



# DivSeek Connect

Official Newsletter of DivSeek International Network

Issue 5, January 2023

## A Message to our Members

DivSeek's ability to extend its network over the past few years has been dependent on financial support from Genome Prairie (Genome Canada) and the Global Institute for Food Research.

Facing the challenges of the pandemic we successfully launched our Strategic Plan, have established active regional and thematic hubs, and begun developing *DivSeek Commons* as a focus for identifying gaps in data interoperability, and for establishment of common data standards.

With our establishment funds due to dry up mid-year, we will be calling on our membership and stakeholders to work with us to identify ongoing funding sources, and so ensure the future reach and impact of DivSeek International Network.



**Graham King**  
DivSeek International Network  
Executive Director





## This issue:

DivSeek in San Diego . . . . .	3
Where global policy meets interoperability: the thorny issue of digital sequence information . . . . .	4
Germinate Hub expands utility for visualizing breeding data through time and space . . . . .	6
Empowering with genomics: DivSeek Hubs deliver workshop series for Latin American germplasm bank community . . . . .	8
What to expect from DivSeek's upcoming Barcodes to Bioeconomy workshop . . . . .	9
Why we need early career researchers to join the conversation on plant genetic resources . . . . .	10
Upcoming Events . . . . .	11

**Welcome to the fifth issue of DivSeek Connect, where we present news, information, and perspectives from the world of plant genetic resources.**

As we commence 2023, we encourage the global plant genomics community to reflect on decisions made during COP15 of the Convention on Biodiversity last month – particularly those concerning digital sequence information (DSI) in the context of Access & Benefit Sharing.

In the decision ([CBD/COP/15/L.30](#)), the COP agrees that benefits derived from DSI should be shared fairly and equitably, and acknowledges the [FAIR](#) and [CARE](#) principles of data governance. Importantly, they agree to develop an 'efficient' and 'feasible' solution for sharing the benefits of DSI (see summary report [here](#)).

Appropriately, this issue of *DivSeek Connect* highlights a new Open Access book chapter that proposes solutions to support future benefit-sharing mechanisms for DSI – by addressing gaps in germplasm traceability and database interoperability (page 4).

We also present on key achievements from three of our Hubs, including a workshop series for Latin American germplasm bank managers, and community-inspired software developments of the Germinate platform (pages 6 and 8).

# DivSeek in San Diego January 2023



## AGM and Community Update

<b>Thu 12 Jan 2023</b>	<b>Courtyard by Marriott, Hotel Circle</b>
<b>Noon – 4pm</b>	<b>Draft Program:</b> <ul style="list-style-type: none"><li>• Arrive, buffet lunch provided</li><li>• Board report (Stephen Kresovich), Finances (Chris Knihnitski), Operational update (Graham King)</li><li>• Member/community updates: 3-minute Flash Talks</li><li>• Break</li><li>• MOU with International Treaty (Daniele Manzella, ITPGRFA)</li><li>• Members' open-floor discussion</li></ul>

## PAG 30 Workshops

<b>Fri 13 Jan 2023</b>	<b>Palm 3 – 4, Town &amp; Country</b> Workshop: <a href="#">Information flows to harness plant genetic diversity</a>
<b>4:05PM</b>	Divseek Commons : Open Data, Analysis Tools and Best Practices for Plant Genetic Resources (Graham King)
<b>4:25PM</b>	Genesys PGR and Management of Genebank Accession Information. (Matija Obreza, Crop Trust)
<b>4:45 PM</b>	West Africa Centre for Crop Improvement, University of Ghana – Training Next Generation Plant Breeders in Africa for Food and Nutrition Security in Africa (Eric Danquah, WACCI)
<b>5:05PM</b>	Germinate: A Common Platform for Management of PGR Experimental Data (Paul Shaw, JHI)
<b>5:25PM</b>	Divseek Hubs Representing the Community of Practice for Latin America (Monica Carvajal-Yepes, Bioversity/CIAT)
<b>5:45PM</b>	Interoperable Platforms for Adding Value to Plant Genetic Resources (Sarah Dyer, EBI)
<b>6 – 7:30 PM</b>	<b>Palm 3 – 4, Town &amp; Country</b> Informal community gathering and networking event
<b>Sat 14 Jan 2023</b> <b>10:30 – 12:30</b>	<b>Pacific D, Town &amp; Country</b> Workshop: <a href="#">Genomics of Genebanks</a>

# Where global policy meets interoperability: the thorny issue of digital sequence information

**Corresponding author:** Pankaj Jaiswal, Oregon State University, Corvallis, OR, USA.

In October last year, a team of experts from FAO, CGIAR and Oregon State University published an Open Access book chapter titled '[Digital Sequence Information and Plant Genetic Resources: Global Policy Meet Interoperability](#).'

The chapter appears in 'Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development', alongside many other insightful perspectives by leading experts in data curation and governance.

The chapter takes stock of the information management systems and global policy frameworks that currently regulate plant genetic resources (PGR). In particular, they examine the status of digital sequence information (DSI) – a contentious issue for policy makers and genomic science communities alike.

DSI is at the heart of modern bioscience. "With the deluge of reference and pan-genome sequence available, researchers have very quickly identified large volumes of SNPs and other genetic markers," explains Pankaj Jaiswal, one of the authors of the chapter.

"These markers are now available for studying genetic diversity, genotype-to-phenotype association studies, integrating sequences with improved traits in plant breeding materials, as well as selecting candidate genes for genetically engineering functional gains, like improving yield or boosting stress resilience."

The strategies that utilise plant genetic DSI (such as accelerated breeding and CRISPR-based approaches) are helping to address numerous global food security challenges.

On the one hand, [F.A.I.R](#) (Findable, Accessible, Interoperable, Reusable) access to nucleotide and protein sequence information (DSI) is guiding and accelerating these developments. But on the other hand, DSI has the potential to undermine some of the global policy frameworks designed to ensure that benefits from agricultural science are equitably shared.

The international regulations on access and benefit-sharing (ABS) from the use of plant genetic resources include the International Treaty on Plant Genetic Resources for Food and Agriculture ([The Treaty](#)) and

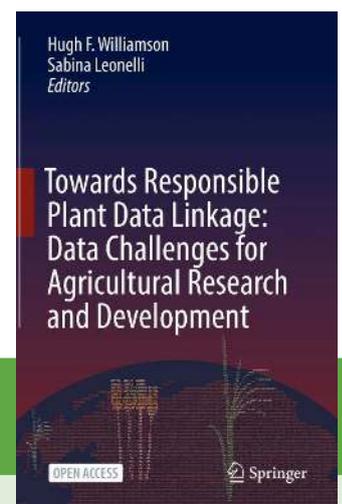
the [Nagoya Protocol](#) of the Convention on Biological Diversity (CBD).

At present, these ABS systems focus on regulating access to the physical, tangible components of genetic resources, i.e. germplasm. "Increasingly, global discussions are taking place on whether to regulate the intangible components (i.e. DSI) within the remit of these agreements," says Jaiswal.

In their chapter, the authors provide an insightful overview of current discussion on DSI and ABS policies up to the fifteenth Conference of the Parties to the CBD ([COP 15](#)), which was held shortly after the chapter was published. They then weave a compelling argument that in order to implement any future legal solutions, we must first enhance the interoperability of relevant data systems.

The authors go on to propose actionable solutions for associating DSI with its source germplasm, enabling reciprocal citations and data exchange between repositories of source material and repositories of genomic data.

Firstly, they propose integrating the federated system of databases of the International Nucleotide Sequence Database Collaboration (INSDC) through the Treaty's Global Information System (GLIS) via the well-known mechanism of digital object identifiers (DOIs).



Manzella, D., Marsella, M., Jaiswal, P., Arnaud, E., King, B. (2023). Digital Sequence Information and Plant Genetic Resources: Global Policy Meets Interoperability. In: Williamson, H.F., Leonelli, S. (eds) Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development. Springer, Cham. [https://doi.org/10.1007/978-3-031-13276-6\\_10](https://doi.org/10.1007/978-3-031-13276-6_10)

"DOIs are an ideal foundation because they have flexible metadata structures," explains Jaiswal. "[This] allows for descriptions specific to each type of knowledge object – from publications, to gene bank accessions, to datasets, to genetic markers and so on."

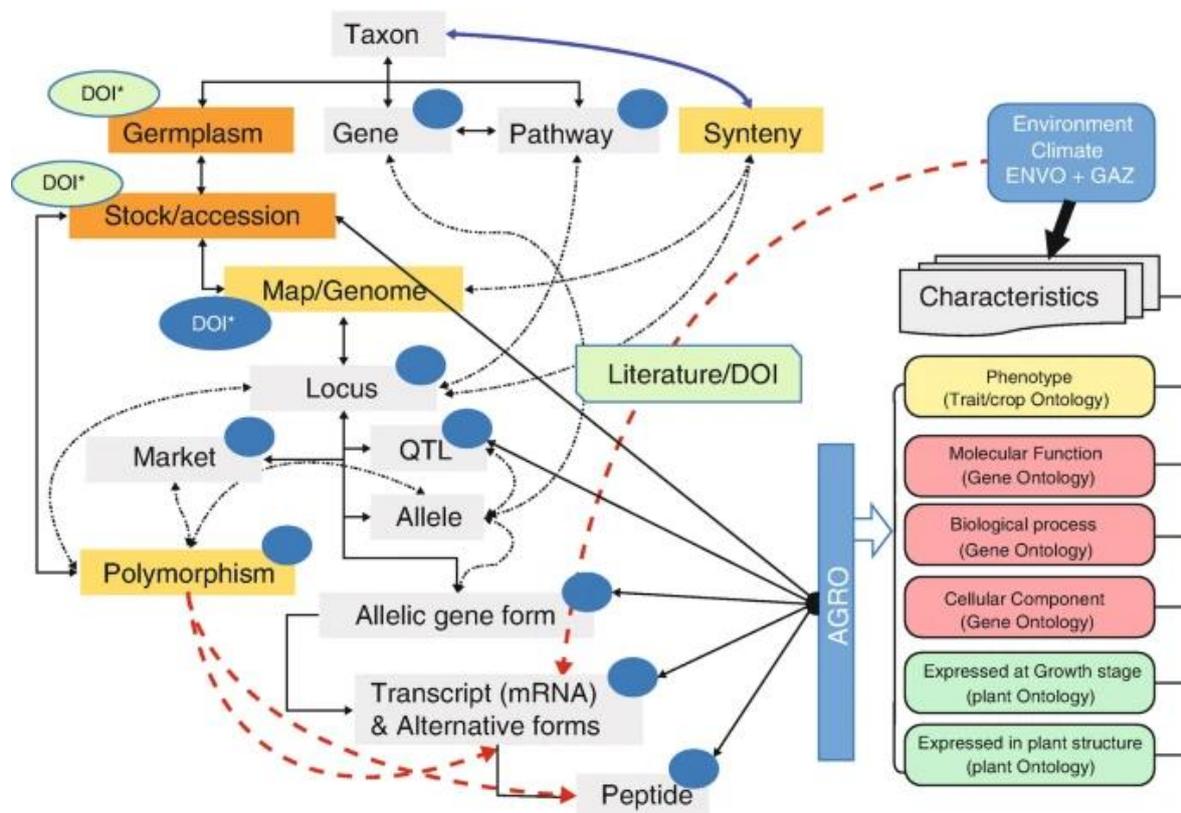
"GLIS already gives researchers the option to assign DOIs to their plant genetic resources" says Jaiswal. The authors suggest that genomics repositories like INSDC could adopt a standard of citing DOIs attributed to specific source material – whether assigned by GLIS, or by another authority. Of course, this will require discussing global metadata standards for DOI-based DSI.

Secondly, the authors propose the use of **digital genetic objects (DGOs)** to address some of challenges arising from the messy ambiguity of what constitutes DSI. Since there are so many different types of information (e.g. nucleic acid sequence reads, data about gene expression and function, unsequenced markers and chromosomal segments), they suggest

assigning a DGO to every distinct piece of DSI. Each of these DGOs can be linked to the DOI of the source material, and thus facilitate discovery via GLIS and interoperability among genomic data repositories.

A major roadblock in implementing these solutions would be user motivation to adhere to new standards – which is why pilot projects, community discussion, training sessions, and raising awareness of the benefits of interoperability is a must. "The concept has far-reaching impacts, not just for the plant community, but across life sciences, and would enable better citation indexing, provenance and benefit sharing," says Jaiswal.

We encourage all of our members to read this thought-provoking book chapter to deepen their understanding of the opportunities for greater data linkage in the world of plant genetic resources, and the benefits this could bring to all stakeholders.



**Figure 1** Schematic showing the utility of digital genetic objects (DGOs), represented with dark blue circles. These knowledge objects can be assigned distinct types of digital sequence information (DSI), and associated with digital object identifiers (DOIs), allowing for greater interoperability between data systems. (Source: Manzella et al. 2023)

# Germinate Hub expands utility for visualizing breeding data through time and space

**Corresponding authors:** Paul Shaw and Sebastian Raubach, James Hutton Institute, Dundee, Scotland.

[Germinate](#) is an open-source software platform that allows users to build, share and explore plant genetic resource databases.

The power of Germinate lies in its flexibility – offering facilities to store both standard collection information and passport data, along with more advanced data types such as phenotypic, genotypic and field trial data.

Importantly, the platform provides visualization tools that help users understand their data and to share it in standard ways with the community. Germinate offers both in-built visualizations, and the ability to export data to other visualization tools like [Flapjack](#), [Helium](#) and [CurlyWhirly](#).

The team behind Germinate is based at the James Hutton Institute in Dundee, Scotland. In 2022, they joined forces with DivSeek International to form the [Germinate Hub](#), which aims to promote and extend Germinate as a common platform and focal point for the plant genetic resources community.

We spoke to the Germinate Hub co-lead, Paul Shaw, about the latest project developments and software features – and how these can meet the needs of

breeders and scientists alike.

“2022 has been a busy year for us,” he says. “We have a pile of new Germinate features to play around with, and we encourage anyone who has not had a look at Germinate to try it out.”

Shaw’s team has been conducting work under the umbrella of the newly established [International Barley Hub](#) (IBH), funded by the Scottish Government/RESAS (JHI-B1-2 BARGAIN).

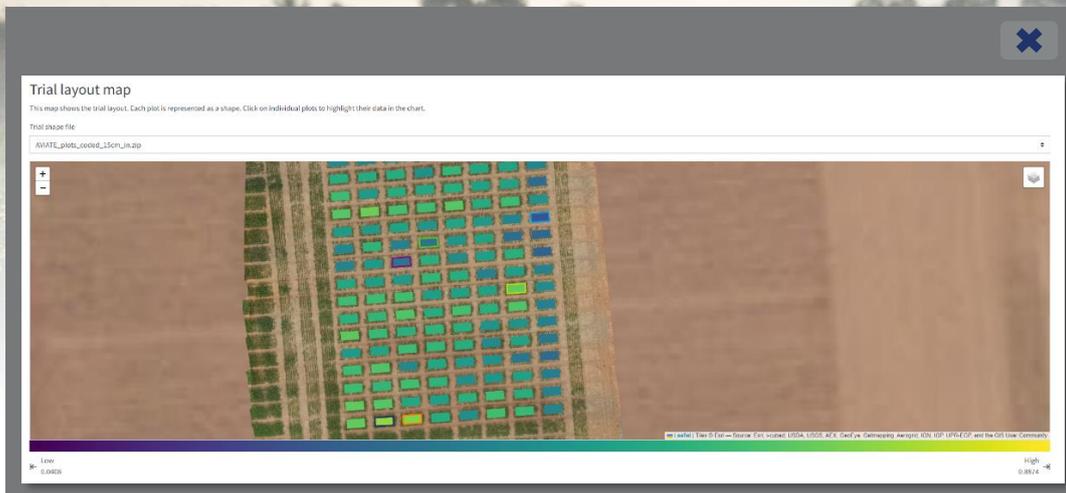
“As part of [this work], we have been developing tools to allow people to explore time-based trials data. In this case, it was drone imagery of barley fields, but the feature could be used for any data with a time aspect to it – and of course, any crop.”

Germinate’s new time series feature allows users to share, explore and download time series data of any kind. Excitingly, they have also added a map feature, which allows users to see how specific plots within a trial compare to other plots, and view this data using several interactive visualizations.

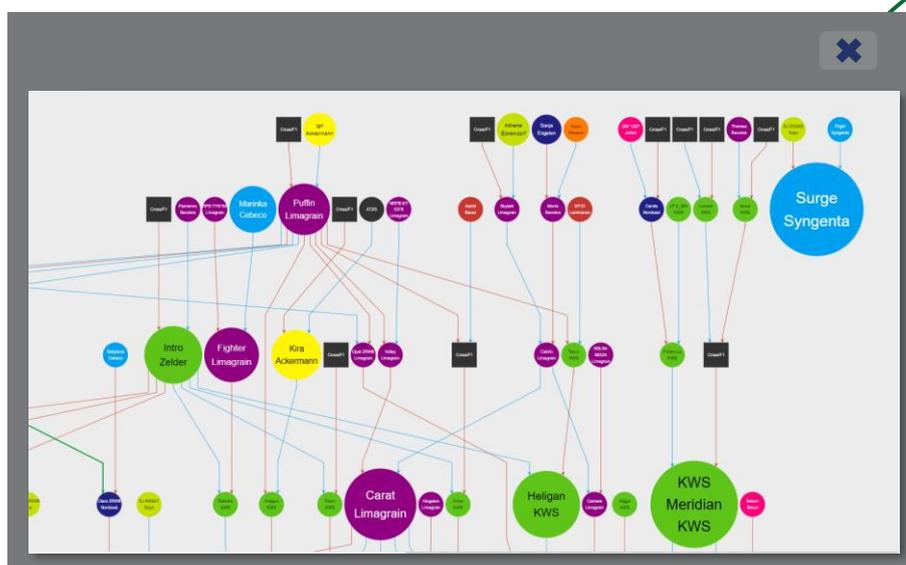
The Information & Computational Sciences Group at James Hutton are continually adding new features to Germinate, and also develop a variety of other software applications aimed at making the plant breeding data pipeline simpler, and more beautiful.



**Figure 2** Germinate’s new time series data view shows trials data across all time points. Individual germplasm can be selected to be highlight on the plot allowing direct comparison of individuals and their development over time.



**Figure 3** The new map visualization added to Germinate shows the trial layout in a geographic context. Trials data is visualized on top of the plots using a colour-coded heatmap. Additionally, drone imagery is displayed in the background.

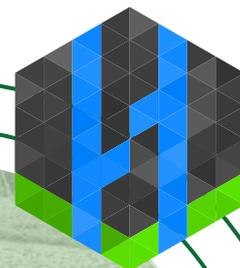


**Figure 4** The new web version of Helium showing an exemplar barley dataset. Circles represent germplasm and lines between them show parental relationships. Circles are coloured to show their original breeder while their size is based on the year they were introduced to the recommended list.

One of these features is Helium – a platform in which various data types can be shown in a pedigree context. It allows users to integrate different datasets and visualize large pedigree structures – helping breeders and researchers to make informed decisions when selecting suitable plant lines for use in subsequent crossings.

This year, the team has complimented the desktop-based version of Helium with a web-based version. “It allows users to quickly access many familiar Helium features without having to install anything on their computers,” says Shaw.

Shaw and his team are excited to offer the DivSeek community the first chance to try out the web-based version of Helium.



**Try Helium**

<https://helium.hutton.ac.uk>

Pass on your comments  
and feedback to

[helium@hutton.ac.uk](mailto:helium@hutton.ac.uk)



***The more feedback we have,  
the better chance we have of  
creating genuinely useful  
tools for the community.***



“We want feedback on our software from the community,” he says. “While we can’t guarantee we can meet everyone’s needs, the more feedback we have, the better chance we have of creating genuinely useful tools for the community.”

Helium, like Germinate and all other James Hutton Information Systems tools, is free to use. We encourage all Members and Observers of DivSeek to explore these tools and pass on their comments and feedback to [helium@hutton.ac.uk](mailto:helium@hutton.ac.uk).

# Empowering with genomics: DivSeek Hubs deliver workshop series for Latin American germplasm bank community

**Corresponding authors:** Carolina Sansaloni, César Petroli, and Sarah Hearne, CIMMYT.

Two of DivSeek International Network's Hubs recently joined forces to deliver a workshop series on the use of genomics and digital tools for the conservation and use of genetic resources for food and agriculture. The workshop series is a joint effort between our Regional Hub for Latin America and our 'Genomics to Democratize and Empower' Hub.

The third and most recent workshop was held from November 14 to 18 at the [International Maize and Wheat Improvement Center \(CIMMYT\)](#) headquarters in Mexico. It highlighted topics such as the analysis of genotypic data derived from the DArTseq platform and the analysis of gaps in germplasm collections.

It was attended by 33 participants from 14 countries and was regarded by many as a useful capacity-building as well as networking exercise. During the workshop, attendees shared the following comments (reproduced here with permission from [CIMMYT](#)):

**“ This workshop has allowed me to get in contact with colleagues from Latin America and to open the possibility of working together to advance in the characterization and contribute to the conservation of the collections we manage. ”**

– Instituto Nacional de Tecnología Agropecuaria, Argentina

**“ This has been a very profitable week for me as curator of the germplasm bank in Brazil. I learnt new concepts and tools that will generate advances in my work. ”**

– Brazilian Agricultural Research Corporation (EMBRAPA)



Photos from the third workshop of the Community of Practice for Latin America and the Caribbean inside the CIMMYT headquarters, Mexico. (Source: CIMMYT)



“It was very gratifying to be able to form this Latin American network of germplasm banks together with our colleagues from the Alliance of Bioversity International and CIAT,” said Carolina Sansaloni, co-lead of the DivSeek ‘Genomics to Democratize and Empower’ Hub.

“Training and knowledge exchange among germplasm banks is essential for better conservation, exploration and utilization of genetic resources in each country.”

The DivSeek Regional Hub for Latin America and the ‘Genomics to Democratize and Empower’ Hub have been working together since their inception.

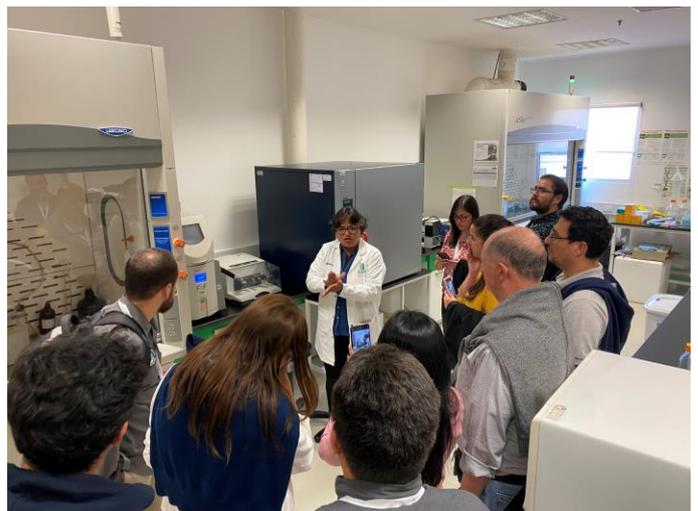
Both are based in Latin America, and share a common goal to foster a regional community of practice to share knowledge and build capacity for using genomic and digital tools to enhance the conservation of plant genetic resources.

The Regional Hub, headed by Monica Carajaval and Peter Wenzl, began its mission by conducting a [scoping survey](#) on the breadth of agricultural germplasm collections, as well as the knowledge gaps and challenges faced by those who curate them.

This survey helped to inform the foci of the workshops, which were organised and coordinated primarily by the co-leads of the ‘Genomics to Democratize and Empower’ Hub: Carolina Sansaloni, César Petrolí and Sarah Hearne.

The workshop was made possible through the support of DivSeek International, CGIAR’s Genebank Initiative, and Analitek-Illumina.

The series is an excellent example of what is possible when groups of researchers form new collaborations based on shared goals, and leverage their combined skills, networks and partnerships to identify and address gaps in the scientific community.



# What to expect from DivSeek's upcoming Barcodes to Bioeconomy Workshop

**Corresponding author:** Chris Richards

'Barcodes to Bioeconomy' will be a small, action-oriented workshop, bringing together scientists from the broad domains of plant gene bank curation, germplasm characterization, knowledge representation, data management and bioinformatics.

Our goal is to identify gaps that currently make it difficult to integrate data describing the physical specimens held in gene banks with derived digital characterization data, and to compare end-use characteristics between crops.

Given the rapid pace of genomic characterization for plant collections around the world, it is timely to convene a workshop on the coordination of such data. The primary outcome will be publication of a draft roadmap (or White Paper) for harnessing plant genetic resources (PGR) – defining the problem space and proposing a unified vision for global co-operation.

Unfortunately, the meeting concept was postponed 3 times during the pandemic. We hope now to convene a **pre-meeting of interested parties during PAG 30** alongside the networking event following [our workshop](#) (Fri Jan 13, 6pm, Pacific Salon 3-4, with refreshments!) to map out and prioritize the agenda for the 'Barcodes to Bioeconomy' workshop in Q2 of 2023.

Ultimately, we expect that development of interoperable information flows with feedback loops to gene banks will both improve curation and conservation, as well as downstream breeding efficiency.

Information flow about genetic resource accessions is often unidirectional, with long time-delays. The challenge of traceability between gene bank accessions and publication of characterization data (such as GWAS) is often exacerbated by subsequent re-generation in research labs.

This represents a missed opportunity because gene bank managers would benefit from accessing information about their materials to improve curation efficiency, and researchers and breeders would benefit from having more information about gene bank

accessions to enable them to identify useful diversity more accurately for research and variety development.

Workshop attendees will identify challenges and look for synergistic solutions for information flows among those involved in gene bank curation activities, genotypic and phenotypic evaluation of PGR, gene discovery programs, and the development of data analysis tools and data management systems. Some of these challenges include:

## ***How to enhance gene bank management?***

- Germplasm sampling strategies for genomic characterization and pan genome development.
- Subsets/core and development of structured populations.

## ***How to add value to gene bank annotation?***

- Best practices for metadata management for efficient applications by gene banks and breeders

## ***How can end-users unskilled in 'omics or informatics access information about PGR?***

- Develop/harness existing workflow platforms with interoperable containerized software/data repositories
- Building API solutions to connect databases, tools and information systems across communities.

## ***How to deliver information that promotes mobilization of variation from gene banks into the field?***

We believe that by developing standard sampling and data management practices, gene banks will become more efficient at conservation, better able to deliver this biodiversity to national and international stakeholders, and more able to engage as partners in research.



# Why we need early-career researchers to join the conversation on plant genetic resources



## Are you a postgrad student or early-career researcher?

We are currently looking for people like you to submit opinion pieces for publication in our regular eNewsletter, *DivSeek Connect*.

The community represented by DivSeek International Network is as diverse as the plant biodiversity we seek to protect. Our network encompasses academics, gene bank managers, software developers, policy experts and many more.

The beauty of this diversity is that we get to share perspectives and expertise from across the many disciplines involved in the utilisation and conservation of plant genetic resources (PGR). For example, our [upcoming PAG 30 workshop](#) features presentations from experts in crop genomics (Sarah Dyer, EMBL-EBI), gene bank information systems (Matija Obreza, Crop Trust), and bioinformatics software development (Paul Shaw, James Hutton Institute).

Encouraging knowledge-sharing across disciplines is a great way to build a more holistic picture of the PGR pipeline. It enables conversations that not only help to identify gaps and possible solutions, but also opportunities to collaborate and to develop those solutions.

But what we really need now is to elevate the voices of our early-career researchers.

Junior academics are an invaluable asset in any research community. They are the engine house of R&D, and can offer a hands-on perspective and fresh insights.

What's more, they represent the generation(s) that will have to live with the consequences of global challenges: climate change, food insecurity, population growth, and depleted natural resources.

Encouraging early-career researchers to join the conversation about evolving the PGR pipeline not only benefits the scientific community, but of course, the researchers themselves.

If you are a postgrad student or post-doc, publishing an opinion piece in a scientific community newsletter, like *DivSeek Connect*, is an excellent way to get your name out there. It's also a great way to network and find potential future collaborators and even employers.

Some 'seeds' of ideas on potential topics (but no need to limit yourself to these!):

- Practical constraints on traceability relating to PGR accessions (seed, data)
- Day to day experiences of interoperability between data systems
- How to make gene bank research more appreciated in the wider community
- How to ensure your project data will be FAIR for future generations
- Different global perspectives on Access & Benefit Sharing (ABS)

Whatever your field of study, we encourage all early-career researchers dealing with PGR to think creatively and critically, and to submit their opinion pieces (500 to 1000 words) to *DivSeek Connect*. Get in touch with us at [info@divseekintl.org](mailto:info@divseekintl.org).



# Upcoming Events



## International Galaxy Community Conference (GCC2023)

24 – 27 April, 2023  
Padova, Italy

This annual gathering of the [Galaxy](#) community includes training, talks, posters, demos, Birds of a Feather meetups, and many other opportunities for collaboration and networking. More information at [biocommons.org.au](https://biocommons.org.au).



**Sorghum in the 21st Century  
Global Sorghum Conference**  
Resiliency and Sustainability in the Face of Climate Change

## Global Sorghum Conference

5 – 9 June, 2023  
Montpellier, France

This conference will connect the broader global sorghum community in identifying key solutions to urgent challenges facing agriculture and food security today. More information at [21centurysorghum.org](https://21centurysorghum.org).



## ISF World Seed Congress

5 – 7 June, 2023  
Cape Town, South Africa

Organized by the International Seed Federation (ISF), this business-centered event incorporates both in-person and virtual opportunities. More information at [congress.worldseed.org](https://congress.worldseed.org).



## AG2PI Conference

15 – 16 June, 2023  
Kansas City, USA

The AG2PI Conference will include talks by community thought leaders, stakeholder roundtables, small group discussions, and networking opportunities, all designed to map the future of AG2P research. Registrations expected to open in February 2023. More information at [ag2pi.org](https://ag2pi.org).



## 16th International Biocuration Conference

July 2023  
Brisbane, Australia

A unique event for curators and developers of biological databases to discuss their work, promote collaborations, and foster a sense of community. More information at [biocuration.org](https://biocuration.org).



**DIVSEEK  
INTERNATIONAL**  
NETWORK INC.

## **Call out for Newsletter Contributions**

We are keen to receive any content, short or long, from Members and Observers, organisations and individuals. We welcome your news updates, opinion pieces, research abstracts and upcoming conferences.

Contact [info@divseekintl.org](mailto:info@divseekintl.org) with 'Connect' in the email subject.

DivSeek International Network Inc.

421 Downey Rd | Saskatoon, SK S7N 4L8 CANADA | +1 306-966-3707

Copyright © 2022 DivSeek International Network Inc. All rights reserved.