arabidopsis mutants or transgenes, which sometimes might need their amplification or rendering them homozygous. For instance, I bought the first 7000 homozygous T-DNA lines that we are currently screening for nutritional traits. I hope to still get the second batch of 7000 T-DNA lines. This is the type of function I would like to see most from TAIR. There, it fills a gap that other databases or gene banks do not fulfill. Thanks.
- A research community in specific (Arabidopsis) and in general (Plants) will only be as efficient and succesful at approaching new and more complicated topics in direct relation to its diversity of researchers. This diversity is directly related to the ability of all academics to equally and easily access what information has been previously generated in any form that it may exist. The limitation of the TAIR resource will end up costing research agencies more in future lost efficiency than will be gained by short term funding concepts. Facillitating the ability of the small to moderate labs in Arabidopsis and other plants to have access to highest quality bioinformatics and genomics analysis and visualization tools should be a TAIR priority as this will free these labs and their grants to work on biology.
- Personally, I would suggest setting priorities based on usage patterns, not by surveys. This provides an unbiased way to establish what features are used most often. No one can argue with the actual usage data. I think the decision to cut TAIRs funding is ludicrous and points to a complete lack of vision at NSF. It infuriates me to no end and makes want to say no to ever sitting on an NSF panel again, or reviewing another NSF grant grant again. Having said that, TAIR needs to accept the current writing on the wall and focus on preserving mission critical features (based on usage patterns). It would be a very bad idea to keep underused and costly projects afloat (my impression is that there may be more than a few of these at TAIR)-- so some tough decisions should be made, and sooner rather than later.
- TAIR IS VERY INFORMATIC. WE NEED TAIR. IT WILL BE BETTER TO GET MORE COMPLETE INFORMATIONS.
- I understand the need to balance the costs of maintaining TAIR plus the freedom of the data access. May I suggest following a simple philosophy: The data must be free; it is not yours, but everybody's. The tools and the infrastructure to analyze and present that data, those have costs that can be recovered. Identify the value-added aspects of your service and recover costs through ads, and if that doesn't work, perhaps a variable fee schedule. Best of luck.

- TAIR is highly needed to plant research community and we must find a way how to help to maintain it and support continuation of its development. I think one way is to get agreement with NASC (and other similar non-commercial organizations) to have similar prices for the same seed (or other material) stocks and then to increase price for the stock materials for TAIR/NASC/etc. users. It is better to pay more and have possibility to get somewhere materials required for research than to loose such a valuable source completely. We can not afford this. On the other hand to my opinion it is important to have freely available all bioinformatic sources that TAIR manage, at least for non-commercial organizations, such as University, etc. I think that commercial organizations can be charged for any kind of information obtained from TAIT. This will give additional financial support to TAIR.
- Excellent resource all round for plant biology researchers particularly focusing on Arabidopsis. Most of my research would not be possible without TAIR in its current format.
- TAIR is a very important tool for Arabidopsis scientists. I would not have done all the analysis without having this tool. Therefore, I hope that TAIR will exist further on.
- keep it going.
- It would be helpful to have a link from the home page that explains how many people are involved in maintaining TAIR, what their roles are and how difficult it is to provide the many different and useful resources offered. I am very grateful for the existence of TAIR!
- Building novel tools to visualize or analyze data will be essential for the future. TAIR is the greatest resource website for plant biologists.
- Keep on the good work. Bests
- Just trying to update more frequently the database.
- Tair has done a great job keeping data accurate and up to date. Thanks.
- Keep up with the good work. I am not working with Arabidopsis any more (use to do that for many years) but I visit the site every now and then to check up the information of certain genes. Best Regards, Biljana Stangeland
- There are many databases to research plant biology. However, it is not convenience that we access each database from the different Web sites. One central Web site that link to various databases is very useful. Please construct the friendly database for amateur of bioinformatics.
- use database more easy,such as tool
- KEEP TAIR FREE FOR ACADEMIC INSTITUTIONS.
- TAIR is an essential part of our daily research life and its contribution to our research activities is immense. If funding is shinking, I think TAIR can consider charging the profit-making companies for sharing Arabidopsis information. Besides, TAIR can consider putting up advertisement in the website and thereby geberate funds.
- No.Thanks

- The wealth of sequence data being generated (genome and expression data) needs to be easily minable through TAIR or other freely available sites (e.g Weigel site), otherwise this will only benefit research groups with established bioinformatics setups/support. Commercial access to TAIR should be charged for - no argument!
- The main problem we have with access to the TAIR website is speed. The search function and linking through to gene model/locus information is very slow and frustrating.
- gene function
- TAIR is essential and should be freely available for everyone.
- I hope TAIR is getting better as handling more.
- I believe that free access must be essential for academics and researchers from developing countries. And I also know that it is very dificult to distinguish the differences between academics and non-academics. So, this site must be kept free. But, I think that TAIR and researchers may accept some advertisements conserning our research.
- It is essential to keep a central information resource for plant biology and maintain its high standard to develop novel ways of visualisation and easy access of information
- Integrating sequencing data from Arabidopsis ecotypes should be the absolute priority.
- Accurate annotation of gene function(s) critical.
- I would like to see all plant data integrated into a new, one-stop-shop entity--The Plant Information Resource--TPIR. This entity could oversee TAIR, TRIR (rice), TMIR (maize), TTIR (tomato), TBIR (beet), etc. and be administered by representatives of all groups.
- The plant biology community needs TAIR to freely communicate with one another.
- Glad we have TAIR facitity. Saves us alot of time and helps tremendously in carrying out research.

Detailed Analysis

Arabidopsis Researchers Viewpoint

To explore the relationship between respondents' research interest and their opinions on other questions, the analysis here divides the respondents into three groups:

1. The respondents that chose Arabidopsis as their most important research organism and chose no other organisms as of any importance (31, 9%)
2. The respondents that chose Arabidopsis as their most important research organism but did rate other organisms as important (214, 62.2%)
3. All other respondents (99, 28.8%)

The following analyses are of relationships that are both significant and interesting. For example, there is little significant difference between researchers in terms of geographic location, workplace type, job titles, or lab size, and those questions do not appear in the following analysis.

Responses to the question on TAIR priorities and to the question on TAIR's free access model did not exhibit significant differences between the respondent groups.

TAIR Necessity

All of the Arabidopsis-only researchers find TAIR either essential (84%) or very useful (16%). Arabidopsis researchers who deal with other organisms are nearly identical in their evaluation (82% and 15%). Non-Arabidopsis researchers, however, were much less in need of TAIR. The only respondent who felt TAIR was not needed at all is a Non-Arabidopsis researcher.

In response to the question "How necessary is TAIR and the data it contains for your research or teaching?", the following table shows how Arabidopsis researchers responded. The top number in the cell is the count of respondents; the second number is the percentage within the row; the third number is the percentage within the column; and the fourth number is the percentage of the cell within all the cells in the table. The Pearson chi-squared test for goodness-of-fit is significant with a p-value of 0.000.

TAIR Necessity	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Essential	25	172	35	232
	10.78	74.14	15.09	100.00
	83.33	82.30	36.08	69.05
	7.44	51.19	10.42	69.05
Very Useful	5	32	35	72
	6.94	44.44	48.61	100.00
	16.67	15.31	36.08	21.43
	1.49	9.52	10.42	21.43
Useful	0	4	17	21
	0.00	19.05	80.95	100.00
	0.00	1.91	17.53	6.25
	0.00	1.19	5.06	6.25
Somewhat Useful	0	1	9	10
	0.00	10.00	90.00	100.00
	0.00	0.48	9.28	2.98
	0.00	0.30	2.68	2.98
Not Needed at All	0	0	1	1
	0.00	0.00	100.00	100.00
	0.00	0.00	1.03	0.30
	0.00	0.00	0.30	0.30
Total	30	209	97	336
	8.93	62.20	28.87	100.00
	100.00	100.00	100.00	100.00
	8.93	62.20	28.87	100.00

Data Sources

Researchers working only or primarily on Arabidopsis used TAIR as a data source with great frequency (most weekdays or once a week or more). They also used SIGnal/Salk with somewhat less frequency. Researchers working primarily on other species used TAIR with less frequency but still used it often. Many of the other data sources show little difference between the two groups of Arabidopsis researchers, though there is a tendency for multiple-organism researchers to use the other data sources more frequently. Researchers working primarily on other species tend toward the "Never" response more than the other groups (except for the TAIR data source, which they use frequently).

In response to the question "How often do you use the following resources to find Arabidopsis data?", the following tables show how the three classes of researcher responded. For clarity, the tables present only the counts. The top number is the count and the bottom number is the percentage within the column. All tables presented have a Pearson chi-squared goodness-of-fit p-value of 0.000 - 0.050, indicating a significant fit with the presumed distribution at least the 95% level of confidence. If a data source does not appear, the chi-squared test indicates no significant fit (making any conclusions from the table inappropriate) (for example, the tabulation for GenBank has a p-value of 0.199, with values distributed relatively randomly through the table; the tables for NCBI and SwissProt are similar).

TAIR Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	15	106	11	132
	51.72	51.21	11.83	40.12
Once a Week or More	12	74	28	114
	41.38	35.75	30.11	34.65
Once a Month or More	2	25	34	61
	6.90	12.08	36.56	18.54
Once a Year or More	0	1	18	19
	0.00	0.48	19.35	5.78
Never	0	1	2	3
	0.00	0.48	2.15	0.91
Total	29	207	93	329

SIGnAL/Salk Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	3	29	2	34
	12.00	15.34	2.44	11.49
Once a Week or More	9	62	6	77
	36.00	32.80	7.32	26.01
Once a Month or More	9	58	19	86
	36.00	30.69	23.17	29.05
Once a Year or More	2	29	24	55
	8.00	15.34	29.27	18.58
Never	2	11	31	44
	8.00	5.82	37.80	14.86
Total	25	189	82	296

AmiGO/GOOSE Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	1	0	0	1
	5.56	0.00	0.00	0.38
Once a Week or More	0	7	5	12
	0.00	4.24	6.49	4.62
Once a Month or More	1	22	4	27
	5.56	13.33	5.19	10.38
Once a Year or More	1	32	15	48
	5.56	19.39	19.48	18.46
Never	15	104	53	172
	83.33	63.03	68.83	66.15
Total	18	165	77	260

AraCyc Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	0	4	0	4
	0.00	2.35	0.00	1.50
Once a Week or More	0	17	3	20
	0.00	10.00	3.85	7.49
Once a Month or More	1	36	9	46
	5.26	21.18	11.54	17.23
Once a Year or More	4	50	21	75
	21.05	29.41	26.92	28.09
Never	14	63	45	122
	73.68	37.06	57.69	45.69
Total	19	170	78	267

MIPS/MAtDB Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	1	10	1	12
	4.55	5.92	1.27	4.44
Once a Week or More	2	24	4	30
	9.09	14.20	5.06	11.11
Once a Month or More	4	38	10	52
	18.18	22.49	12.66	19.26
Once a Year or More	8	49	18	75
	36.36	28.99	22.78	27.78
Never	7	48	46	101
	31.82	28.40	58.23	27.41
Total	22	169	79	270

NASC/AtEnsembl Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	1	13	1	15
	4.55	7.34	1.28	5.42
Once a Week or More	4	33	2	39
	18.18	18.64	2.56	14.08
Once a Month or More	5	39	12	56
	22.73	22.03	15.38	20.22
Once a Year or More	2	46	22	70
	9.09	25.99	28.21	25.27
Never	10	46	41	97
	45.45	25.99	52.56	35.02
Total	22	177	78	277

RIKEN/RARGE Frequency	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Most Weekdays	2	0	0	2
	9.52	0.00	0.00	0.73
Once a Week or More	0	15	5	20
	0.00	8.82	6.10	7.33
Once a Month or More	2	49	9	60
	9.52	28.82	10.98	21.98
Once a Year or More	9	52	27	88
	42.86	30.59	32.93	32.23
Never	8	54	41	103
	38.10	31.76	50.00	37.73
Total	21	170	82	273

Time Saved

Researchers who worked only on Arabidopsis tend to save more time by using TAIR than those who work on Arabidopsis and other organisms. Researchers who don't work with Arabidopsis save some time but not much by working with TAIR.

In response to the question "Can you estimate the amount of time TAIR saves you in a typical week?" the following table shows how the three classes of researcher responded. The top number is the count of respondents, and the second number is the percentage of the column count (which shows the differences between the different groups of researchers). The Pearson chi-squared test for goodness-of-fit is significant with a p-value of 0.000.

Time Saved	Only Arabidopsis	Primarily Arabidopsis	Primarily Other Species	Total
Less Than 10 Minutes	0	9	25	34
	0.00	4.31	25.77	10.12
10 Minutes to 1 Hour	8	50	31	89
	26.67	23.92	31.96	26.49
1-3 Hours Per Week	7	83	28	118
	23.33	39.71	28.87	35.12
4-8 Hours Per Week	4	37	4	45
	13.33	17.70	4.12	13.39
Over 8 Hours Per Week	5	25	6	36
	16.67	11.96	6.19	10.71
Other	6	5	3	14
	20.00	2.39	3.09	4.17
Total	30	209	97	336
	100.00	100.00	100.00	100.00

The following table shows the median for each group and its 95% confidence interval.

Group	Median	Lower 95% Bound	Upper 95% Bound
Only Arabidopsis	3	2	3
Primarily Arabidopsis	3	3	3
Primarily Other Species	2	2	2

Appendix A: Data Quality

Most of the quality issues in the data come from the "other" categories, when the respondent classifies themselves with a more specific version of one of the allowed choices. I have reclassified the response into one of the allowed choices where it seems reasonable to do so.

For the question on position, the other responses were reclassified as follows:

Other Position Response	Reclassified As Response
Research scientist, PI	Research Group Leader
Senior researcher (National Institute)	Research Group Leader
Research associate/Lab manager	Other Researcher
Lab leader	Research Group Leader
Research assistant	Other Researcher

For the question on workplace, the other responses were reclassified as follows:

Other Workplace Response	Reclassified As Response
University	Academic institution
Nonprofit government-funded research institution	Nonprofit research institution
Government governed university	Academic institution
university	academic institution

For the question on geographic location, the other responses were reclassified as follows:

Other Geographic Location Response	Reclassified as Response
Israel, in Asia. but in some cultures referred to be located at the middle east	Middle East
China	Asia

For the question on daily activity, the other responses were not reclassified. Several of the "other" responses were combinations of the allowed values, but all the responses were distinct in themselves from any single allowed value.

Appendix B: Research Organisms

Respondents listed the following research organisms as alternative choices to the ones listed in the survey. Unless otherwise indicated, each organism was mentioned by only one respondent. Multiple mentions appear as a number in parentheses after the species name.

Agave and Fructan accumulating plants
Artemisia annua
Bacillus thuringiensis, carnivorous plants, E. coli
bacteria, e.g. Bacillus subtilis
bamboo and orchid
Beneficial bacteria and insects
Beta vulgaris (beet)
Beta vulgaris, Lactuca sativa, Cucurbitaceae, Trialeurodes and Bemisia
Brachypodium distachyon (3)
BY-2 cell line
Camellia sinensis
Cassava
Catharanthus roseus
Chlamydomonas reinhardtii (2)
Chrysanthemum
coffea ssp
Common bean (2)
Comparative genomics in angiosperms generally, asterids in particular
cucumber
Cucurbitaceae and Aquifoliaceae
Cyanobacteria
cyanobacterium Synechocystis sp. PCC6803
E. coli (4)
Eucalyptus (2)
Eucalyptus, Psychotria
first other=many bacteria and archaeal species, second other=liliopsida
Fragaria x ananassa
Gerbera
goat grasses, hazelnuts, grass pea
grass pollens
Halophyte Salicornia europaea
hops
Impatiens balsamina, Proteaceae
Italian ryegrass
Jatropha curcas
Latuca sativa, Spinacia oleracea
Liverwort
Lotus japonicus (2)
Marine diatoms, Fragaria (strawberries)
Marrubium
medicinal plants

Melilotus sp.
Mesembryanthemum crystallinum
Morus sp. and Maclura pomifera
moss and liverwort
Most: Synechocystis sp. PCC 6803; Second: Thermosynechococcus sp. BP-1
oat (avena sativa)
Orchid
Other (secind): Antirrhinum; Other (third): Lettuce
Other = Chlamydomonas reinhardtii
P.patens
Papaver species (poppies)
pea
pea for chloroplast isolation and biochemical work, E. coli for protein expression
pea, bean (vulgaris)
pea, E.coli
Pepper (2)
Petunia
Physcomitrella patens (3)
Pinus spp., Solanum tuberosum
Pinus taeda, P. palustris, Caretta caretta, Chelonia mydas, Picoides borealis
Piper species (Black pepper and its relatives).
Pisum sativum
Plant viruses
plants in general
Plumbago
Potato
potato, Ginkgo
Primula vulgaris and Petunia inflata
Proteaceae, Xerophyta humilis
Pseudomonas syringae, a bacterial plant pathogen
pumpkin (C. max.)
rat
Salicornia europaea L.
sandalwood and banana
Selaginella moellendorffi (2)
sinapis alba
Solanaceae
Stenocereus gummosus, Pachycereus pringlei, Ferocactus peninsulae
stone fruit (peach, plum, almond)
Strawberry
sugarcane
teaching = soybean; research = Trifolium, Physcomitrella, Arabidopsis
Thellungiella halophila

Tomato
trypanosoma and leishmania
Ustilago maydis
Willow (Salix sp.)