AgBioData Consortium

2022 Survey of Genomic, Genetic, and Breeding (GGB) Database Stakeholders

Summary of Baseline Data

V1.4 (7/21/2022)

Survey Sample, Participant Characteristics, and Familiarity and Experience with GGB Databases

A total of 80 usable survey responses were received during Spring 2022, with complete or mostly complete responses to the substantive questions that were asked after the initial self-descriptors at the beginning of the survey. Sample sizes for specific questions may vary slightly since a few people may have left particular questions unanswered. In some of the following tables, reported percentages may not sum to 100 due to rounding.

The responding group of stakeholders was not a random sample of all stakeholders, since the invitation to participate was circulated widely and those who responded were self-selected, i.e. they chose to participate. We can't know the extent to which this group may be representative of all stakeholders, so we can't confidently generalize the responses of this group to the entire population of GGB database stakeholders. We can, however, be confident that we know with reasonable clarity what this group of 80 GGB database stakeholders reported about their experiences, opinions, and recommendations with regard to these databases.

Almost half of respondents (49%) reported working in a university that offers related PhD degrees; 20% reported working in a Land Grant University State Agricultural Experiment Station (see Table 1). About a sixth of respondents (16%) reported working for USDA and 11% reported working for a non-profit organization. Others reported working in a primarily undergraduate institution (10%), a private company or industry organization (4%), or for other types of institutions. As displayed in Table 2, most reported their primary professional role as being a research scientist (60%), post-doctoral intern (9%), or bioinformatics professional (8%).

Most respondents (85%) reported a professional focus on plants, including major plant crops (28%), horticultural specialty crops (19%), and plants grown for other purposes besides human consumption (9%); see Table 3. Animals were a primary focus for 9% of respondents, including major livestock animals (5% of respondents) and minor livestock animals (6%). Pests, diseases, physiological stressors, and other threats were a focus for 18% of respondents,; 15% reported working on understanding wild organisms not directly used in agriculture. Respondents could indicate more than one area of focus.

The survey asked participants to rate their interest in or familiarity with a list of 34 specific GGB databases; see Table 4. Respondents were most familiar with certain specific databases such as TAIR, Gramene, GRIN, MaizeGDB and Solanacea Genomics Network; for each of these, more than a quarter of survey participants reported being "somewhat familiar," "very familiar," or having "used this a lot and could teach others to use it." Other databases that were rated as having similar levels of familiarity by 15 percent or more of respondents included Soybase, Planteome, AgBase, Genome Database for Rosaceae, and GrainGenes. Table 5 lists participant recommendations for other specific databases that should be included in the next iteration of the survey, along with other related comments.

As displayed in Table 6, over half of participants had not tried to submit data to a GGB database, while 90 percent or more had attempted to search for or retrieve data from a GGB database, and 76 percent had attempted to reuse data that had been retrieved from a GGB database. Approximately 20 percent reported that they had attempted to perform these functions and found it to be impossible, very difficult, or not easy to do in a satisfactory way.

Table 1. Organizational Affiliations of Survey Respondents

Which terms best describe your organization? *(mark all that apply)*	
16%	US Department of Agriculture (USDA)
20%	Land Grant University State Agricultural Experiment Station (SAES)
49%	University offering related PhD degrees
2%	University offering related Masters (but not PhD) degrees
10%	Primarily undergraduate institution (PUI)
2%	Minority-serving institution (MSI)
1%	Historically Black College or University (HBCU)
4%	Private company or industry organization
11%	Nonprofit organization
6%	Other: *"International organization for R&D" "Non US government" "SAAS"*

Table 2. Primary Professional Role of Survey Respondents

Please indicate your primary role:	
60%	Research scientist
9%	Post-doc
8%	Bioinformatics professional
2%	Technical staff member
2%	Pathologist, physiologist, entomologist, or other related scientific professional
4%	Graduate student
1%	Professional plant breeder
1%	Undergraduate student
0%	Food and agriculture industry professional
0%	Professional animal breeder
12%	Other: *"Assistant Professor" "Retired" "Data Analyst" "Faculty" "Instructional faculty" "Librarian/Data curation services" "PI" "Professor" "Professor (teaching and research)" "Professor"*

Table 3. Professional Focus of Survey Respondents

Please indicate your primary area(s) of professional focus:	
85%	Plants
9%	Animals
26%	Major plant crops grown as food for people on a large percentage of farm land (e.g. maize, soybeans, rice etc.)
19%	Horticultural, specialty, or other plant crops grown as food for people on a smaller percentage of farm land (e.g. vegetables, fruits, nuts, etc.)
9%	Plants grown for animal feed, fiber, lumber, industrial use, or ecosystem services/cover crops
5%	Major livestock animals grown as food for people on a large percentage of ranch/farm land (e.g. beef or dairy cattle, pork, chicken, turkey)
6%	Minor livestock animals such as sheep or goats; fish, shellfish or other aquatic animals raised as food for people; honey bees, etc.
15%	Wild organisms not directly used in agriculture (as potential germplasm sources for agriculture or for understanding biology and ecosystems independent of potential agricultural applications)
18%	Pests, diseases, physiological stressors, other threats to agriculture or ecosystems
6%	Other: *"Edible insects" "Model plants" "Pet animals which consume resources" "Theory and methods: data science and bioinformatics" "Undergraduate education"*

Table 4. Interest in and Knowledge of Specific GGB Databases

Please indicate your familiarity with or interest in specific genomics, genetics and breeding databases:	I don't know much about this and I don't think it's relevant for me	I don't know much about this but I'm interested, it might be useful	I'm a little familiar with this	I'm somewhat familiar with this	I'm very familiar with this	I have used this a lot and could teach others to use it	*Sum: somewhat familiar, very familiar, or have used a lot and could teach others*
AgBase	37 %	29	13	15	4	1	*20*
Agroportal	42	41	7	5	3	3	*11*
Alfalfa Toolboxb	70	21	4	5	-	-	*5*
Animal QTLdb -- Animal Quantitative Trait Loci (QTL) Database	82	10	1	3	3	1	*7*
BGD – Bovine Genome Database	84	7	3	3	4	-	*7*
CassavaBase	49	26	14	8	3	-	*11*
CGD – Citrus Genome Database	54	29	11	4	3	-	*7*
Citrusgreening	66	26	5	1	1	-	*2*
CottonGen	59	21	9	8	3	-	*11*
CuGenDB	67	22	7	1	1	1	*3*
GDR – Genome Database for Rosaceae	45	25	14	9	4	3	*16*
GDV – Genome Database for Vaccinium	64	25	6	3	3	-	*6*
GrainGenes	42	26	16	10	5	1	*16*
Gramene	25	20	15	24	12	4	*40*
GRIN – Germplasm Resources Information Network	27	18	14	13	21	6	*40*
HWG – Hardwood Genomics Project	60	27	6	1	5	-	*6*
HGD – Hymenoptera Genome Database	74	21	3	-	3	-	*3*
i5K Workspace@NAL	70	18	1	8	1	1	*10*
KitBase	76	24	-	-	-	-	-
LIS – Legume Information System	49	29	12	5	3	3	*11*
MaizeGDB	32	18	17	14	6	12	*32*
MusaBase	61	26	6	4	3	-	*7*
PeanutBase	55	27	9	5	1	3	*9*
Planteome	30	31	17	10	6	5	*21*
PulseDB	55	26	8	9	1	-	*10*
SGN – Solanaceae Genomics Network	32	18	22	10	15	3	*28*
SoyBase	41	22	15	9	9	4	*22*
SweetPotatoBase	61	26	7	4	3	-	*7*
T3 – The Tritaceae Toolbox	51	31	4	9	4	1	*14*
TAIR -- The Arabidopsis Information Resource	8	5	10	13	33	32	*78*
TreeGenes	49	32	9	3	5	1	*9*
VectorBase	58	29	8	4	1	-	*5*
WheatIS	49	34	8	8	1	-	*9*
YamBase	64	26	4	3	3	-	*6*

Note. N = 80. Each row contains the proportion (percentage) of participants who gave each response to the question; the last column contains the proportion who responded with any one of the three highest levels of familiarity. Row percentages may not add to exactly 100% due to rounding.

Table 5. Recommendations for Inclusion of Specific GGB Databases and Related Comments

Are there other GGB databases that we should include in this list? If so, how familiar are you with those? Please share any other comments to help us understand your ratings above or your general interest in and knowledge of these databases.
▪ *Banana Genome Hub, Rice Genome Hub, Germinate and others but from the list it looks like database are mostly linked to US partners.*
▪ *Ensembl https://www.ensembl.org (I am very familiar with this) highly used genome browser for agricultural species in Europe e.g. https://www.ensembl.org/Sus_scrofa/Info/Index Also Ensembl Plants https://plants.ensembl.org/index.html (I'm very familiar with this. European Farm Animal Biodiversity Information System (EFABIS) https://www.fao.org/dad-is/regional-national-nodes/efabis/en/ (somewhat familiar) FAANG Data Portal https://data.faang.org/ (Used a lot and can teach).*
▪ *Ensembl Plant.*
▪ *FlyBase, BeetleBase, BeeBase, AntWeb, InsectBase.*
▪ *https://www.ncbi.nlm.nih.gov/search/ https://phytozome-next.jgi.doe.gov http://gigadb.org https://bioinformatics.psb.ugent.be/plaza/*
▪ *I don't know much about these databases yet I know they are very important for methods development and validation.*
▪ *Phytozome is a good database. I used it a lot.*
▪ *BrassicaDB https://brassicadb.cn/*
▪ *Brassica DB*
▪ *phylogenes - I am somewhat familiar, phytozome - I am somewhat familiar.*
▪ *Ag Data Commons @ NAL?*
▪ *Ensembl, Phytozome*
▪ *As a plant biologist studying basic science questions, I'm most familiar with TAIR. But we are starting to work in legume species, so I'm interested to learn more about other databases.*
▪ *Is Phytozome considered a GGB database. I am familiar with this database.*
▪ *PLAZA and Proteomics DB.*
▪ *Phytozome*
▪ *Cyverse, Expression Atlas, SorghumBase -- I am somewhat familiar with all 3.*
▪ *I have been retired for 14 years, and though I have spent much of the intervening time advising, and even doing bench research, I was never comfortable with database management. I counted on the young pups in the lab.*
▪ *These over 1 dozen databases are stand alone and have very limited interoperability of information let aside asking an AI based question which can query all databases and present a possible answer.*
▪ *Lotus base - Lotus japonicus database, https://lotus.au.dk/ I am fairly familiar with it. CamRegBase https://camregbase.org/*

Table 6. Participant Experience with Specific GGB Database Functions

Please indicate your experience with specific genomics, genetics and breeding database functions:	I have not tried to do this	I've tried this and found it very difficult or impossible	I've done this but it was not easy to do it in a satisfactory way	I've done this and it was somewhat easy to do	I've done this and it was very easy
Submitting data to a GGB database	51 %	4	18	16	11
Searching for data in a GGB database	8	3	15	51	24
Retrieving data from a GGB database	10	3	16	54	16
Reusing data retrieved from a GGB database	26	4	19	35	17

Comments or clarifications on the above:

- *There often isn't all the required metadata (i.e. how the data was generated, the exact DNA sample/accession used, program arguments, etc.) that I need to determine it is correct to combine datasets.*
- *Searching and retrieving data is easier than reusing; reuse depends on metadata and understanding context of underlying study.*
- *This varies by database, so it is difficult to generalize.*
- *All these databases were built with old frame work in mind and need to evolve to keep the user interface in mind!*
- *It really depends on the database. Some are easier than others.*
- *The distinction between "retrieving" data from a database and "reusing" data is not clear. If the implication is that sometimes data that is retrieved is not useful, perhaps that could be framed more directly as a question such as "are the data retrieved from GGB databases always in usable formats?"*

Note. N = 80. Each row of numbers contains the proportion (percentage) of participants who gave each response to the question. Row percentages may not add to exactly 100% due to rounding.

Baseline Ratings and Recommendations On FAIR Data and GGB Databases

Survey participants were asked to rate their level of agreement with a series of statements about their knowledge and experience with FAIR data and GGB databases. They were also asked to rate their priorities for future development of these databases, and to provide related comments and written recommendations. The distributions of their responses are displayed in Tables 7 – 12; some highlights are noted below.

In some cases, ratings on multiple survey questions will eventually be combined into composite index scores, which will provide more reliable, psychometrically useful baseline measures to compare with follow up data that will be collected near the end of the current project. This step has not yet been done for this report; such indices will not be very useful until follow up data are available for formal statistical comparisons, but some notes about this are included below.

The first three questions displayed in Table 7 ask about baseline familiarity with FAIR data and value placed on these principles. Although the distributions are somewhat skewed, with a majority of respondents already reporting agreement with these question stems, there is room for improvement (greater agreement), so hopefully these measures will be sensitive enough to detect any differences at the follow up. (Though the follow up survey will include a different group of respondents, so any interpretation of differences from baseline to follow up should be carefully framed in that context.)

This pattern appears to hold for the remainder of the survey questions in Table 7 that asked respondents to rate the baseline implementation quality or value and usefulness of FAIR principles and related tools and procedures in GGB databases. There is little outright disagreement that these databases are currently doing a good job with these issues, but there were enough ratings of "slightly agree" or "moderately agree" that the survey should be sensitive to any improvement in participant appraisal of these tools and practices. In other words, it appears that the 2022 data has reasonable psychometric properties and will serve as a useful baseline measure for the project.

There are various ways to scan these ratings to identify which areas appear to be most ripe for improvement. We could add the two (or three) right-most columns to produce a quick sum that would show those items with the highest ratings (GGB database strengths), or conversely, add the two (or three) left-most columns to produce a quick sum that would show those items with the lowest ratings (weaknesses). For now, a quick color-coding scheme has been applied to give some indication of how this might be interpreted; darker blue indicates relative strengths, darker red indicates relative weaknesses that could be important targets for improvement. Some participant comments on the issues that were addressed in Table 7 are listed in Table 8.

Table 7. Baseline Appraisal of FAIR Data Awareness, Implementation and Value in GGB Databases

	Based on your knowledge of and experience with FAIR data and GGB databases, please rate how much you agree with these statements:	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
C1	I am very familiar with the concept of FAIR data (Findable, Accessible, Interoperable and Reusable) and could explain this to others.	8 %	10	9	23	19	33
C2	I am very familiar with FAIR data management (what metadata to collect, how to use standardized names and methods, etc.) and could supervise implementation of these practices.	16	11	13	26	19	15
C3	If I have a choice among data resources I am likely to work with those that best implement the FAIR data principles.	3	10	6	19	19	44
C4	GGB databases related to my work highlight the importance of FAIR data principles for researchers and provide educational resources to help researchers understand and follow FAIR practices.	3	8	13	27	27	24
C5	GGB databases related to my work make it easy to submit data following FAIR principles and practices.	4	10	18	29	23	16
C6	GGB databases related to my work provide data in common formats that are easy to work with.	-	11	10	27	29	23
C7	GGB databases related to my work curate and catalog data using standard terms, making it easy to search for data.	1	5	10	30	35	18
C8	GGB databases related to my work provide good guidelines on what standard nomenclatures to use when preparing/submitting data.	-	7	23	32	28	11
C9	GGB databases related to my work provide good guidelines on what metadata to provide when preparing/submitting data.	1	6	19	36	22	15
C10	GGB databases related to my work provide good guidelines on what file formats to use when preparing/submitting data.	-	6	8	39	28	18
C11	GGB databases related to my work provide good guidelines on how to provide information about how the data have been generated and cleaned, how missing data has been handled, and other process transparency issues.	1	6	23	37	21	11
C12	The process of contributing data to GGB databases helps researchers and breeders improve their understanding and ability to share their data according to FAIR principles	-	6	7	24	35	28
C13	The process of retrieving data from GGB databases helps researchers and breeders improve their understanding and ability to work with data according to FAIR principles.	1	4	3	31	26	35
C14	GGB databases enable their users to do things they couldn't have done otherwise, or enable them to do some things more quickly, easily, or inexpensively than would otherwise be possible.	-	3	5	23	22	47
C15	GGB databases provide cost-sharing efficiencies for individual users, projects, or institutions, enabling them to shift resources to other priorities.	1	8	7	22	30	32
C16	GGB database(s) related to my work provide useful resources for learning how to use them: tutorials, FAQs, how-to documents and videos, etc.	-	7	15	32	28	18
C17	Agricultural graduate students would benefit from a formal introduction during their courses (lectures and labs) to using GGB databases.	-	1	3	9	27	59

Table 8. Comments on FAIR Data Awareness, Implementation and Value in GGB Databases

Please add comments, extensions, clarifications or recommendations related to any of the questions above:
▪ *The databases are well designed for their purpose, but if you are doing anything else they are either too general or too specialized for something unrelated. So some of the statements above would be "strongly agree" for the people who developed the database and their close peers. I haven't really gotten much value myself.* ▪ *Many questions above are about submitting data, which I don't do.* ▪ *Database often use Gene-identifiers, nice for machine/bioinformatics users but unfriendly to biologists as they often look for a function and with missing annotation that becomes a hassle.* ▪ *GRIN does not provide information on how to submit data back to them. Stakeholders must contact crop curator directly.* ▪ *Emphasis of data must be on trait/phenotype leading to gene and mutation!* ▪ *I haven't submitted data to any GGB databases so I cannot speak to issues about data submission.*

Baseline Priorities for Further Development of FAIR Data Practices in GGB Databases

Survey participants were asked to rate the importance of six potential priorities for improved data curation; their responses are summarized in Table 9. All six were rated as being "very important" or "highest priority" by more than 60 percent of respondents. The highest ratings were given to "timely and up-to-date availability of curated data," "visualization of integrated data," and "training materials for FAIR data (for data submission).

Additional comments on these priorities are listed verbatim in Table 10. Some responses have been color coded to highlight themes that reflect similar comments from multiple participants, including better quality and consistency of curation and more educational opportunities for students and other database users.

Table 9. Baseline Priorities for Further Development of FAIR Data Practices in GGB Databases

Please indicate what you believe to be the most pressing priorities for improved data curation:		Not a Priority for Funding or Development	A Minor Priority	A Somewhat Important Priority	A Very Important Priority	Highest Priority for Funding or Development
D1	Timely and up-to-date availability of curated data	-	3 %	12	53	32
D2	Visualization of integrated data (e.g. genetic maps to whole genomes to expression)	1	1	25	47	26
D3	Hyperlinks and interconnectedness among databases	-	4	27	40	29
D4	Training material for FAIR data (for data submission)	-	5	24	39	32
D5	Training and support for database users	-	5	27	35	32
D6	Formal educational materials (i.e. lectures, labs) about FAIR databases for undergraduate and graduate courses	1	9	24	39	27

Table 10. Comments on Priorities for Further Development of FAIR Data Practices in GGB Databases

Please add any comments about your answers above or your recommendations for priorities. What are the most important avenues for improving the GGB databases you are aware of?
▪ *Well curated submissions.*
▪ *The data submission pipeline needs some serious work to ensure that all important metadata is submitted. Currently, data scientists can just exclude important information with no consequences.*
▪ *Accurate/sufficient metadata checks - Highest Priority for Funding or Development.*
▪ *Interconnectedness across databases, timely availability as this best serves the research community.*
▪ *Functional annotations instead of domain names.*
▪ *We will not have robust use of such tools until the science fields as a whole make it a priority for funding and publication requirements. When society (policy makers) understand we need these tools for food security/human health/economic sustainability.*
▪ *I have very limited use of GGB databases, but I understand their importance within the larger context of incorporating data science into life sciences education.*
▪ *Permanent funding.*
▪ *The educational materials would be immensely useful so instructors don't have to all write their own guides from scratch. I think in general, the features of many of these databases are underutilized because PIs assume their graduate students either know how to use them or will figure it all out on their own. The result is a community of users not entirely comfortable with the platforms or secure in their understanding of what the data means. I think a lot of users miss out on some super useful info and features because they use the platforms in a "trial by error" sort of way rather than an intentional, well-thought out systematic way.*
▪ *Important avenues for improving GGB databases, and access to them? Youth, or rejuvenation. In the old days I ground up tissues. Now, the day starts with turning on the computer.*
▪ *One item not mentioned above is data download capabilities. The databases can't provide all possible analysis capabilities that individual users will need, so users need to be able to extract data for download and local analysis.*
▪ *About how to use these databases, I think it is important for students to learn how to use the genomic information to design molecular experiments. For example, you are working on a gene that plays certain roles in Arabidopsis root. You can use TAIR JBrowser to find which isoform is expressing in root, and where the 5'UTR is. You can design your experiment based on the correct information instead of using the "representative isoform" from the annotation that sometimes does not even express in the condition of interest. It would save researchers time and grant agencies' money if we use genomic resources to help studying individual genes.*
▪ *A few use cases that lead a student to realize importance of FAIR data.*

Recommendations for Topics and Formats of Training Opportunities for Users of GGB Databases

Survey participants were asked an open-ended question about "what sort of training opportunities/formats or content/topics for users of these databases would be most helpful?" Their verbatim answers are listed in Table 11, and many are color-coded based on the following themes:

- Synchronous (live) educational events online such as webinars or asynchronous online video presentations, demonstrations or tutorials (these are coded together because after being recorded, events like webinars can be cut into segments and made into brief asynchronous, "static" video recordings. It is often helpful to design and organize webinars or similar online presentations with this segmentation and re-use in mind.) (20 comments)
- Static online educational materials such as tutorials or manuals (11 comments)
- Online courses – again, these can be synchronous and/or asynchronous or static, and an initial synchronous event or event series can later be posted online as an asynchronous or static learning resource. (6 comments) (Note that curriculum materials developed for an online or face to face course can then be repurposed and provided to instructors for use in their own courses; curriculum and assistance for course instructors was specifically mentioned in some comments.)
- Face to face workshops, stand alone or in conjunction with conferences or meetings (5 comments)
- Synchronous "office hours" or asynchronous "discussion forums" in which people can ask questions and get timely advice and assistance (3 comments)
- Several comments highlighted the importance of regularly updating any static online tutorials, manuals, or similar materials so that they reflect the current status of the databases they reference, even as those databases are updated. (2 comments)

In some cases it may be difficult to classify specific comments into adjacent themes, for example, a comment about "workshops" could mean online or face-to-face workshops. It may be best to combine the first two themes since the boundary between an online event and a later offering of the recorded version of that event is fuzzy and perhaps not important. The counts presented above might have some error but the themes themselves may still be useful. In addition, there are some individual comments that were not echoed by others and thus did not become part of a theme, but that may be worth considering, so it may be a good idea for readers to review all of the verbatim comments to find worthwhile suggestions or revise their personal interpretation of the themes above.

The comments and the coded themes above were dominated by format issues, e.g. how and where to deliver the topics and content. However, the comments in Table 11 do contain various recommendations for specific topics and content that would be useful. These are repeated here for ease of viewing:

- *How to generate and submit quality data.*
- *… specific workflows (i.e. association analysis) that cover the whole process from searching GGB for the data, how to know what datasets are suitable and can be combined, how to do the analysis, how to prepare the resulting data for submission, and how to actually submit it.*
- *Statistic and R tutorial would be a useful topic.*
- *… how to use tools in each database.*
- *Instructor training workshops so that educators can envision how to incorporate these materials into their courses.*
- *… training to keep abreast of changes to the databases.*
- *step-by-step user guides with explanatory info/definitions that are kept up to date.*
- *Actual research problem tutorials.*
- *Training on SQL, data visualization, data mining, statistics.*

Table 11. Recommendations for Training of GGB Database Users

What sort of training **opportunities/formats** or **content/topics** for users of these databases would be most helpful? (e.g. more static online tutorials or manuals, videos, live webinars, online courses over a period of weeks, face to face workshops, training on specific topics, etc.)
▪ *Static online tutorials or manuals.*
▪ *How to generate and submit quality data.*
▪ *A mixture of static online tutorials and live webinars, the latter supported by "office hours" or "drop in" times for continued support.*
▪ *Online courses over the summer.*
▪ *Online courses on specific workflows (i.e. association analysis) that cover the whole process from searching GGB for the data, how to know what datasets are suitable and can be combined, how to do the analysis, how to prepare the resulting data for submission, and how to actually submit it.*
▪ *Webinars and tutorials.*
▪ *Online tutorials on the GGB websites, more user friendly websites (so your use wouldn't depend on a kind person uploading a video tutorial), "common use" examples.*
▪ *Online tutorials are fine and can include videos.*
▪ *Online YouTube videos.*
▪ *Face to face workshops. Static online tutorials.*
▪ *Online training courses and recorded videos would be helpful. Statistic and R tutorial would be a useful topic.*
▪ *Face to face workshops by data curators at crop-specific or society meetings (Maize Genetics, Crop Science). Reduced registration cost for graduate students.*
▪ *Video and live virtual tutorials on how to use tools in each database.*
▪ *Static online tutorials, some videos.*
▪ *Courses in curriculum, free webinars tied to specific courses, professional workshops at annual meetings in all societies, funding security (it has to be sustainable).*
▪ *Instructor training workshops so that educators can envision how to incorporate these materials into their courses.*
▪ *SHORT videos that are searchable by topic and kept up-to-date.*
▪ *Easy-to-access online training to keep abreast of changes to the databases.*
▪ *Tutorials, summer camps, workshops.*
▪ *Online tutorials.*
▪ *Short videos paired with step-by-step user guides with explanatory info/definitions that are kept up to date. As an instructor, it is frustrating that every time updates are conducted, my class documents are out of date. Any documents created need to be reviewed regularly to keep up with updates to the sites.*
▪ *Static online tutorials or manuals.*
▪ *Videos, Q&A lists.*
▪ *Actual research problem tutorials.*
▪ *Online tutorials and videos!*
▪ *I have taught use of those databases in class, but it would be nice to have materials and free video lectures available for students who cannot participate. I also recommend making them in more than one language (English, Spanish, Chinese)?*
▪ *Online tutorials available at any time are more helpful than classes available only at limited times. Keeping tutorials up to date with database changes is critical, though.*
▪ *Tutorial videos and discussion forums (e.g. something like stackoverflow or biostars.org for bioinformaticians. You can ask questions there and people can help.)*
▪ *Training on SQL, data visualization, data mining, statistics.*
▪ *More static online tutorials or manuals, face to face workshops, and online courses that are always available.*
▪ *Graphic based workflow is more useful than text based materials.*
▪ *There should practical training courses for graduate students on specific topics that will enhance efficiency in the use of databases for research as it applies to plant, animals and microorganisms.*
▪ *All of these are useful.*
▪ *More live online training sessions.*
▪ *Online tutorials or manuals, videos.*
▪ *Video explanations.*
▪ *Online course for graduate level and training modules for basic understanding at the advanced undergraduate level.*
▪ *Online tutorials, short user cases.*
▪ *Hands-on training and office hours.*

At the end of the survey, participants were asked to offer any last comments or recommendations. These are listed verbatim in Table 12.

Table 12. Additional Comments

Please add any other comments, observations or recommendations you'd like to share about the development of genomic, genetic, and breeding databases for shared research community use:
▪ *For those scientists who publish but fail to finalize submissions, we must allow 3rd party annotations.*
▪ *Please fund them! These databases are extremely important to being able to do my work and I need to be able to rely on their continued existence and improvement!*
▪ *We will not have robust use of such tools until the science fields as a whole make it a priority for funding and publication requirements. And when society (policy makers) understand we need these tools for food security/human health/economic sustainability.*
▪ *Generally, the databases are outstanding.*
▪ *Inter linking of these data bases must be a priority to apply this knowledge to agricultural improvement and all data must be freely available!*
▪ *This will accelerate crop improvement to feed the future populations and also to achieve SDGs.*
▪ *They should avoid duplication of work and join hands to share resources, workload and curated data.*