Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Wednesday, October 22, 2014 11:43 AM
To: Chris Richards
Subject: Re: skype call on genepool coverage of ex situ collections

Ok, great! Ex situ genetics meeting sounds interesting! Who is organizing it?

H

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Wed, Oct 22, 2014 at 5:33 PM, Chris Richards <crichard@lamar.colostate.edu> wrote:

> Hey Hannes-
>
> OK, Wednesday I'll be in San Diego (San Diego Zoo ex situ genetics meeting). I can Skype therefore at 7:00 PDT.
>
> Cheers,
>
> C.
>
> **From:** Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> **Sent:** Wednesday, October 22, 2014 8:23 AM
> **To:** crichard
> **Subject:** Re: skype call on genepool coverage of ex situ collections
>
> We can't do Monday since it's our donor council meeting. The suggestion now is Wednesday...?
>
> H

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Wed, Oct 22, 2014 at 4:21 PM, crichard <crichard@lamar.colostate.edu> wrote:

On Wed, 22 Oct 2014 12:29:24 +0200, Hannes Dempewolf <hannes.dempewolf@croptrust.org> wrote:

Good on Monday....traveling on Tuesday

Dear all,

It seems like not all of us can make the proposed date and time. I therefore suggest we reschedule to the same time (4pm Bonn time (3pm Ibadan, 9am Mexico City, 8am Fort Collins)) either on Tuesday, the 4th of November or Wednesday, the 5th of November.

Please let me know if you would be available then.
Thanks,

Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org [1]

Conserving Crop Diversity, Forever

On Tue, Oct 21, 2014 at 10:16 AM, Hannes Dempewolf wrote:

Dear Michael, Tom and Chris,

Charlotte mentioned that at the AGM you had some interesting discussions around the question of what could be a reasonable target for genepool coverage of ex situ collections? We had also been discussing related issues with Chris Richards (USDA) and were thinking it would make sense to have a bit of a discussion via skype to explore some of the ways that this question could be tackled.

It seems like Friday the 31st of October would be a good date for a call, since we are all in office then. Would this also work for you? Say 4pm Bonn time (3pm Ibadan, 9am Mexico City, 8am Fort Collins)?

Please let me know.
Thanks,
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115 [3]
Mobile [redacted]7 [4]
www.croptrust.org [5]

Conserving Crop Diversity, Forever

Links:
---
[1] http://www.croptrust.org
[2] mailto:hannes.dempewolf@croptrust.org
[3] http://acnsmail.colostate.edu/tel:%2B49%20228%2085427%20115
[4] http://acnsmail.colostate.edu/tel:%2B49%20171%201839227
[5] http://www.croptrust.org

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Wednesday, October 22, 2014 10:23 AM
To: crichard
Subject: Re: skype call on genepool coverage of ex situ collections

We can't do Monday since it's our donor council meeting. The suggestion now is Wednesday...?

H

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Wed, Oct 22, 2014 at 4:21 PM, crichard <crichard@lamar.colostate.edu> wrote:

> On Wed, 22 Oct 2014 12:29:24 +0200, Hannes Dempewolf <hannes.dempewolf@croptrust.org> wrote:
>
> Good on Monday....traveling on Tuesday
>
> Dear all,
>
> It seems like not all of us can make the proposed date and time. I therefore suggest we reschedule to the same time (4pm Bonn time (3pm Ibadan, 9am Mexico City, 8am Fort Collins)) either on Tuesday, the 4th of November or Wednesday, the 5th of November.
>
> Please let me know if you would be available then.
> Thanks,
>
> Hannes
>
> ---
> Hannes Dempewolf
> Scientist and Project Manager
> Global Crop Diversity Trust

duplicate email trail deleted

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Wednesday, October 22, 2014 6:50 AM
To: Charlotte Lusty
Cc: Michael Abberton (IITA); Thomas Payne; Luigi Guarino; Chris Richards
Subject: Re: skype call on genepool coverage of ex situ collections

Thanks for the quick reply, Charlotte.

OK, so let's propose Wednesday 5th of November 4pm Bonn time (3pm Ibadan, 9am Mexico City, 8am Fort Collins)).

Please let me all know whether you can make it then.
Thanks!
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Wed, Oct 22, 2014 at 12:33 PM, Charlotte Lusty <charlotte.lusty@croptrust.org> wrote:

Dear Hannes

I would only be able to make Wednesday.

Charlotte

On 22/10/14 12:29, Hannes Dempewolf wrote:

> Dear all,
>
> It seems like not all of us can make the proposed date and time. I therefore suggest we reschedule to the same time (4pm Bonn time (3pm Ibadan, 9am Mexico City, 8am Fort Collins)) either on Tuesday, the 4th of November or Wednesday, the 5th of November.
>
> Please let me know if you would be available then.
> Thanks,
> Hannes

duplicate email trail deleted

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Tuesday, October 21, 2014 4:03 AM
To: Chris Richards
Subject: Re: skype

OK, great. Perhaps a paragraph of two that summarizes your thinking about the challenge? That would be very useful.

Thanks!
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Mon, Oct 20, 2014 at 6:12 PM, Chris Richards <crichard@lamar.colostate.edu> wrote:

> That would be fine, Hannes. Anything I should prepare?
>
> Cheers,
>
> C.
>
> **From:** Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> **Sent:** Monday, October 20, 2014 7:47 AM
> **To:** Chris Richards
> **Subject:** skype
>
> Hi Chris,
>
> We were wondering whether you may be available for a chat (via skype)with Charlotte and me (and possibly also a couple of CGIAR genebank managers) about the question of genepool coverage of collections on Friday, the 31st of October?

Please let me know. If so, can you please let me know what time would work for you?

Thanks,
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Friday, October 17, 2014 5:51 AM
To: Chris Richards
Subject: Re: PAG talk about DivSeek

Hi Chris,

Ok, I'll ask Peter to put something together and send it to you asap.

I am hearing through the grapevine that the USDA will want to change some of the language of this non-binding (!)yet expression of interest.... Not quite sure how we'll handle that one... We are just now waiting to receive their official notice.

Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Fri, Oct 17, 2014 at 12:08 AM, Chris Richards <crichard@lamar.colostate.edu> wrote:

> Hannes-
>
> Please have Peter Wenzel send me a title and an abstract for his ~30 minute slot at PAG Genomics of Genebanks workshop. I can enter the data in the PAG website. I would like it all ready to go ASAP.
>
> I hope Peter Bretting figures out who can bloody sign this letter of interest with DivSeek! So embarrassing that USDA (the 800 lb gorilla in the corner) can't get out in front of this.
>
> Hope you are doing great.

Cheers,

Chris

Hutchison, Stasia

From: Bretting, Peter
Sent: Friday, October 10, 2014 12:29 PM
To: Richards, Chris
Cc: Buckler, Ed
Subject: RE: DivSeek partners coming on board

Need to determine who can sign for ARS. Working on it,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Richards, Chris
Sent: Friday, October 10, 2014 11:17 AM
To: Bretting, Peter
Cc: Buckler, Ed
Subject: FW: DivSeek partners coming on board

Any movement towards a signature for our agency?

Cheers,
C.

From: DivSeek [mailto:divseek@croptrust.org]
Sent: Friday, October 10, 2014 2:42 AM
Cc: Marie Haga; Bhatti, Shakeel (AGDT); Powell, Wayne (CGIAR Consortium); Ruth Bastow; Div Seek; pgrfa-treaty@fao.org
Subject: DivSeek partners coming on board

Dear recipients of the DivSeek expression of interest letter,

This is to let you know that 23 partner institutions have already expressed their interest in the DivSeek initiative (see list below). We are very happy to see so many and such a diverse group of institutions already engaged. We also know from some others that the process of signature is under way and we are therefore very much looking forward to receiving further signed letters ahead of the October 31st deadline.

Given the considerable interest in the DivSeek initiative at this point already, we have decided to go ahead with plans for the first partners assembly, which has now been scheduled for **Friday, the 9th of January** to take place in San Diego (USA). This date was chosen, as it is convenient for many who are also planning to attend

duplicate email trail deleted

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Friday, October 10, 2014 6:34 AM
To: Chris Richards
Subject: Re: PAG speaker

Hi Chris,

Yeah, it will for sure have to be a guestimate for now, but that doesn't mean that we wouldn't be interested to look at ways to better understand the issue more generally. Unfortunately, at this time we don't have funds that could be used for a major study on this. I'll let you know about the outcome of our discussions next week.

Cheers,
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Fri, Oct 10, 2014 at 12:22 AM, Chris Richards <crichard@lamar.colostate.edu> wrote:

> As to this metric, sounds like you need a quick project target, based on some reasonable set of data...but that the metric itself can continue to be refined and validated.
>
> Parameterizing a genetic simulations from SDMs is a way of developing arrange of plausible structures. Ultimately breeders are interested in adaptive variation...and calculating the scale of localized selection will be difficult without knowing the biology of the trait.
>
> I'm willing to engage with you guys but it doesn't look like you're willing to invest anything more than a short study...something that may be insufficient to really validate the methods and get them in a shape for their general application.

I will pursue this kind of metric in the course of our genetic geographic simulation objectives and will keep you posted.

Let me know what else I might offer to get you towards your short term goal....which will invariably be just guestimation!

Cheers,

C.

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Thursday, October 09, 2014 2:40 PM

To: Chris Richards
Subject: Re: PAG speaker

Hi Chris,

You are right, there is an immediate need for us to come up with a somewhat realistic performance target for the CG genebanks and that is the CGIAR's SRF (Strategy and Results Framework), which is currently under development and which we are contributing to: http://www.cgiar.org/resources/strategy-and-results-framework/

The figure that is floating around is 60% of crop diversity conserved by year XXXX. Charlotte is much more in the thick of this than I, so I am not sure how much it makes sense for us to have a chat tomorrow - though I'd of course be happy to do so, if yo think it'll be useful. Luigi, Charlotte and I are scheduled to discuss this next week, once they are back from Arusha, so it would be good to receive any further ideas you may have (and that you are willing to share) before then.

Cheers,
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7

53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Thu, Oct 9, 2014 at 10:11 PM, Chris Richards <crichard@lamar.colostate.edu> wrote:

Hi Hannes-

As to the performance indicator research I'd really like to have a chat about this. The basis for the research in this area is already underway in my lab with the coalescent simulations parameterized with niche models form CWR taxa in NA.. The work is part of my lab's project plan. I have an idea for developing a collaborative project with a colleague who works in this area of theory that might be useful in the short-term. I will be speaking with this person at the end of the month while they are here on a visit.

I sense you have a particular back-story here. What is it that you want in the short term and why the abbreviated work schedule? Sounds like you have a donor who wants benchmarks in place. I want to work with you, so give me the context here and I'll see what I can come up with.

Can we Skype sometime tomorrow?

Cheers,

Chris

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Thursday, October 09, 2014 2:02 AM
To: Chris Richards
Subject: Re: PAG speaker

Hi Chris,

Yes, Charlotte and I have read and briefly discussed your email - though only remotely, since she is at the AGM in Arusha and I am back here in the offices. As you had indicated, we are indeed under some time pressure here, since the goals for these indicators have to be set soon - so even though I tend to agree with you that this would warrant a multi-year PostDoc lead research study, I am afraid this won't be possible at this time. Do you have any alternative suggestions on how we could drive this forward in the space of a few weeks or months?

Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Wed, Oct 8, 2014 at 10:29 PM, Chris Richards <crichard@lamar.colostate.edu> wrote:

Hi Hannes-

Unrelated to the PAG workshop, did you read through my email about your performance metric? I'm interested in following up with this only since you seemed to want to get going on it rather quickly. Let me know what more information I can provide.

Cheers,

Chris

Hutchison, Stasia

From:	DivSeek <divseek@croptrust.org>
Sent:	Friday, October 10, 2014 4:42 AM
Cc:	Marie Haga; Bhatti, Shakeel (AGDT); Powell, Wayne (CGIAR Consortium); Ruth Bastow; Div Seek; pgrfa-treaty@fao.org
Subject:	DivSeek partners coming on board

Dear recipients of the DivSeek expression of interest letter,

This is to let you know that 23 partner institutions have already expressed their interest in the DivSeek initiative (see list below). We are very happy to see so many and such a diverse group of institutions already engaged. We also know from some others that the process of signature is under way and we are therefore very much looking forward to receiving further signed letters ahead of the October 31st deadline.

Given the considerable interest in the DivSeek initiative at this point already, we have decided to go ahead with plans for the first partners assembly, which has now been scheduled for **Friday, the 9th of January** to take place in San Diego (USA). This date was chosen, as it is convenient for many who are also planning to attend the annual International Plant & Animal Genome XXIII Conference, which takes place in San Diego from the 10th to the 14th of January. We will distribute more information about the partners assembly to all partners after October 31st.

Here the list of partners so far (in alphabetical order):

1 Biotechnology and Biological Sciences Research Council (BBSRC)
2 Bioversity International
3 Centre for Genetic Resources, the Netherlands (CGN)
4 Centro Agronómico Tropical de Investigación y Enseñanza (CATIE)
5 CGIAR Consortium Office
6 Dalhousie University
7 Genome Prairie
8 Global Crop Diversity Trust (Crop Trust)
9 Global Plant Council (GPC)
10 Indonesian Agency for Agricultural Research and Development (IAARD)
11 International Center for Tropical Agriculture (CIAT)
12 International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
13 International Institute of Tropical Agriculture (IITA)
14 International Potato Center (CIP)
15 International Rice Research Institute (IRRI)
16 Jülich Research Centre (Forschungszentrum Jülich)
17 Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
18 Queensland Alliance for Agriculture & Food Innovation (QAAFI)
19 Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA)
20 The University of British Columbia (UBC)
21 The University of Georgia, College of Agricultural and Environmental Sciences
22 University of Missouri, Division of Plant Sciences
23 World Agroforestry Centre (ICRAF)

Thank you again for your interest and support in the DivSeek initiative.

Best wishes,
Hannes Dempewolf *(on behalf of the joint facilitation unit)*

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Thursday, October 09, 2014 4:40 PM
To: Chris Richards
Subject: Re: PAG speaker

Hi Chris,

I realize it's silly to keep the name from you, so I trust you won't share it further for a few more days. It's Peter Wenzl, who as you know has during his time as lead of the Seeds of Discovery project gained much relevant expertise.

You are right, there is an immediate need for us to come up with a somewhat realistic performance target for the CG genebanks and that is the CGIAR's SRF (Strategy and Results Framework), which is currently under development and which we are contributing to: http://www.cgiar.org/resources/strategy-and-results-framework/

The figure that is floating around is 60% of crop diversity conserved by year XXXX. Charlotte is much more in the thick of this than I, so I am not sure how much it makes sense for us to have a chat tomorrow - though I'd of course be happy to do so, if yo think it'll be useful. Luigi, Charlotte and I are scheduled to discuss this next week, once they are back from Arusha, so it would be good to receive any further ideas you may have (and that you are willing to share) before then.

Cheers,
Hannes

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

On Thu, Oct 9, 2014 at 10:11 PM, Chris Richards <crichard@lamar.colostate.edu> wrote:

> Hi Hannes-
>
> OK, I will keep a slot open for He Who Should Not Be Named. Can I assume that your new hire will, in fact, be able to give a talk about DivSeek or has this even been discussed? I'm in the process of finalizing the speaker list.

duplicate email trail deleted

Hutchison, Stasia

From: Richards, Chris
Sent: Monday, October 06, 2014 7:49 PM
To: Hannes Dempewolf
Cc: Charlotte Lusty
Subject: RE: crop diversity indicator
Attachments: postdocproposal_final.docx

Dear Hannes and Charlotte-

I've been considering your interest in getting a handle on a metric that could be used to rank the efficacy of an ex situ collection (roughly Genetic Diversity in ex situ conservation/ total genetic diversity of the species). There is something of conundrum for estimating the unsampled. This has been propounded through gap analysis where we might know the geography of missing samples, but we don't know if these missing regions contain novelty to the collection. I have been thinking of an approach that might extract more spatial information out a niche model including the ecological distance (least cost path type metric) which can parameterize the migration matrix in a set of structured genetic coalescent simulations. We could even parameterize regional selection. The simulations could be run across a range of scenarios and the results would be an estimated genetic layer (like a climatic layer). One could use this simulated data to estimate (or just rank) an indicator of collection completeness. In addition a complementary metric would seek to estimate the global effective population size which has its own challenges.

The simulation data might be a first step at least in understanding the behavior of the metric. One could easily envision where geography or climate are not the best predictors of genetic diversity (selection and history are idiosyncratic and hardly come to equilibrium) but I think those cases would be the exception. Importantly this set of initial simulation studies could explore where these metrics are useful, and where they fail. That in itself would be valuable. Where the approach works, however, it still may not have enough predictive power to be used objectively as a performance indicator without additional information data.

Much of this kind of analysis involves some pretty serious programming. Finding the right person for this is not obvious. I thought I have someone in mind, but after more thought I realized that this project is more complex.

Attached was a post-doc proposal I wrote that is related to this issue in that predictive modelling is ties to some empirical validation. The proposal was not funded, but a proposal like this hardly is.

I'm not sure what the timeline is for enacting this kind of study, we are going forward with this without external funding.

More later,
Cheers,
C.

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Friday, September 26, 2014 1:44 AM
To: Richards, Chris
Cc: Charlotte Lusty
Subject: Re: crop diversity indicator

Hi Chris,

OK, sounds good. A skype chat today is difficult for me, since my afternoon is packed. Do you have any specific questions that I could help answer via email? Perhaps you could draft a brief paragraph with some ideas

that I could discuss here internally and give you feedback on? The genebank manager's AGM is the week after next, so it would be neat if we could have a bit of a plan ready by then. I am copying here also Charlotte, since she is very much involved in our thinking around this.

Cheers,
Hannes

On Fri, Sep 26, 2014 at 12:43 AM, Richards, Chris <Chris.Richards@ars.usda.gov> wrote:

Hi Hannes-

Gosh I left for [redacted] ans just got back today. You bet I'm interested!

I think I might have someone in mind I'll need to check.

Perhaps a conversation via Skype tomorrow? Sorry for the delay....I've had no email out on this [redacted].

Cheers,

C.

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Thursday, September 25, 2014 8:01 AM
To: Richards, Chris
Subject: Re: crop diversity indicator

Hi Chris,

Have you had a chance to consider this?

Thanks,
Hannes

On Wed, Sep 17, 2014 at 9:53 AM, Hannes Dempewolf <hannes.dempewolf@croptrust.org> wrote:

Hi Chris,

Do you have a suggestion on how we could move swiftly forward with this? We were thinking it may make sense to engage an intern / grad student / research assistant to do some of the necessary groundwork. Is there someone you have in mind who may be suitable/interested? We could probably help out with some financial support for a short desk study if necessary.

Hannes

Am Sep 17, 2014 um 6:28 AM schrieb "Richards, Chris" <Chris.Richards@ars.usda.gov>:

> Hey Hannes!
>
> Cool question!...incredibly difficult to generalize conceptually, and virtually no accepted standards in PGR that I know about. Certainly there are some surrogates or proximal metrics that could be implemented but in practice these are not often considered. Challenging but worthwhile endeavor. Of course you can count on me to assist in any way I can!
>
> Cheers,
>
> C
>
> **Christopher Richards, Ph.D.**
>
> Population Geneticist
>
> United States Department of Agriculture
>
> National Center for Genetic Resources Preservation
>
> Colorado State University
>
> 1111 South Mason Street
>
> Fort Collins, CO 80521
>
> USA
>
> Chris.richards@colostate.edu
>
> 970 495 3201
>
> http://www.ars.usda.gov/pandp/people/people.htm?personid=42033

http://orcid.org/0000-0002-9978-6079

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Tuesday, September 16, 2014 6:00 AM
To: Richards, Chris
Subject: crop diversity indicator

Hi Chris,

I have a question for you. As you know, the Trust is putting quite a bit of thinking into defining 'performance indicators' for genebanks. We have been toying with the idea of some higher-level indicators that could be relevant not just for the CG genebanks but the global system of PGR conservation more generally. One indicator that seems to excite a lot of people is "global diversity of crop x conserved". This is of course very tricky to quantify since for few (any?) crops we have a good enough understanding of the amount of intraspecific diversity that exists, nor are there any well defined, universally accepted ways to measure/quantify it crop diversity. Large-scale genotyping data sets may be able to get us closer to an answer, but I am not sure if anyone has recently tried to do this for a set of different crops? Have you/ someone in your group? If not, would you be interested to engage with us on this?

Hannes

--

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

--

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

--

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Richards, Chris
Sent: Sunday, October 05, 2014 11:08 PM
To: Edward S. Buckler; Bretting, Peter
Cc: Buckler, Ed
Subject: RE: DivSeek expression of interest

Thanks, too. I realize this can be logistically confusing...but it's good to see the ARS getting behind this effort.
-Chris

From: Edward S. Buckler [mailto:esb33@cornell.edu]
Sent: Sunday, October 05, 2014 5:01 PM
To: Bretting, Peter
Cc: Richards, Chris; Buckler, Ed
Subject: Re: DivSeek expression of interest

Thanks very much for following up on this.
-Ed

On Oct 5, 2014, at 3:35 PM, Bretting, Peter <Peter.Bretting@ARS.USDA.GOV> wrote:

Hi Chris and Ed—just an update. Conferring with State Dept. and USDA/ARS, we believe that ARS can sign the DivSeek expression of interest. But now we must identify whom in ARS should actually sign it!

So, keep tuned, I'll continue to work on it this week.

Thanks,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Bretting, Peter
Sent: Friday, September 26, 2014 4:16 PM
To: Richards, Chris; Edward S. Buckler; Buckler, Ed
Subject: FW: DivSeek expression of interest

Hi Chris and Ed—would you like me to reply to Hannes on behalf of ARS?

Thanks,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Tuesday, September 23, 2014 11:01 AM
To: Bretting, Peter; Richards, Chris; Edward Buckler
Cc: Marie Haga; Bhatti, Shakeel (AGDT); Ruth Bastow; Powell, Wayne (CGIAR Consortium)
Subject: DivSeek expression of interest

Dear Drs. Peter Bretting, Christoper Richards and Ed Buckler,

After two scoping meetings in January and April 2014, DivSeek is becoming a reality. We have created a small steering and working group to develop a white paper, to initiate the scoping phase and to develop the plan for launching the initiative. A website includes all our work to date (http://www.divseek.org).

A first Partners' Assembly will take place in early 2015, to establish the governance of DivSeek. For your institution to take part in the Assembly, we ask you to return a signed copy of the letter that is attached, by 31 October. Please identify the appropriate person who can sign on behalf of your organization. Signature of the letter constitutes a simple and non-binding expression of interest in DivSeek. Please submit the signed letter to divseek@croptrust.org and pgrfa-treaty@fao.org

A draft Charter, setting forth the principles and institutional mechanisms of DivSeek, will be distributed in advance of the Partners' Assembly.

We continue to promote DivSeek as an open and transparent initiative, and look forward to your further engagement.

Kind regards,
Marie Haga, Shakeel Bhatti, Ruth Bastow and Wayne Powell

On behalf of the joint facilitation unit:
--
Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Hutchison, Stasia

From: Bretting, Peter
Sent: Sunday, October 05, 2014 3:35 PM
To: Bretting, Peter; Richards, Chris; Edward S. Buckler; Buckler, Ed
Subject: RE: DivSeek expression of interest

Hi Chris and Ed—just an update. Conferring with State Dept. and USDA/ARS, we believe that ARS can sign the DivSeek expression of interest. But now we must identify whom in ARS should actually sign it!

So, keep tuned, I'll continue to work on it this week.

Thanks,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Bretting, Peter
Sent: Friday, September 26, 2014 4:16 PM
To: Richards, Chris; Edward S. Buckler; Buckler, Ed
Subject: FW: DivSeek expression of interest

Hi Chris and Ed—would you like me to reply to Hannes on behalf of ARS?

Thanks,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

duplicate email trail deleted

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Monday, September 29, 2014 1:02 PM
To: Richards, Chris
Subject: Re: PAG workshop

Ok. In the worst case, I guess we could do a space holder name and then change later if necessary?

H

> Am Sep 29, 2014 um 6:59 PM schrieb "Richards, Chris" <Chris.Richards@ars.usda.gov>:
>
> Well, the sooner the better but typically I like to have the slate of speakers finalized by mid-October.
>
> C.
>
> -----Original Message-----
> From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> Sent: Monday, September 29, 2014 10:56 AM
> To: Richards, Chris
> Subject: Re: PAG workshop
>
> Hi Chris,
>
> Ok, great! Let me get back to you about the name. When do you need to know this by?
>
> Hannes
>
>> Am Sep 29, 2014 um 6:53 PM schrieb "Richards, Chris" <Chris.Richards@ars.usda.gov>:
>>
>> Hannes-
>>
>> Yeah, I think the proposal period for new workshops is well past....but I would be HAPPY to have a talk about DivSeek at the Genomics of Genebanks workshop....who would you suggest be the speaker?
>>
>> Cheers,
>> C.
>>
>> -----Original Message-----
>> From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
>> Sent: Monday, September 29, 2014 10:33 AM
>> To: Richards, Chris
>> Subject: PAG workshop
>>
>> Hi Chris,
>>
>> On a slightly different topic, I was wondering whether you have already finalized your plans for the genebank genomics workshop at PAG next year? We were at one point thinking to organize a DivSeek workshop but of course wouldn't want to overlap with yours. Also I think it may be too late to still put up proposals for new ones.

>> Another option could be to try and give a talk about DivSeek during your workshop. Would that be an option at all?
>>
>> Please advise.
>> Thanks,
>> Hannes
>>
>>
>>
>>

Hutchison, Stasia

From: Richards, Chris
Sent: Monday, September 29, 2014 12:59 PM
To: Hannes Dempewolf
Subject: RE: PAG workshop

Well, the sooner the better but typically I like to have the slate of speakers finalized by mid-October.

C.

-----Original Message-----
From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Monday, September 29, 2014 10:56 AM
To: Richards, Chris
Subject: Re: PAG workshop

Hi Chris,

Ok, great! Let me get back to you about the name. When do you need to know this by?

Hannes

> Am Sep 29, 2014 um 6:53 PM schrieb "Richards, Chris" <Chris.Richards@ars.usda.gov>:
>
> Hannes-
>
> Yeah, I think the proposal period for new workshops is well past....but I would be HAPPY to have a talk about DivSeek at the Genomics of Genebanks workshop....who would you suggest be the speaker?
>
> Cheers,
> C.
>
> -----Original Message-----
> From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> Sent: Monday, September 29, 2014 10:33 AM
> To: Richards, Chris
> Subject: PAG workshop
>
> Hi Chris,
>
> On a slightly different topic, I was wondering whether you have already finalized your plans for the genebank genomics workshop at PAG next year? We were at one point thinking to organize a DivSeek workshop but of course wouldn't want to overlap with yours. Also I think it may be too late to still put up proposals for new ones.
> Another option could be to try and give a talk about DivSeek during your workshop. Would that be an option at all?
>
> Please advise.
> Thanks,
> Hannes
>

>
>
>
> This electronic message contains information generated by the USDA solely for the intended recipients. Any unauthorized interception of this message or the use or disclosure of the information it contains may violate the law and subject the violator to civil or criminal penalties. If you believe you have received this message in error, please notify the sender and delete the email immediately.

Hutchison, Stasia

From: Bretting, Peter
Sent: Friday, September 26, 2014 4:21 PM
To: Richards, Chris; Edward S. Buckler; Buckler, Ed
Subject: RE: DivSeek expression of interest

Haven't read it yet, as I just returned from this year's NGRAC meeting, held in Ames.

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Richards, Chris
Sent: Friday, September 26, 2014 4:17 PM
To: Bretting, Peter; Edward S. Buckler; Buckler, Ed
Subject: RE: DivSeek expression of interest

Yes please. Let's sign it.

C.

From: Bretting, Peter
Sent: Friday, September 26, 2014 2:16 PM
To: Richards, Chris; Edward S. Buckler; Buckler, Ed
Subject: FW: DivSeek expression of interest

Hi Chris and Ed—would you like me to reply to Hannes on behalf of ARS?

Thanks,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191

duplicate email trail deleted

Hutchison, Stasia

From:	Bretting, Peter
Sent:	Friday, September 26, 2014 4:16 PM
To:	Richards, Chris; Edward S. Buckler; Buckler, Ed
Subject:	FW: DivSeek expression of interest
Attachments:	DivSeek expression of interest.pdf

Hi Chris and Ed—would you like me to reply to Hannes on behalf of ARS?

Thanks,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Tuesday, September 23, 2014 11:01 AM
To: Bretting, Peter; Richards, Chris; Edward Buckler
Cc: Marie Haga; Bhatti, Shakeel (AGDT); Ruth Bastow; Powell, Wayne (CGIAR Consortium)
Subject: DivSeek expression of interest

Dear Drs. Peter Bretting, Christoper Richards and Ed Buckler,

After two scoping meetings in January and April 2014, DivSeek is becoming a reality. We have created a small steering and working group to develop a white paper, to initiate the scoping phase and to develop the plan for launching the initiative. A website includes all our work to date (http://www.divseek.org).

A first Partners' Assembly will take place in early 2015, to establish the governance of DivSeek. For your institution to take part in the Assembly, we ask you to return a signed copy of the letter that is attached, by 31 October. Please identify the appropriate person who can sign on behalf of your organization. Signature of the letter constitutes a simple and non-binding expression of interest in DivSeek. Please submit the signed letter to divseek@croptrust.org and pgrfa-treaty@fao.org

duplicate email trail deleted

SUBJECT: Call for non-binding expressions of interest in DivSeek

We are writing to engage you in Diversity Seek (DivSeek), a ground-breaking initiative which we believe will unlock the wealth of crop diversity for global food and nutritional security.

DIVSEEK MISSION AND GOALS

DivSeek's mission is to enable breeders and researchers to mobilize a vast range of plant genetic variation to accelerate the rate of crop improvement and furnish food and agricultural products to the growing human population.

This initiative will work with existing, emerging and future efforts to characterize crop diversity and develop a unified, coordinated and cohesive information management platform to provide easy access to genotypic and phenotypic data associated with genebank germplasm.

DivSeek will help bridge the gap between the information requirements of genebank curators, plant breeders and more targeted upstream biological researchers, to support forward-looking breeding programs, germplasm curation and strategic research.

A white paper describing the initiative in more detail has been developed and is available at www.divseek.org.

This initiative also contributes to on-going efforts for the creation of a global information system on plant genetic resources, under the aegis of the International Treaty on Plant Genetic Resources for Food and Agriculture.[1]

VALUE PROPOSITIONS

DivSeek will generate the following benefits.

1. DivSeek will unleash the full potential of crop diversity by linking germplasm with data through the promotion of common data standards and best practices. This will:

(i) enable interoperability and broaden usability of data;

(ii) facilitate exploration by breeders, researchers and other users;

(iii) enable managers and curators of genebanks and other collections to better leverage large-scale characterization efforts for effective curation; and

(iv) lessen transaction costs and enhance risk management by providing an accepted framework for 'rights management' that enables projects to align with internationally agreed policy frameworks and data sharing principles.

[1] http://www.planttreaty.org/content/GIS

2. DivSeek will promote the open sharing of data, while ensuring transparency and allowing a level of discretion regarding the distribution and use of information and knowledge.

3. DivSeek will support capacity building and training of qualified professionals to implement the DivSeek mission.

4. As DivSeek reaches a critical mass, funding agencies may require their grantees to ensure adherence to best practices for data storage and curation, and thus maximize impact of project outputs.

5. DivSeek will engage with publishers to encourage authors to use the DivSeek framework to make data available for analysis and peer validation.

PRINCIPLES AND ACTIVITIES

DivSeek is founded on a core set of operating principles. This collaborative venture is designed to facilitate sharing, engagement and synergy. Operating in an open, transparent, and inclusive way, the initiative will advance an initial set of activities, including:

1) a scoping exercise to map current practices, identify potential partners and extend the dialogue with the community;

2) the organization of the first Partners' Assembly to review and approve a Charter for DivSeek, which will establish the governance framework; and

3) a kick-off meeting to discuss potential case studies and pilot projects that will build upon current efforts for linking data with germplasm.

In the long term, DivSeek will build a platform that provides access to genotypic, phenotypic, and other types of information linked to physical germplasm.

DivSeek is a community-driven initiative and a facilitation unit has been established to advance the DivSeek initiative. The Global Crop Diversity Trust hosts and implements the facilitation unit jointly with the Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture, and operates it on a day-to-day basis with additional inputs provided by the CGIAR Consortium, the Global Plant Council and other experts/organizations.

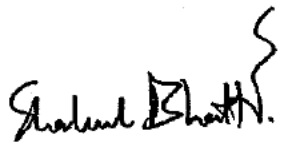
INVITATION

We encourage you to help advance this initiative. On behalf of the joint facilitation unit, we invite your organization to become a recognized partner in DivSeek. If you agree, please countersign and return this letter by 31 October, as a non-binding expression of interest. We will then publicly acknowledge your organization as a DivSeek partner, including through the DivSeek website.

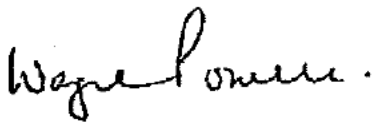
We will follow up by circulating a draft Charter for your preliminary review and with further details concerning the first Partners' Assembly, which is tentatively scheduled for early 2015.

Yours sincerely,

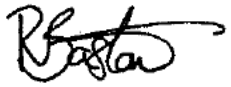
On behalf of the Joint Facilitation Unit
(Marie Haga, Executive Director, Global Crop Diversity Trust)



On behalf of the Joint Facilitation Unit
(Shakeel Bhatti, Secretary, International Treaty on Plant Genetic Resources for Food and Agriculture)



(Wayne Powell, Chief Science Officer, Consortium of the Consultative Group on International Agricultural Research)



(Ruth Bastow, Executive Director, Global Plant Council)

Countersigned on behalf of _____.

DATE

NAME, POSITION AND SIGNATURE

_____ _____

_____

_____

Hutchison, Stasia

From: Richards, Chris
Sent: Friday, September 26, 2014 12:16 PM
To: Hannes Dempewolf
Subject: RE: crop diversity indicator

That's fine...let me craft a papragraph outlining the approach I was thinking about and get back to you and Charlotte.

Cheers,
C.

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Friday, September 26, 2014 1:44 AM
To: Richards, Chris
Cc: Charlotte Lusty
Subject: Re: crop diversity indicator

Hi Chris,

OK, sounds good. A skype chat today is difficult for me, since my afternoon is packed. Do you have any specific questions that I could help answer via email? Perhaps you could draft a brief paragraph with some ideas that I could discuss here internally and give you feedback on? The genebank manager's AGM is the week after next, so it would be neat if we could have a bit of a plan ready by then. I am copying here also Charlotte, since she is very much involved in our thinking around this.

Cheers,
Hannes

On Fri, Sep 26, 2014 at 12:43 AM, Richards, Chris <Chris.Richards@ars.usda.gov> wrote:

Hi Hannes-

Gosh I left for [redacted] ans just got back today. You bet I'm interested!

I think I might have someone in mind I'll need to check.

Perhaps a conversation via Skype tomorrow? Sorry for the delay....I've had no email out on this [redacted].

Cheers,

C.

duplicate email trail deleted

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Friday, September 26, 2014 3:44 AM
To: Richards, Chris
Cc: Charlotte Lusty
Subject: Re: crop diversity indicator

Hi Chris,

OK, sounds good. A skype chat today is difficult for me, since my afternoon is packed. Do you have any specific questions that I could help answer via email? Perhaps you could draft a brief paragraph with some ideas that I could discuss here internally and give you feedback on? The genebank manager's AGM is the week after next, so it would be neat if we could have a bit of a plan ready by then. I am copying here also Charlotte, since she is very much involved in our thinking around this.

Cheers,
Hannes

On Fri, Sep 26, 2014 at 12:43 AM, Richards, Chris <Chris.Richards@ars.usda.gov> wrote:

> Hi Hannes-
>
> Gosh I left for [redacted] ans just got back today. You bet I'm interested!
>
> I think I might have someone in mind I'll need to check.
>
> Perhaps a conversation via Skype tomorrow? Sorry for the delay....I've had no email out on this [redacted]
>
> Cheers,
>
> C.
>
> **From:** Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> **Sent:** Thursday, September 25, 2014 8:01 AM
> **To:** Richards, Chris
> **Subject:** Re: crop diversity indicator

duplicate email trail deleted

Hutchison, Stasia

From:	Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent:	Tuesday, September 23, 2014 11:01 AM
To:	Bretting, Peter; Richards, Chris; Edward Buckler
Cc:	Marie Haga; Bhatti, Shakeel (AGDT); Ruth Bastow; Powell, Wayne (CGIAR Consortium)
Subject:	DivSeek expression of interest
Attachments:	DivSeek expression of interest.pdf

Dear Drs. Peter Bretting, Christoper Richards and Ed Buckler,

After two scoping meetings in January and April 2014, DivSeek is becoming a reality. We have created a small steering and working group to develop a white paper, to initiate the scoping phase and to develop the plan for launching the initiative. A website includes all our work to date (http://www.divseek.org).

A first Partners' Assembly will take place in early 2015, to establish the governance of DivSeek. For your institution to take part in the Assembly, we ask you to return a signed copy of the letter that is attached, by 31 October. Please identify the appropriate person who can sign on behalf of your organization. Signature of the letter constitutes a simple and non-binding expression of interest in DivSeek. Please submit the signed letter to divseek@croptrust.org and pgrfa-treaty@fao.org

A draft Charter, setting forth the principles and institutional mechanisms of DivSeek, will be distributed in advance of the Partners' Assembly.

We continue to promote DivSeek as an open and transparent initiative, and look forward to your further engagement.

Kind regards,

Marie Haga, Shakeel Bhatti, Ruth Bastow and Wayne Powell

On behalf of the joint facilitation unit:

--

Hannes Dempewolf
Scientist and Project Manager
Global Crop Diversity Trust
Platz der Vereinten Nationen 7

53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

SUBJECT: Call for non-binding expressions of interest in DivSeek

We are writing to engage you in Diversity Seek (DivSeek), a ground-breaking initiative which we believe will unlock the wealth of crop diversity for global food and nutritional security.

DIVSEEK MISSION AND GOALS

DivSeek's mission is to enable breeders and researchers to mobilize a vast range of plant genetic variation to accelerate the rate of crop improvement and furnish food and agricultural products to the growing human population.

This initiative will work with existing, emerging and future efforts to characterize crop diversity and develop a unified, coordinated and cohesive information management platform to provide easy access to genotypic and phenotypic data associated with genebank germplasm.

DivSeek will help bridge the gap between the information requirements of genebank curators, plant breeders and more targeted upstream biological researchers, to support forward-looking breeding programs, germplasm curation and strategic research.

A white paper describing the initiative in more detail has been developed and is available at www.divseek.org.

This initiative also contributes to on-going efforts for the creation of a global information system on plant genetic resources, under the aegis of the International Treaty on Plant Genetic Resources for Food and Agriculture.[1]

VALUE PROPOSITIONS

DivSeek will generate the following benefits.

1. DivSeek will unleash the full potential of crop diversity by linking germplasm with data through the promotion of common data standards and best practices. This will:

(i) enable interoperability and broaden usability of data;

(ii) facilitate exploration by breeders, researchers and other users;

(iii) enable managers and curators of genebanks and other collections to better leverage large-scale characterization efforts for effective curation; and

(iv) lessen transaction costs and enhance risk management by providing an accepted framework for 'rights management' that enables projects to align with internationally agreed policy frameworks and data sharing principles.

[1] http://www.planttreaty.org/content/GIS

2. DivSeek will promote the open sharing of data, while ensuring transparency and allowing a level of discretion regarding the distribution and use of information and knowledge.

3. DivSeek will support capacity building and training of qualified professionals to implement the DivSeek mission.

4. As DivSeek reaches a critical mass, funding agencies may require their grantees to ensure adherence to best practices for data storage and curation, and thus maximize impact of project outputs.

5. DivSeek will engage with publishers to encourage authors to use the DivSeek framework to make data available for analysis and peer validation.

PRINCIPLES AND ACTIVITIES

DivSeek is founded on a core set of operating principles. This collaborative venture is designed to facilitate sharing, engagement and synergy. Operating in an open, transparent, and inclusive way, the initiative will advance an initial set of activities, including:

1) a scoping exercise to map current practices, identify potential partners and extend the dialogue with the community;

2) the organization of the first Partners' Assembly to review and approve a Charter for DivSeek, which will establish the governance framework; and

3) a kick-off meeting to discuss potential case studies and pilot projects that will build upon current efforts for linking data with germplasm.

In the long term, DivSeek will build a platform that provides access to genotypic, phenotypic, and other types of information linked to physical germplasm.

DivSeek is a community-driven initiative and a facilitation unit has been established to advance the DivSeek initiative. The Global Crop Diversity Trust hosts and implements the facilitation unit jointly with the Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture, and operates it on a day-to-day basis with additional inputs provided by the CGIAR Consortium, the Global Plant Council and other experts/organizations.

INVITATION

We encourage you to help advance this initiative. On behalf of the joint facilitation unit, we invite your organization to become a recognized partner in DivSeek. If you agree, please countersign and return this letter by 31 October, as a non-binding expression of interest. We will then publicly acknowledge your organization as a DivSeek partner, including through the DivSeek website.

We will follow up by circulating a draft Charter for your preliminary review and with further details concerning the first Partners' Assembly, which is tentatively scheduled for early 2015.

Yours sincerely,

On behalf of the Joint Facilitation Unit
(Marie Haga, Executive Director, Global Crop Diversity Trust)

On behalf of the Joint Facilitation Unit
(Shakeel Bhatti, Secretary, International Treaty on Plant Genetic Resources for Food and Agriculture)

(Wayne Powell, Chief Science Officer, Consortium of the Consultative Group on International Agricultural Research)

(Ruth Bastow, Executive Director, Global Plant Council)

Countersigned on behalf of _____.

DATE

NAME, POSITION AND SIGNATURE

_____

_____

_____

_____

Hutchison, Stasia

From: Bretting, Peter
Sent: Tuesday, September 16, 2014 4:42 PM
To: Richards, Chris
Subject: RE: DivSeek - meeting of governance and legal group - report

Will do. I'd expected to receive a document by now, but haven't. Not sure what's up.

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Richards, Chris
Sent: Tuesday, September 16, 2014 4:08 PM
To: Bretting, Peter
Subject: RE: DivSeek - meeting of governance and legal group - report

Hey Peter-

What kinds of changes did you recommend and have they been incorporated? Please keep me posted on whether we're going to be a part of this.

Hope all is well with you.

Cheers,
C.

From: Bretting, Peter
Sent: Thursday, August 28, 2014 1:59 PM
To: Richards, Chris
Subject: RE: DivSeek - meeting of governance and legal group - report

Thanks, Chris. I just checked with Hannes regarding whether the status of the letter of intent. I'd suggested some possible changes to the text.

Peter

Peter Bretting
USDA/ARS Office of National Programs

Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Richards, Chris
Sent: Thursday, August 28, 2014 1:48 PM
To: Bretting, Peter
Subject: FW: DivSeek - meeting of governance and legal group - report

Hey Peter-

Are we signing on to this letter of interest and including the USDA in the 'momentum'? I think it would be beneficial to become a member of the governance structure instead of watching from the sidelines. What do you think?

Cheers,

Chris

From: Scott Allen Jackson [mailto:sjackson@uga.edu]
Sent: Wednesday, August 27, 2014 6:47 AM
To: Manzella, Daniele (AGDT)
Cc: daniele.manzella@croptrust.org; Robbie Waugh; Michael Abberton (IITA); Sarah Ayling; Bhatti, Shakeel (AGDT); Paula Bramel; David Ellis; Ellul, Philippe (CGIAR Consortium); Marie Haga; Peter Langridge; Lopez, Francisco (AGDT); Emily Marden; Lisette Mascarenhas; Susan McCouch; Matija Obreza; Peter Phillips; Powell, Wayne (CGIAR Consortium); Richards, Chris; Loren Rieseberg; Ruaraidh Hamilton; Ulrich Schurr; Theo van Hintum; p.wenzl@cgiar.org; Ruth Bastow; Luigi Guarino; Gruissem Wilhelm; Hannes Dempewolf; Reno Pontarollo; Carolina Roa
Subject: Re: DivSeek - meeting of governance and legal group - report

THanks Daniele, This is great and will help to maintain momentum.

Scott Jackson
On Aug 25, 2014, at 4:02 AM, Manzella, Daniele (AGDT) <Daniele.Manzella@fao.org> wrote:

Dear all,

Please find attached the informal report of the meeting of the DivSeek governance and legal group. Thank you again to the experts who have participated in the meeting.

As you may recall, the participants in the Bonn meeting had asked the group to advise on the transitioning of DivSeek from its current interim governance structure to a more permanent setting. In this concise report, you will find an indication of the next steps, which consist of: a) a letter to call for non-binding expressions of interest by the participating stakeholders; b) the first DivSeek partners' meeting where a Charter, setting forth the principles and institutional mechanisms of DivSeek, will be reviewed and approved.

The DivSeek joint facilitation unit will soon be dispatching the letter, which contains an indicative timeline for signature and return (15 October), so please feel free to alert the counterparts within your respective institutions.

The governance and legal experts will prepare a draft Charter for in-advance circulation to those who will reply positively to the letter. The tentative schedule for the first DivSeek partners' meeting is in early 2015, possibly in conjunction with the annual PAG in San Diego.

Please do not hesitate to contact me for any question related to the work of the governance and legal group. I am looking forward to continue working with you towards making DivSeek a reality.

Kind regards,

Daniele
(on behalf of the joint facilitation unit)

<image001.png>
Daniele Manzella
Joint Liaison Officer (Global Crop Diversity Trust)
Secretariat of the International Treaty on PGRFA
Viale delle Terme di Caracalla, 1
00153 Rome, Italy
Phone +39 06 5705 6180
Skype [redacted]
www.planttreaty.org
www.croptrust.org

<Vancouver report.pdf>

Hutchison, Stasia

From: Richards, Chris
Sent: Saturday, August 02, 2014 3:55 PM
To: Hannes Dempewolf
Cc: Greg Baute
Subject: RE: Greg visit

Wonderful. I'm all for it. Greg, as we get closer to the date, keep in touch and keep me posted. It would be great to get together and discuss some projects I'm working on. Thanks for being a good catalyst, Hannes! Hope you are not moving among times zone more than once every few days!

Cheers,
Chris

Christopher Richards, Ph.D.
Population Geneticist
National Center for Genetic Resources
United States Department of Agriculture
Colorado State University
Fort Collins
USA

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Saturday, August 02, 2014 8:50 AM
To: Richards, Chris
Cc: Greg Baute
Subject: Greg visit

Hi Chris,

I am sure you'll remember Greg Baute, a PhD student of Loren's who also attended the Asilomar meeting and spoke at your PAG workshop etc. When I mentioned to Greg your work on hops diversity and your work on the expanded gap analysis approach he was very interested and told me about an upcoming trip to Colorado that he is planning in late October/early November. In case youa are around then, it seems like it would be a perfect opportunity for the two of you to meet and discuss the possible use of some of the Rieseberg lab's sunflower data sets to test your predictions. I understand Greg would even be able to come to Ft Collins and meet you at the USDA facilities. I have copied him here for easy follow up.

Cheers,
Hannes

--
Hannes Dempewolf
Scientist
Global Crop Diversity Trust

Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Bretting, Peter
Sent: Thursday, August 21, 2014 6:34 AM
To: Richards, Chris
Subject: RE: Big data and "data wrangling"

Thanks for the reminder, Chris. I'll check on the progress of the letter of intent. I corresponded with Hannes about it some time ago.

It's been a much a calmer latter part of the summer as compared to earlier, which featured [redacted]

[redacted]

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Richards, Chris
Sent: Wednesday, August 20, 2014 6:43 PM
To: Bretting, Peter
Subject: RE: Big data and "data wrangling"

Hi Peter-

Thanks for the link...yes, saw that article on Monday, too.
The premise seemed to be that in order to have software work, one still needs "data janitors" to mop up the messy formats and disparate categories with human judgment to put the data together. I read it as more like man vs. the machine...and if Ford hasn't gotten the answer yet with all of their investments in consumer data it might be that data wrangling will be here to stay at some level.

You're so right, the NPGS is one big data wrangling operation!

Speaking of data wrangling, any movement on the Letter of Intent from the DivSeek folks? It seems to me if we are really thinking seriously about integrating these kinds of data (physical accessions and digital genomes) we're going to have to invest in a sizable data janitorial staff!

[redacted]

[redacted]

Cheers,
Chris

From: Bretting, Peter [mailto:Peter.Bretting@ARS.USDA.GOV]
Sent: Monday, August 18, 2014 6:09 AM
To: pgoc@grin.barc.usda.gov; curators@ars-grin.gov
Subject: Big data and "data wrangling"

See article in NY Times this AM http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?emc=edit_th_20140818&nl=todaysheadlines&nlid=14285777

"Data wrangling" is surely a vivid term for a major occupation for NPGS curators, information managers, etc. I wonder if any of the new software tools under development by software start-up companies could help with managing and analyzing information associated with plant genetic resources?

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Saturday, August 02, 2014 10:50 AM
To: Richards, Chris
Cc: Greg Baute
Subject: Greg visit

Hi Chris,

I am sure you'll remember Greg Baute, a PhD student of Loren's who also attended the Asilomar meeting and spoke at your PAG workshop etc. When I mentioned to Greg your work on hops diversity and your work on the expanded gap analysis approach he was very interested and told me about an upcoming trip to Colorado that he is planning in late October/early November. In case youa are around then, it seems like it would be a perfect opportunity for the two of you to meet and discuss the possible use of some of the Rieseberg lab's sunflower data sets to test your predictions. I understand Greg would even be able to come to Ft Collins and meet you at the USDA facilities. I have copied him here for easy follow up.

Cheers,
Hannes

--
Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Thursday, July 17, 2014 8:56 PM
To: Hannes Dempewolf
Subject: DivSeek white paper and website

Dear participants of the San Diego meeting,

This is a brief update on some recent developments with regards to the Diversity Seek initiative. First of all, thank you for providing feedback on the draft DivSeek white paper, which has now been finalized. For the purpose of making the white paper available publicly, we developed an initial DivSeek website that describes the goals of the emerging initiative. The white paper is now available here: www.divseek.org

As you know, Divseek is a community-driven initiative and we are now in the process of developing an 'expression of interest' letter that will soon be distributed to many organizations and institutions in the community who may be interested to join the initiative. We then hope to draft a more formal governance structure by the end of the year with inputs from all those who have expressed their interest to join the initiative.

As part of the interim governance structure of DivSeek, a facilitation unit was formed. The Global Crop Diversity Trust hosts and implements the facilitation unit jointly with the Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture, and operates it on a day to day basis with additional inputs provided by the CGIAR consortium, the Global Plant Council and other experts/organizations.

On a related note, as part of the consultation on a global information system for plant genetic resources, the ITPGRFA secretariat has just launched a survey on genomics, phenomics and data sharing that we would like to encourage you to complete. The idea is that the results of this survey will also be analyzed in the context of a scoping study for DivSeek. You can find the survey in three languages here: English:
[redacted]
French: [redacted]
Spanish: [redacted]

So, thank you again for your continued interest and support of the DivSeek initiative. We will be in touch with you again shortly, once the expression of interest letter has been finalized.

Warm regards,
Hannes (on behalf of the DivSeek facilitation unit)

--
Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]

www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Richards, Chris
Sent: Monday, July 07, 2014 12:46 PM
To: Hannes Dempewolf
Subject: RE: thanks!

Thanks for the information, Hannes. It was great to host you here...even for a little while.

Hope your travels are going to plan.

I will run down this information shortly. Really appreciate your ideas on this front. It's hard to find this kind of data easily. I think using these data for validating these sampling models would be an entirely complimentary objective to these research groups and as such wouldn't be viewed as a form of competition.

I will continue to advocate for reference genomes and will also convene a meeting of curators (hopefully at the next PGOC/curator meeting) in ways to prepare collections for the kinds of characterization the DivSeek envisions.

Stay well!

Cheers,
Chris

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Friday, July 04, 2014 8:13 AM
To: Richards, Chris
Subject: thanks!

Hi Chris,

Thanks again for hosting me yesterday in Ft Collins. I really enjoyed my visit and learning a lot more about what you guys are up to. All very impressive.

I just wanted to send you a some information that we had discussed:

This is the barley 1K website: https://sites.google.com/site/barley1k/intro
Attached please find the draft of the chapter on CWR genomics (would be good if you could not share this further at this time, since it's still in the publication pipeline).

I have looked through my email record to see whether I could pull out a general project description of Doug Cook's effort at Davis. However, I wasn't able to find any document that describes it. I therefore copy and pasted below a description that he had sent me once in an email. It should give you a good understanding of whether they are hoping to take their effort. Again, it would probably best if you couldn't share this further, but if you are interested to learn more about it, you can get in touch with him directly (drcook@ucdavis.edu), or I can put you in touch also, if you feel this would be helpful.

Happy 4th of July!
Hannes

Here the excerpt of Doug's email:

Our project is just starting, so though you may have read bits on the web its unlikely you have a sense of what we are actually about.

We are basically re-inventing how wild germplasm is collected, analyzed, and utilized. When I say re-inventing, I mean relative to the ways in which crop researchers have traditionally made their collections. Ecologists have used principles like ours for years, but without the advent of genomics and certain other enabling technologies. We start with the premise that collections must encompass the breadth of functional diversity present in wild populations -- I use the word "populations" intentionally, because populations are the functional entities that comprise species, more so than are individuals. Populations are the leverage point for selection, and they are what respond in evolutionary terms.

We combine ecology, community genomics, quantitive modeling and climate/edaphic surveys to drive our collations. In chickpea, we began with a climate model for the species and spent two months continuously on the ground prospecting. We focused not only on climate variables, but also on soil type variation and altitudinal gradients. Chickpea's wild relatives have a narrow geographical distribution, which simplifies the task. But the area is still 60K square km of space, so the task is non-trivial.

To date we have identified physical sites spanning this 60 sq km area. The sites encompass and expand well beyond the origin of the current and rather limited material in the international collections. We have surveyed ~1200 individuals, mostly from C. reticulatum and a few from C. echinospermum. These individuals derive from 20 locations, 15 of which are Cr and 5 of which are Ce. The sites span altitudinal, temperature, soil type and rainfall extremes for the species. At each site we collected DNA from every individual that we could identify -- in other words, we wanted to be comprehensive. We have GPS mapped each sample to meter scale resolution. From each site we collected and have now analyzed chemical characteristics in 5 soil samples (we would like to have more soil samples, but resources are limiting). For the Cr sites, we planted 100 micro sensors in the soil, so that we have a seeds-eye view of temperature and humidity at 4 hr intervals through the year. Taken together, these data allows us to know the spatial scales of genomic, soil and climatic variation, which is critical to associating genomic variation with evolutionary pressures. Among the simple analyses we conduct will be to determine co-variation of all variables within and among sites, leading to nomination of candidate genes.

For the 1200 plant samples that we currently have, we have sequenced a common 5% of each genome, which allows us to assay genomic variation at 300,000 loci among all accessions, for which we find ~58k segregating sites. This is the first phase of a more detailed genomic/genetic analysis, that will ultimately permit us to infer which regions of the plant genome have been under selection in which populations. The information will also drive rational re-surveying for new populations over the next 2 years. From the first data set, we know that Cr represents ~7 populations among the 15 sites, and that the 5 Ce sites represent 2 populations. We have calculated genetic diversity, Fst, population structure, major and minor alleles, and identified the major and minor haplotypes in each site/populations.

The next phase of genomics will be to sequence the full genomes of most of the 1200 individuals, which will permit more detailed and informative analyses of LD, allele frequencies, selective sweeps, etc.

We have also begun phenotyping with several goals in mind initially. Phenotypic variation, if properly and rigorously collected, can be a useful complementary data set to genetic variation. Among the initial traits are seed characters, plant architecture, flowering time and N-responsiveness, not all of which are collected yet. Ultimately, of course, we will explicitly focus on adaptations to heat, drought, frost, etc.

We are currently selecting a set of 20 representative Cr wild accessions to develop experimental populations as a prelude to breeding. Our populations will be nested association mapping panels and advanced backcross introgression lines. We will not use the simple and relatively uninformative pair-wise diversity metrics that are common in prior work; instead we are using more informative measures of genome diversity and allele representation, including population-level sampling based an frequent alleles, which permits us to enrich for functional loci that under natural selection within and among sites.

So this is where we are at generally. We are only 3 months into the project, so there is much more to come, and undoubtedly many obstacles to over come.

--
Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Friday, July 04, 2014 10:13 AM
To: Richards, Chris
Subject: thanks!
Attachments: Chapter_19_on_Genomics _CWR_Feb_19.docx

Hi Chris,

Thanks again for hosting me yesterday in Ft Collins. I really enjoyed my visit and learning a lot more about what you guys are up to. All very impressive.

I just wanted to send you a some information that we had discussed:

This is the barley 1K website: https://sites.google.com/site/barley1k/intro
Attached please find the draft of the chapter on CWR genomics (would be good if you could not share this further at this time, since it's still in the publication pipeline).

I have looked through my email record to see whether I could pull out a general project description of Doug Cook's effort at Davis. However, I wasn't able to find any document that describes it. I therefore copy and pasted below a description that he had sent me once in an email. It should give you a good understanding of whether they are hoping to take their effort. Again, it would probably best if you couldn't share this further, but if you are interested to learn more about it, you can get in touch with him directly (drcook@ucdavis.edu), or I can put you in touch also, if you feel this would be helpful.

Happy 4th of July!
Hannes

Here the excerpt of Doug's email:

> Our project is just starting, so though you may have read bits on the web its unlikely you have a sense of what we are actually about.
>
> We are basically re-inventing how wild germplasm is collected, analyzed, and utilized. When I say re-inventing, I mean relative to the ways in which crop researchers have traditionally made their collections. Ecologists have used principles like ours for years, but without the advent of genomics and certain other enabling technologies. We start with the premise that collections must encompass the breadth of functional diversity present in wild populations -- I use the word "populations" intentionally, because populations are the functional entities that comprise species, more so than are individuals. Populations are the leverage point for selection, and they are what respond in evolutionary terms.
>
> We combine ecology, community genomics, quantitive modeling and climate/edaphic surveys to drive our collations. In chickpea, we began with a climate model for the species and spent two months continuously on the ground prospecting. We focused not only on climate variables, but also on soil type variation and altitudinal gradients. Chickpea's wild relatives have a narrow geographical distribution, which simplifies the task. But the area is still 60K square km of space, so the task is non-trivial.

To date we have identified physical sites spanning this 60 sq km area. The sites encompass and expand well beyond the origin of the current and rather limited material in the international collections. We have surveyed ~1200 individuals, mostly from C. reticulatum and a few from C. echinospermum. These individuals derive from 20 locations, 15 of which are Cr and 5 of which are Ce. The sites span altitudinal, temperature, soil type and rainfall extremes for the species. At each site we collected DNA from every individual that we could identify -- in other words, we wanted to be comprehensive. We have GPS mapped each sample to meter scale resolution. From each site we collected and have now analyzed chemical characteristics in 5 soil samples (we would like to have more soil samples, but resources are limiting). For the Cr sites, we planted 100 micro sensors in the soil, so that we have a seeds-eye view of temperature and humidity at 4 hr intervals through the year. Taken together, these data allows us to know the spatial scales of genomic, soil and climatic variation, which is critical to associating genomic variation with evolutionary pressures. Among the simple analyses we conduct will be to determine co-variation of all variables within and among sites, leading to nomination of candidate genes.

For the 1200 plant samples that we currently have, we have sequenced a common 5% of each genome, which allows us to assay genomic variation at 300,000 loci among all accessions, for which we find ~58k segregating sites. This is the first phase of a more detailed genomic/genetic analysis, that will ultimately permit us to infer which regions of the plant genome have been under selection in which populations. The information will also drive rational re-surveying for new populations over the next 2 years. From the first data set, we know that Cr represents ~7 populations among the 15 sites, and that the 5 Ce sites represent 2 populations. We have calculated genetic diversity, Fst, population structure, major and minor alleles, and identified the major and minor haplotypes in each site/populations.

The next phase of genomics will be to sequence the full genomes of most of the 1200 individuals, which will permit more detailed and informative analyses of LD, allele frequencies, selective sweeps, etc.

We have also begun phenotyping with several goals in mind initially. Phenotypic variation, if properly and rigorously collected, can be a useful complementary data set to genetic variation. Among the initial traits are seed characters, plant architecture, flowering time and N-responsiveness, not all of which are collected yet. Ultimately, of course, we will explicitly focus on adaptations to heat, drought, frost, etc.

We are currently selecting a set of 20 representative Cr wild accessions to develop experimental populations as a prelude to breeding. Our populations will be nested association mapping panels and advanced backcross introgression lines. We will not use the simple and relatively uninformative pair-wise diversity metrics that are common in prior work; instead we are using more informative measures of genome diversity and allele representation, including population-level sampling based an frequent alleles, which permits us to enrich for functional loci that under natural selection within and among sites.

So this is where we are at generally. We are only 3 months into the project, so there is much more to come, and undoubtedly many obstacles to over come.

--
Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Chapter 19

Using genomic approaches to unlock the potential of CWR for crop adaptation to climate change

Gregory J. Baute[1]
Hannes Dempewolf[2]
Loren H. Rieseberg[1,3]

1. Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada
2. Global Crop Diversity Trust, Platz der Vereinten Nationen 7, 53113 Bonn, Germany
3. Department of Biology, Indiana University, 1001 East Third Street, Bloomington, Indiana 47405, USA

Email: gbaute@biodiversity.ubc.ca

a First level

19.1 Introduction

Ultimately, it is the genome of a crop wild relative (CWR) that will be utilized to improve our crops, not its phenotype, collection locality, or its history of local adaptation. With this in mind, we argue that germplasm resources in gene banks or in the wild may eventually be best explored, surveyed, or mapped, at the level of the genome as we become better at predicting the breeding value of individual accessions and/or alleles. A growing array of genomic technologies is becoming more widely available and affordable. Genome-wide surveys of large numbers of genetic markers in large populations are starting to allow us to truly understand what diversity is available to breeders within crop wild relatives. Faced with limited resources, plant breeders must choose carefully which germplasm to evaluate and incorporate into their improvement programs. This selection is especially difficult with wild relatives. Germplasm curators are also faced with difficult issues with regard to CWRs, having to decide which materials to conserve over the long-term and where to focus germplasm acquisition efforts. Genomic data offers solutions to a number of practical issues

surrounding the conservation and use of exotic and wild germplasm, as well as more effective and efficient ways to use these resources to help adapt crops to a changing climate.

The plummeting cost of DNA sequencing is making genomic data more accessible to a larger number of researchers. Many plant reference genomes have been sequenced and are now available and even more are underway or are planned (Hamilton and Robin Buell 2012). The assembly of reference genomes, and their annotation, greatly facilitates population genomic studies. Many projects are currently re-sequencing the genomes (or large portions of genomes) of numerous cultivars and their wild relatives. Much of this work so far has focused on elite material, but some efforts are now underway to sequence entire gene bank holdings. This means hundreds of thousands, and eventually millions, of plant genomes will be sequenced at least to some degree. This information may facilitate the efficient use of CWRs and other gene bank material to adapt our crops to a changing climate. It also presents researchers and plant breeders with new opportunities and challenges.

Here we discuss how genome technologies may be used to enhance conservation and use of CWRs to breed crops for a changing climate. There are many obvious uses of genomic data and novel uses are likely to become apparent in the coming decades. After discussing current and potential uses of genomic technologies, we will briefly review challenges to the generation and analysis of genomic data. We also offer perspectives on how both sequencing technologies, data storage and analysis could develop in the future. The level of adoption of genomic approaches will largely depend on the gains in efficiency and effectiveness that crop improvement programs are able to achieve. Genomic methods promise to facilitate the maintenance and characterization of current CWRs, help breeders select CWRs to use in their breeding programs and improve the speed and efficiency of their use once in a breeding program.

a First level

19.2 Use of genomic technologies

b Second level

19.2.1 Gene bank curation

Gene banks hold material of immense economic value (Smale and Koo 2003) as they contain germplasm that is key to future global food production. In spite of this great value, the exact number of accessions held in gene banks remains unknown. Estimates of the number of unique accessions vary widely, but the figure most commonly quoted is circa 2 million (FAO 2010). The natural history of the germplasm can vary greatly, with different collections capturing different amounts of diversity. Many accessions have complicated histories of collection, rounds of regeneration, involvement in breeding cycles and sometimes movement among institutions around the globe. Complicating the situation further, handling errors are nearly unavoidable when dealing with large collections. This leaves germplasm curators and potential users with many important but largely unanswered questions. How much diversity is in each accession? Does the passport information accurately describe the identity of an accession? How similar are accessions? How much diversity can be found within a crop's wild relatives? What subset of a collection captures the most diversity? Which set of accessions captures unique diversity? Which accessions are duplicates of one another?

Genomic data offer an unprecedented level of accounting of a gene bank's holdings, thereby allowing questions pertaining to the diversity and nature of accessions to be more confidently addressed. The uniqueness and diversity of accessions can be identified using phylogenetic

approaches, population genetics statistics (Huang et al. 2012), and multivariate analyses (Cavanagh et al. 2013; Romay et al. 2013). These analyses may justify the splitting or lumping of accessions, or may identify material that has been duplicated, mislabelled or misidentified. Various molecular markers have been used to identify and remove redundant lines from germplasm collections (van Treuren et al. 2010). Genomic technologies offer the marker density required to screen germplasm for rare alleles, which may have not have been identified previously. The genomic data from CWRs can also help to resolve taxonomic questions, possibly resolving species relationships or identifying cryptic species, sub-populations or ecotypes of a taxa.

Analyses of genomic data can facilitate CWR germplasm management through the calculation of statistics that summarize how genetic diversity is partitioned within and among accessions. For each CWR, collection size should scale positively with species diversity and negatively with its potential usefulness. A larger collection does not necessarily capture more diversity given both the idiosyncratic distribution of genetic variation within and among accessions and well-known collection biases (Hijmans et al. 2000; Smale and Koo 2003). Although it is not yet possible to predict the breeding value of a species based solely on genomic information, genetic diversity can be used to assess whether the amount of diversity in a gene bank scales according to estimated population sizes, range sizes, or habitat diversity. Identifying distinct genetic units is critical for ensuring that important, unique and locally adapted, material is safely stored in gene banks. With the advent of high-density genotyping it will now become possible to assess, with considerable confidence, whether collections for a given taxa are saturated in terms of genetic diversity. Collection of new materials should only be considered when substantial gains in diversity are to be expected.

b Second level

19.2.2 Germplasm collection

The acquisition of new CWR germplasm through field collecting is an important process that can also be aided by genomic information. Collections are often incomplete and the diversity of the CWRs of many crops are not captured to a sufficient extent in current gene bank holdings (FAO 2010). Resource limitations and plant life histories often lead to small windows of opportunity for collecting, mandating maximum efficiency. Effective and efficient collecting missions are particularly important for diversity that is threatened in the wild or in the field. Considering genomic information for collecting may be key to capturing under-represented germplasm. Genomic information can be combined with information on the geographic and eco-geographic distribution of a species to maximize the effectiveness of collecting. Current 'gap analyses' use occurrence records from herbaria and other sources to generate models of where a species is likely to occur. The distribution of existing germplasm collections is then overlaid on these distribution models and gaps (areas where a species is likely to occur but there are no collections) identified(Ramírez-Villegas et al. 2010). However, accession abundance and collection locality alone are not truly representative of species diversity and distribution, as the genetic diversity of a species is rarely spatially homogeneous. Species frequently undergo range changes and population size fluctuations. Some areas may also contain more diversity, such as glacial refugia (Petit and Excoffier 2009), or they could contain unique alleles not found elsewhere. It can therefore be expected that the integration of genomics information into gap analyses will yield more precise and accurate estimates of current 'gaps' in gene bank collections.

b Second level

19.2.3 Germplasm selection

Phenotyping is one of the biggest bottlenecks in crop improvement programs, especially when

CWRs are involved, because they may require multiple generations of introgression into a cultivated genomic background before evaluation. Collections of CWRs are often larger than what can be realistically evaluated and many have limited or incomplete passport information. Breeders can only select a subset of the material for evaluation and for incorporation into a breeding program. Germplasm selection by breeders is typically based on natural history records, previous phenotypic and genotypic evaluations, curator knowledge about a given accession's characteristics, prior use in breeding programs and the breeders' needs. Candidates for use are sometimes drawn from a core collection, which represents subsets of accessions that are designed to capture the greatest amount of genetic diversity in as few accessions as possible(Brown 1989).

Core collections are generally made by dividing accessions into phenotypic, life history, taxonomic or ecogeographic groups and then selecting representatives from each group. Often several metrics are used together to establish a core collection, but molecular markers are particularly useful for this process. Genetic diversity can be assessed with many different kinds of molecular markers, but the marker density obtained from high-throughput sequencing can be several orders of magnitude greater than other methods and will likely give the clearest picture of the patterns of diversity within a collection. With genotypic data, selections can be made in an automated fashion with software that aims to maximize the number of alleles in a selected group of samples(Gouesnard et al. 2001). Not only will deep genome coverage help identify rare alleles, but with high density genetic maps the specific recombination events and allelic combinations found in an accession can be determined too. This means that core collections can be designed not just to maximize allelic diversity but also the number of different allelic combinations or haplotypes.

Another approach for identifying germplasm that may contain alleles conferring tolerance to a given abiotic stress, herbivore, or pathogen is the Focused Identification of Germplasm Strategy (FIGS).

The FIGS approach uses collection locality information to select accessions most likely to have desired traits (Mackay and Street 2004). However, a crucial assumption underlying FIGS and similar approaches is that the material is locally adapted, which is not always the case, especially for crop wild relatives where populations can be small and/or migration rates are high (Rieseberg and Willis 2007; Mallet 2007). Local adaptation can be identified with reciprocal transplant experiments, with the expectation that the material collected near the evaluation site will have superior performance. There are many well-characterized examples of local adaptation and it appears to be frequent in plants. However, many studies fail to detect evidence of local adaptation and it may occur less often than generally assumed (Leimu and Fischer 2008). Although transplantation experiments can identify locally adapted germplasm, having to grow and evaluate material to justify the use of the FIGS approach defeats its purpose, which is to reduce the amount of phenotyping needed for germplasm selection.

Population genomic scans have proven to be effective in detecting local adaptation and so could be used as additional criteria for a FIGS-like selection approach. Many such scans examine variation among loci in F_{ST}, which is one of the main statistics used to study population differentiation. Loci with exceptionally high F_{ST} values may be the targets of divergent selection, as they are more highly differentiated between populations than the rest of the genome. This "genome scan" method of detecting selection by identifying outlier high-F_{st} markers has been gaining popularity with the advent of high-throughput sequencing technologies (Roesti et al. 2012; Stölting et al. 2012). An F_{ST} distribution with a fat tail is evidence that local adaptation may have occurred. Furthermore, it identifies specific candidate genes that may be the targets of selection. Placing the selected markers onto a genetic map can reveal regions of the genome that have undergone selection. Under some conditions the reduced effective recombination rates near the selected sites could result in "islands"

of differentiation in the genome (Feder and Nosil 2010; Feder et al. 2012). However, it is important to keep in mind that some demographic processes, such as population bottlenecks and range expansion, can produce F_{st} distributions that are similar to those generated by local adaptation (Excoffier et al. 2009).

When selection has recently favoured a new mutation, and that allele is brought to a high frequency, linked sites will have reduced diversity, creating a longer haplotype in that region. This type of selection, known as a hard sweep, can create areas of elevated linkage disequilibrium and can be detected with methods such as extended haplotype homozygosity (Sabeti et al. 2002) or integrated haplotype score (Voight et al. 2006). Soft sweeps, which involve alleles already present in a population, are much more difficult to detect (Przeworski et al. 2005; Teshima 2006). Alleles that are adapted to specific climatic factors may be detected by their association with those climatic factors (Hancock et al. 2011). Such modeling of allele frequencies with climatic variables has been used to detect subtle signatures of selection in human populations (Hancock et al. 2010), and is a promising method to identify locally adapted alleles in CWRs. When multiple clines along the same environmental factors are present, for example, multiple colonisations of dry environments, this approach will be particularly powerful. Understanding the presence, size, and number of genomic regions that have experienced selection will be an essential component of understanding how local adaptation happens and how it can be harnessed in breeding programs.

When locally adapted alleles are present, several further issues need to be considered before they can be utilized effectively in a breeding program. First, the traits may be highly complex, making introgression into elite material difficult. Secondly, even if the responsible alleles are successfully introduced, nearby maladaptive alleles (i.e., linkage drag) may be a problem. In addition, novel genetic interactions may prevent the expression of the locally adapted traits. Lastly, locally adapted

phenotypes must be useful in the context of cultivation. There are many ways in which plants can evolve tolerances to a given environmental factor. For example, it is possible that a population that is adapted to dry environments harbours alleles for more efficient photosynthetic machinery, better membrane transporters or tolerance to reactive oxygen species that arise during drought stress, all of which may benefit cultivars (Cruz de Carvalho 2008; Schroeder et al. 2013). It is also possible that the population responds to drought by closing its stomata, shutting down and waiting for water to become available, or by escaping drought by flowering early in the season. Although the latter strategies may help wild populations survive, and leave a corresponding signature of selection in the genome, such strategies are not useful in the context of breeding drought tolerance because they may not lead to yield advantages, creating a pitfall for both FIGS and genomics-enabled approaches.

b Second level

19.2.4 Mining CWR diversity

Genomic approaches can also be used to link genes in genomes of CWRs with phenotypes of interest to breeders. Both association mapping and quantitative trait locus (QTL) mapping have benefited greatly from the use of high-throughput genotyping approaches. Indeed, the number of recombination events, rather than marker density, is now the limiting factor in most genetic mapping studies. The number of recombination events assayed can be increased through the use of more distantly related individuals from natural populations, larger mapping populations, or populations from advanced generation crosses. Examples of the latter include Nested Association Mapping (NAM) (McKhann et al. 2004) and Multiparent Advanced Generation Intercross (MAGIC) populations (Cavanagh et al. 2008). NAM uses multiple recombinant inbred line populations that are developed using a common parental line. MAGIC uses a diverse set of founders and involves successive

generations of inter-mating. Both approaches could be powerful tools for mapping traits of interest in CWRs, although some modifications may be required. Many CWRs must undergo one or more generations of backcrossing to elite material before evaluations can take place, in order to reveal traits that will be expressed in an elite background. However, backcrossing can be readily incorporated into NAM and MAGIC-like populations. With respect to the latter, CWRs can be backcrossed into a diverse set of elite lines to provide founders for the MAGIC population, making it possible to assess genetic background effects, as well as to incorporate numerous CWRs into a single mapping population. These types of populations and sufficiently high marker density across the genome can facilitate fine mapping of traits for subsequent introgression into other elite material. Genomic selection, a method which can facilitate more rapid and lower cost gains from breeding, may be useful in reducing the amount of time required for pre-breeding as well (Bernardo 2009). Genomic information could also help during the development of populations, to select lines that capture specific recombination events or to select lines that yield a more even coverage of the wild donor parents' genome, as in chromosomal segment substitution lines, for example.

b Second level

19.2.5 Investigating previous CWR use

Many crop genomic projects have focused primarily on cultivated material. Although the diversity of CWRs cannot be deduced from these sequencing projects, this information can provide insights regarding the genetic basis of domestication and how CWRs have been used for crop improvement in the past.

Crop plants were selected from wild plants over millennia, leaving a signature of selection in the genome, which can be detected using genome scans similar to those employed to detect local

adaptation (see above). Numerous candidate genes have been identified in a variety of crops, including rice (Huang et al. 2012), maize (Hufford et al. 2012), soybeans (Lam et al. 2010), tomato (Koenig and Jiménez-Gómez 2013) and watermelon (Guo et al. 2012). Measures of differentiation or diversity can be used independently, or in combination with metrics such as the cross-population composite likelihood ratio test (Chen et al. 2010), to detect these genes. Groups working with CWRs may consider exploiting this information to rapidly remove undesirable wild traits during pre-breeding.

The improvement of crops following domestication was not only done through recurrent rounds of selection, but also through the introduction of new genetic material from wild species. Some of this introgression from CWRs happened in pre-historic times. Rice was domesticated in southern China, for example, and as it moved into South and Southeast Asia it experienced rampant introgression from the local wild rice species. The introgression from the local wild *Oryza rufipogon* was so extensive that there has been much debate as to whether or not the two main subspecies of present-day cultivated rice arose from separate domestication events. Genomic data has helped resolve this issue (Huang et al. 2012). In another example, high altitude landraces of maize have obtained alleles from a high altitude wild teosinte species. This introgression was also identified using genomic data and STRUCTURE (Pritchard et al. 2000) and HAPMIX (Price et al. 2009) analyses (Hufford et al. 2013).

Plant breeders have made intentional use of wild relatives since at least the 1800s (Knight 1806) and many modern breeding programs make substantial use of wild relatives. Often, however, useful alleles are introgressed without knowledge of their molecular basis or genomic location. With genomic techniques, we can now identify the genomic location and size(s) of these historic introgressions. The profile of a particular introgression alongside pedigree and registration information may enable useful predictions about the function of the introgression. For example, an

elite line may be known to contain unmapped disease resistant alleles from a particular wild relative; sequencing of this line as well as the wild donor could reveal which alleles have been introgressed and bypass the need for mapping. This method is promising but has yet to be carefully tested and would depend on the extent of backcrossing involved and the strength of selection. Phylogenetic analyses correspond closely to pedigrees and can indicate the origin of each component of a cultivar's genome. Pedigrees, when available, can be used to create expectations of relationships that can validate genomic data or vice versa (Lorenz and Hoegemeyer 2013).

a First level

19.3 Technical issues in applying genomics approaches to CWRs

Several issues have to be considered before embarking on a CWR genomics project, including which sequencing technologies and approaches to use, important biological differences between species, and the availability of genomics resources, financial resources and bioinformatics capacity.

b Second level

19.3.1 Genomic technologies

Numerous methods have been developed for high throughput genotyping. Many single nucleotide polymorphism (SNP) chip array platforms exist that are capable of assaying thousands of sites across the genome. SNP arrays must be designed for a specific population or taxon and can be made to assay any number of sites, from hundreds to hundreds of thousands. SNP chips can be cost effective and require low computational capacity. These technologies are reviewed elsewhere (LaFramboise 2009). Here, we focus on sequencing approaches, since we believe that in the future essentially all genotyping will be sequence-based. This is especially true for highly diverse material,

such as CWRs, since sequence based technologies have no inherent biases against highly divergent alleles such as is the case for SNP arrays. SNP chips require prior knowledge of markers and a 'discovery phase' that is usually sequence-based, before the chip can be designed and the population scored or 'detected'. A further advantage of high-throughput sequencing is that it allows simultaneous discovery and detection of markers.

High throughput sequencing is currently dominated by Illumina, Inc., a biotechnology company headquartered in San Diego, CA, whose machines currently yield > 10 megabases of sequence per dollar (www.genome.gov/sequencingcosts July 31 2013). A number of new 'third generation' sequencing platforms may become available in the near future. Amongst other things, such as longer reads of higher quality, these third generation approaches promise to sequence single molecules, which may greatly reduce technical artefacts from PCR amplification, as well as require fewer reagents and less starting material for sequencing (Quail et al. 2012; Schatz et al. 2012).

Sequencing is still sufficiently expensive that researchers must consider trade offs between the numbers of individuals sequenced, sequence depth and the number of sites assayed. The gold standard is whole genome shotgun sequencing (WGS), in which the genome is sheared randomly and all sites in the genome are sampled. However, WGS sequencing may be cost prohibitive for organisms with large genomes, even with modest sample sizes. Transcriptome sequencing and exome capture methods target genic regions for sequencing, which lowers sequencing costs, but libraries are expensive and labour intensive to prepare. Reduced representation methods, like genotyping by sequencing (GBS) (Elshire et al. 2011) and Restriction site Associated DNA sequencing RAD (Baird et al. 2008), sequence from restriction enzyme sites to sample thousands or millions of sites in a large number of individuals. GBS and RAD sequencing libraries are straightforward and inexpensive to prepare, and depth per sample can be adjusted by varying levels of multiplexing. The main downside

of the reduced representation approaches is obvious: many genes and other functionally important components of the genome will not be sampled. Also, the amount of missing data (sites where not every sample gets sequenced) can be high because of polymorphism in the presence/absence of restriction sites (especially in genetically diverse samples like CWRs), and because of the stochastic sampling process across sites and samples.

b Second level

19.3.2 Sequencing considerations

Most sequence-based population genomics research involves two components, reference genome assembly and re-sequencing of population samples. The sequencing and assembly of a reference genome for an organism is often conducted as a large collaborative effort. This reference genome is usually based on a single popular elite cultivar. When more individuals are sequenced for mapping populations, population genomic studies or CWR germplasm, the reads are aligned to the reference genome. In some cases, with highly diverse material or related taxa, *de novo* assemblies or reference guided *de-novo* assemblies may be carried out (Schneeberger et al. 2011).

Genome size is highly variable across plants (Ohri 1998) and dictates how much sequence data will be needed to adequately sample the genome. Large genomes are doubly troublesome because they are usually composed of large amounts of repetitive elements (Flavell et al. 1974). Bread wheat has one of the largest crop genomes at 17 Gigabases, 80% of which is made up of repeated sequences, largely retroelements (Brenchley et al. 2012). The total number of repeated elements may be less important than the age of those repeats. The sunflower genome contains a very high proportion of a single retro-element, which has undergone a very recent expansion, making it difficult to resolve the number and location of repeats (Staton et al. 2012). In addition to the amplification of repetitive

elements, genes can be duplicated by a variety of mechanisms (Flagel and Wendel 2009), creating paralogs that can also be difficult to differentiate. Polyploids, especially recent polyploids like *Brassica napus* or bread wheat, contain entire genomes in duplicate, which are difficult to differentiate. Deeper sequencing may not be enough to obtain high quality assemblies in the face of repetitive elements and duplication, and high density genetic maps, physical maps (International Barley Genome Sequencing Consortium 2012) and or a BAC-by-BAC sequencing approach (Schnable et al. 2009) may be required. Resolving duplicated genes and other repetitive elements in the genome should be a priority as copy number variation may underlie many QTL (Chia et al. 2012), and wild relatives are likely to contain a great deal of copy number diversity (Swanson-Wagner et al. 2010; Muñoz-Amatriaín et al. 2013).

With increasing diversity and complexity of a given genome, the amount of sequencing required to obtain high confidence markers also increases. Inbred bi-parental mapping populations, especially in cases where the actual parental genotypes are available, can be genotyped in a more straightforward manner than, for example, wild collections. In mapping populations derived from inbred parental lines there should only be two possible alleles at each locus, and clear expectations for allele frequencies and heterozygosity can be made. Low coverage WGS sequencing may be the best-suited approach for genotyping mapping populations; deep coverage is not required at any given site and the increased marker density can help resolve recombination locations (Renaut et al. 2013). For genetically or taxonomically diverse samples, however, it is difficult to predict allele frequencies and, because of the likely occurrence of many rare alleles and heterozygosity, more sequencing depth is required at each site to confidently identify bases. Reduced representation methods, RAD and GBS, may be well suited for diverse material and could be the best choice for genotyping wild collections (Hohenlohe et al. 2010; Lu et al. 2013).

As sequencing costs decrease, and higher levels of multiplexing become possible, WGS sequencing will continue to gain in popularity. For some studies, only 1x coverage or less may be required to genotype populations(Pasaniuc et al. 2012). A number of computational methods have been developed that help address problems arising from missing data in such low coverage sequencing. By sampling the entire genome it will be less likely that key rare variants will be missed. A recent example in maize highlights this point where a large effect size QTL was initially missed before whole genome sequence was available and a single SNP was found to be associated with the trait (Romay et al. 2013). Ideally, new technologies with improved quality and longer sequence length will reduce the need to sequence the same fragments repeatedly.

b Second level

19.3.3 Data analysis

The sheer amount of data from high-throughput sequencing runs can be overwhelming. Many projects are underway to sequence thousands of samples with WGS or reduced representation sequencing, which means terabytes of raw data need to be stored and processed. This requires considerable computational power, as well as computational expertise. An additional challenge is that much of the analytical software built for these types of analyses was not developed with the diversity of plant genomes in mind. Although these challenges exist, we believe they are surmountable, and as demand for plant genomics data continues to grow, more appropriate analytical approaches and software solutions will be developed that allow biologists with little training in bioinformatics to confidently handle this type of information.

Often, populations are sequenced using genetic barcoded adapters so that the samples can be pooled together for sequencing. The recovered reads are demultiplexed and aligned to a reference.

Compared to assemblies, which are quite computationally intensive, alignments for one or a few lanes of Illumina sequence can be performed on a standard desktop computer in a matter of hours or days using one of a number of programs, such as BWA (Li and Durbin 2009) or Bowtie (Langmead et al. 2009). After alignment, variants are identified; these will often be SNPs but could also include insertions and deletions, or information on copy number variation that can be quantified as read depth variants (Chia et al. 2012). At each of these stages (demultiplexing, alignment and variant identification) more information is generated and the amount of hard drive space required increases. Including additional hard-drive capacity to enable experimentation with different methodologies, as well as safety backups, can lead to substantial costs and often requires the purchase of dedicated storage machines.

With highly diverse sets of samples, such as many CWRs, there are several potential sources of analytical bias. Aligning to a common reference sequence, although attractive from an analysis perspective, may result in selection against highly diverse sites in the genome. Highly divergent reads may not align at all, ultimately resulting in an underestimate of the actual amount of diversity in the sample. This trend can even be observed within cultivated maize, where lines that are less closely related to the reference have more missing data (Romay et al. 2013). This type of bias may be avoided by conducting reference guided assemblies or completely *de-novo* assemblies. Different mapping parameters or mapping programs can yield different results (Zook et al. 2013). Methods for identifying variant sites without using a reference sequence, using k-mer counting for example(Nordström et al. 2013), are being developed. For reduced representation methods of sequencing, there are software solutions that allow variant discovery in the absence of a reference sequence (Catchen et al. 2011; Lu et al. 2013). Researchers often find with reduced representation methods that they have a great deal of missing data, sites where only a few of the individuals are

scored. To address this problem, methods to impute the missing data are available (Spindel et al. 2013; Rutkoski et al. 2013), but these may not be appropriate with diverse and unrelated material, such as CWRs.

a First level

19.4 Future perspective

b Second level

19.4.1 Making the most of it: Databasing and standards

An important consideration when generating genomic data is where and how the data will be stored. Not only should the data be put into long-term storage to facilitate its future use by other groups, but clear links need to be made between the genomic data and other types of information, including the physical germplasm samples and phenotypic data. Gene banks often have their own information systems, containing passport and characterization data for many accessions, yet in many cases there are few or no links to relevant genomic data. Likewise, databases dedicated to the storage of genomic data often store raw data that is of little help to many researchers who could benefit from more processed information.

To increase the usefulness of genomic data, it must be provided alongside important associated information. This should include the meta data that goes along with sequencing itself, such as: (1) how was the library prepared? (2) what kind of indexing was used? (3) what machine was used for sequencing? (4) how are the quality scores encoded? and (5) which accession and which seed lot did the individual plants come from? In addition, the sequence information should be connected to

biologically relevant data such as: (1) what breeding schemes has this material been used in? (2) was the sequenced individual used for any crosses? (3) was it selfed/purified? (4) for mapping populations, is the genomic data from the parental lines of a given population, from the actual parents or from their sibs or their offspring? (5) where were the plants collected? and (6) does any relevant phenotypic data exist?

For researchers to use data from other labs, they must trust it. This means open and transparent research is paramount and established standards for experimentation, analysis and record keeping have to be followed. Standards for genomics datasets may involve describing the settings of each parameter for all the software used, quality filtering, contamination removal and metrics to gauge assembly quality (Salzberg et al. 2012). For each type of analysis, different standards must be established, and reproducibility could be a useful metric to gauge how well those standards are being adhered to. For example, a recent meta-analysis showed that 30% of studies using the population genetic program, STRUCTURE, could not be reproduced, due to highly variable and unreported parameter selection. This result highlights the need for reporting of all parameters used in analyses, as well as both raw and filtered data sets (Vines et al. 2013). Making the filtered data available will not only facilitate attempts to reproduce the results, but it also can facilitate more rapid use of the data for other analyses. Establishing standards and best practices in these early days of gene bank sequencing will ensure the long-term usefulness of the data and greatly increase its future impact.

b Second level

19.4.2 Germplasm cataloguing

We envision a future where gene bank curators and plant breeders will have access to a

genomics parts catalogue to facilitate the use of CWRs and all other available germplasm. This catalogue would include far more information than an accession's genome sequences, although that would be its basis. It would include how much genetic diversity is in each accession, how accessions are related and how similar they are. It would quantify the diversity found in a given set of accessions or within a particular species, sub-species or ecotype. It would allow selections to be made based on the presence of desired alleles for particular QTLs. A breeder could also select improved material that already contains particular wild alleles of interest. Information on the collecting site, such as latitude, longitude as well as climate and soil parameters, would be integrated into the catalogue as well. With appropriate computational support, this would allow not just the selection of accessions, but of alleles, associated with climatic factors. For wild accessions, there would be information as to the material's history of introgression, local adaptation and selection. This catalogue would also include an accessible and extensive list of markers for marker-assisted selection.

Genomic information is attractive not just because it represents the true value of individual CWR accessions, but because it is possible to obtain it for every sample in a collection, including both existing and future collections. For many existing collections, gathering complete passport data, including information on phenotype, collection locality, soil type, exposure and ecology, will not be possible and it is likely that many future collection trips will not be able to capture all of this relevant information either. However, it is possible to collect genomic information for any material, no matter its origin. Thus, genomic data could become the most common 'currency' of germplasm curation and use. As the climate changes, crop improvement must become more efficient. For the genomics revolution to reach its full potential, we need to train scientists who are skilled in both biology and computational science, we need to allow data and software to be generated and transferred quickly and transparently, and we need to work on long-term solutions for the storage and curation of both

raw and processed data.

a First level

19.5 References

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3(10): e3376. doi:10.1371/journal.pone.0003376.

Bernardo, R. 2009. Genomewide Selection for Rapid Introgression of Exotic Germplasm in Maize. *Crop Science* 49:419-425.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L.A., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F.X., Edwards, K.J., Bevan, M.W., Hall, N. 2012.Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491:705–710.

Brown, A. 1989. Core collections: a practical approach to genetic resources management. *Genome* 31:818–824.

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3* 1:171–182.

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences USA* 110:8057–8062.

Cavanagh, C., Morell, M., Mackay, I., Powell, W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* 11:215–221.

Chen, H., Patterson, N., Reich, D. 2010. Population differentiation as a test for selective sweeps. *Genome Research* 20:393–402.

Chia, J., Song, C., Bradbury, P., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., Glaubitz, J.C., Gore, M., Guill, K.E., Holland, J., Hufford, M.B., Lai, J., Li, M., Liu, X., Lu, Y., McCombie, R., Nelson, R., Poland, J., Prasanna, B.M., Pyhajarvi, T., Rong, T., Sekhon, R.S., Sun, Q., Tenaillon, M.I., Tian, F., Wang, J., Xu, X., Zhang, Z., Kaeppler, S.M., Ross-Ibarra, J., McMullen, M.D., Buckler, E.S., Zhang, G., Xu, Y., Ware, D.2012. Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics* 44:803-807.

International Barley Genome Sequencing Consortium. 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491:711–716.

Cruz de Carvalho, M.H. 2008. Drought stress and reactive oxygen species: Production, scavenging and signaling. *Plant Signal Behavior* 3:156–165.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. 2011. A

Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6(5): e19379. doi:10.1371/journal.pone.0019379.

Excoffier, L., Foll, M., Petit, R.J. 2009. Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics* 40:481-501.

Feder, J.L., Nosil, P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64:1729–1747.

Feder, J.L., Egan, S.P., Nosil, P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* 28:342–350.

Flagel, L.E., Wendel, J.F. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist* 183:557–564.

Flavell, R.B., Bennett, M.D., Smith, J.B., Smith,D.B. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochemical Genetics* 12:257–269.

FAO2010. *Second report on the state of the world's plant genetic resources for food and agriculture.* Rome.

Gouesnard, B., Bataillon, T.M., Decoux, G., Rozale, C., Schoen, D.J., David, J.L. 2001. MSTRAT: An algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. *Journal of Heredity* 92:93–94.

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z., Min, J., Guo, X., Murat, F., Ham, B., Zhang, Z., Gao, S., Huang, M., Xu, Y., Zhong, S., Bombarely, A., Mueller, L A., Zhao, H., He, H., Zhang, Y., Zhang, Z., Huang, S., Tan, T., Pang, E., Lin, K., Hu, Q., Kuang, H., Ni, P., Wang, B., Liu, J., Kou, Q., Hou, W., Zou, X., Jiang, J., Gong, G., Klee, K., Schoof, H., Huang, Y., Hu, X., Dong, S., Liang, D., Wang, J., Wu, K., Xia, Y., Zhao, X., Zheng, Z., Xing, M., Liang, X., Huang, B., Lv, T., Wang, J., Yin, Y., Yi, H., Li, R., Wu, M., Levi, A., Zhang, X., Giovannoni, J.J., Wang, J., Li, Y., Fei, Z., Xu, Y. 2012. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics* 45:51–58.

Hamilton, J.P., Buell, R.C. 2012. Advances in plant genome sequencing. *The Plant Journal* 70:177–190.

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., Toomajian, C., Roux, F., Bergelson, F. 2011. Adaptation to Climate Across the *Arabidopsis thaliana* Genome. *Science* 334:83–86.

Hancock, A.M., Alkorta-Aranburu, G., Witonsky, D.B., Di Rienzo, A. 2010. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:2459–2468.

Hijmans, R.J., Garrett, K.A., Huaman, Z., Zhang, D.P., Schreuder, M., Bonierbale, M. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology* 14:1755–1765.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics* 6(2): e1000862. doi:10.1371/journal.pgen.1000862.

Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., Feng, L., Furuumi, H., Kubo, T.,

Miyabayashi, T., Yuan, X., Xu, Q., Dong, G., Zhan, Q., Li, C., Fujiyama, A., Toyoda, A., Lu, T., Feng, Q., Qian, Q., Li, J., Han, B. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497–501.

Hufford, M.B., Lubinksy, P., Pyhäjärvi, T., Devengenzo,M.T., Ellstrand,N.C., Ross-Ibarra, J. 2013. The Genomic Signature of Crop-Wild Introgression in Maize. *PLoS Genetics* 9:e1003477.doi:10.1371/journal.pgen.1003477.

Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhajarvi, T., Chia, J., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., Lai, J., Morrell, P.L., Shannon, L.M., Song, C., Springer, N.M., Swanson-Wagner, R.A., Tiffin, P., Wang, J., Zhang, G., Doebley, J., McMullen, M.D., Ware, D., Buckler, E.S., Yang, S., Ross-Ibarra, J.2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics* 44:808-811.

Knight, T.T. 1806. Observations on the Method of producing new and early fruit*Transactions of the Horticultural Society* 2:30-39.

Koenig, D., Jiménez-Gómez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeples, M., Brady, S.M., Pauly, M., Weigel, D., Usadel, B., Fernie, A.R., Peng, J., Sinha, N.R., Maloof J.N.2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato.*Proceedings of the National Academy of Sciences USA*doi:10.1073/pnas.1309606110.

LaFramboise, T. 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Research 37:4181–4193.

Lam, H., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F., Li, M., He, W., Qin, N., Wang, B., Li, J., Jian, M., Wang, J., Shao, G., Wang, J., Sun, S.S., Zhang, G.2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics* 42:1053–1059.

Langmead, B., Trapnell, C., Pop, C., Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25.

Leimu, R., Fischer, M. 2008. A Meta-Analysis of Local Adaptation in Plants. *PLoS ONE* 3:e4010. doi:10.1371/journal.pone.0004010.

Li, H., Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Lorenz, A., Hoegemeyer, T. 2013. The phylogenetic relationships of US maize germplasm. *Nature* 45:844–845.

Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., Buckler, E.S., Costich, D.E. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS Genetics 9:e1003215.doi:10.1371/journal.pgen.1003215.

Mallet, J. 2007. Hybrid speciation. *Nature* 446:279–283.

McKhann, H.I., Camilleri, C., Bérard, A., Bataillon, T., David, J.L., Reboud, X., Le Corre, V., Caloustian,C., Gut, I.G., Brunel, D. 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *The Plant Journal* 38:193–202.

Muñoz-Amatriaín, M., Eichten, S.R., Wicker, T., Richmond, T.A., Mascher, M., Steuernagel,B., Scholz,

U., Ariyadasa, R., Spannagl, M., Nussbaumer, T., Mayer, K.F., Taudien, S., Platzer, M., Jeddeloh, J.A., Springer, N.M., Muehlbauer, G.J., Stein, N. 2013. Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biology* 14:R58.

Nordstrom, K.J.V., Albani, M.C., James, G.V., Gutjahr, C., Hartwig, B., Turck, F., Paszkowski, U., Coupland, G., Schneeberger, K. 2013. Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nature Biotechnology* 31:325–330.

Ohri, D. 1998. Genome size variation and plant systematics. *Annals of Botany* 82:75–83.

Pasaniuc, B., Rohland, N., McLaren, P.J., Garimella, K., Zaitlen, N., Li, H., Gupta, N., Neale, B.M., Daly, M.J., Sklar, P., Sullivan, P.F., Bergen, S., Moran, J.L., Hultman, C.M., Lichtenstein, P., Magnusson, P., Purcell, S.M., Haas, D.W., Liang, L., Sunyaev, S., Patterson, N., de Bakker, P.I.W., Reich, D., Price, A.L. 2012. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies *Nature Genetics*44:631–635.

Petit, R.J., Excoffier, L. 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution* 24:386–393.

Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., Myers, S. (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics* 5: e1000519doi:10.1371/journal.pgen.1000519.

Pritchard J.K.,Stephens M, Donnelly P.J. 2000.Inference of population structure using multilocus genotype data *Genetics* 155:945–959.

Przeworski, M., Coop, G., Wall, J.D. 2005. The signature of positive selection on standing genetic variation. *Evolution* 59:2312–2323.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, T.R., Swerdlow, H.P., Gu, Y. 2012. A tale of three next generation sequencingplatforms: comparison of Ion Torrent, PacificBiosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:1–1.

Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L., 2010. A Gap Analysis Methodology for Collecting Crop Genepools: A Case Study with Phaseolus Beans. *PLoS ONE* 5:e13497.doi:10.1371/journal.pone.0013497.

Renaut, S., Grassa, C., Rieseberg, L.H. 2013. The number and size of genomic islands of divergence do not vary with geography of speciation.*Nature Communications* 4:1827 doi:10.1038/ncomms2833.

Rieseberg, L.H., Willis, J.H. 2007. Plant Speciation. *Science* 317:910–914.

Roesti, M., Hendry, A.P., Salzburger, W., Berner, D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* 21:2852–2862.

Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-Garcia, S.A., McMullen, M.D., Holland, J.B., Buckler, E.S., Gardner, C.A. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology* 14:R55.

Rutkoski, J.E., Poland, J., Jannink, J.L., Sorrells, M.E. 2013. Imputation of Unordered Markers and the Impact on Genomic Selection Accuracy. *G3*3:427–439.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.

Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M., Marcais, G., Pop, M., Yorke, J.A. 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research* 22:557–567.

Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., Koren, S., Phillippy, A.M. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology* 30:693–700.

Schnable, P.S., Ware., D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney L., Kruchowski S.S., Tomlinson C., Strong C., Delehaunty K., Fronick C., Courtney B., Rock S.M., Belter E., Du F., Kim K., Abbott R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., Wilson, R.K. 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326:1112–1115.

Schneeberger, K., Ossowski, S., Ott, F., Klein, J.D., Wang, X., Lanz, C., Smith, L.M., Cao, J., Fitz, J., Warthmann, N. 2011. Reference-guided assembly of four diverse Arabidopsis thaliana genomes. *Proceedings of the National Academy of Sciences USA* 108:10249–10254.

Schroeder, J.I., Delhaize, E., Frommer, W.B., Guerinot, M.L., Harrison, M.J., Herrera-Estrella, L., Horie, T., Kochian, L.V., Munns, R., Nishizawa, N.K., Tsay, Y.F., Sanders, D. 2013. Using membrane transporters to improve crops for sustainable food production. *Nature* 497:60–66.

Smale, M., Koo, B. 2003. Biotechnology and Genetic Resource Policies: What is a Genebank Worth? *IFPRI: IPGRI: SGRP.*

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N., McCouch, S. 2013. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theoretical and Applied Genetics* doi:10.1007/s00122-013-2166-x.

Staton, S.E., Bakken, B.H., Blackman, B.K., Chapman, M.A., Kane, N.C., Tang, S., Ungerer, M.C., Knapp,

S.J., Rieseberg, L.H., Burke, J.M. 2012. The sunflower (*Helianthus annuus*L.) genome reflects a recent history of biased accumulation of transposable elements. *The Plant Journal* 72:142–153.

Stölting, K.N., Nipper, R., Lindtke, D., Caseys, C., Waeber, S., Castiglione, S., Lexer, C. 2012. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* 22:842–855.

Swanson-Wagner, R.A., Eichten,S.R., Kumari, S., Tiffin, P., Stein, J.C., Ware, D., Springer, N.M. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Research* 20:1689–1699.

Teshima, K.M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Research* 16:702–712.

van Treuren, R., de Groot, E.C., Boukema, I.W., van de Wiel, C.C.M., van Hintum, T.J.L. 2010. Marker-assisted reduction of redundancy in a genebank collection of cultivated lettuce. *Plant Genetic Resources* 8:95–105.

Vines, T.H., Andrew, R.L., Bock, D.G., Franklin, M.T., Gilbert ,K.J., Kane, N.C., Moore, J-S., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T., Yeaman, S. 2013. Mandated data archiving greatly improves access to research data. *The FASEB Journal* 27:1304–1308.

Voight, B.F., Kudaravalli,S., Wen,X., Pritchard,J.K. 2006. A Map of Recent Positive Selection in the Human Genome. *PLoS Biology*4:e72.doi:10.1371/journal.pbio.0040072

Zook, J.M., Chapman, B., Wang, J., Mittelman, D., Hofmann, O., Hide, W., & Salit, M. 2013. Integrating sequencing datasets to form highly confident SNP and indel genotype calls for a whole human genome. *arXiv preprint arXiv:1307.4661*.

Mackay, M.C.,Street, K.A. 2004. Focused Identification of Germplasm Strategy – FIGS. Proceedings 11th Wheat Breeding Assembly, 20–24 September 2004, Canberra.

Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).(GSP) Available at: www.genome.gov/sequencingcosts. Accessed July 31 2013.

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Wednesday, July 02, 2014 5:45 PM
To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

OK, sounds good. Though I guess I could also stay parked at the Hilton, since it seems like an easy walk? Or do I need my car to head to the genebank after?

Cheers,
Hannes

On Wed, Jul 2, 2014 at 5:43 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

> Sounds great Hannes- I'll see you there at 7:00. BTW parking is around back (turn on Lake street and go down alley). We will catch up with Chris around 8:00 at the genebank. See you soon!
>
> Stephanie
>
> **From:** Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> **Sent:** Wednesday, July 02, 2014 3:38 PM
>
> **To:** Greene, Stephanie
> **Cc:** Richards, Chris
> **Subject:** Re: visit on July 3
>
> Hi Stephanie and Chris,
>
> Am about to leave Dulles for my flight to Denver tonight and just wanted to reconfirm that we'll meet tomorrow at 7am at the Wild Boar?
>
> Looking forward to it!
>
> Cheers,
>
> Hannes

On Mon, Jun 2, 2014 at 5:01 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

Hi Hannes:

Let's meet at the Wild Boar (1510 S. College Ave)- it's close to the Hilton. Will 7:00 am work? Chris , you are welcome to join us as well. Looking forward to talking about alfalfa CWR with you!

Stephanie

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Sunday, June 01, 2014 1:07 PM
To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

Hi Stephanie,

A breakfast meeting sounds great! Do you know a good spot somewhere in Ft. Collins? I have a rental car so I'd be happy to go wherever suits you!

Let me check with Matija on whether he wants me to discuss Grin Global with folks while I am up there. I'll get back to you about this asap. I also have a particular interest to discuss alfalfa wild relatives with you!

Thanks!

Hannes

On Fri, May 30, 2014 at 6:50 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

Hi Hannes:

Chris told me you would be in our neck of the woods on July 3rd. I would love to visit- I am flying out at 11:00 am from Denver, but could meet you first thing in the morning or maybe we could have breakfast together? Let me know what works. I will also have Pat Conine give you an in-depth tour of our facilities. Of course, Chris can give you the low down on the research being conducted. If you are interested in touching base with any GRIN folks let me know too. Josef Pohl is our IT specialist that developed our local apps to interface with GRIN global and has been developing a prototype taxonomy wizard for Grin global. He can provide you with insight regarding how the NPGS (and specifically NCGRP)is gearing up to adopt GRIN global. Let me know your time frame for the visit and we'll get things set up.

Stephanie

~~~~~~

Stephanie L. Greene, PhD

Seed Curator

USDA, ARS

National Center for Genetic Resources Preservation

1111 South Mason St.

Fort Collins, CO 80521

970-495-3227 (tel)

[redacted] (cell)

970-221-1427 (fax)

--

Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org
~~~~~~

Conserving Crop Diversity, Forever

--

Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

--

Hannes Dempewolf
Scientist
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 115
Mobile: [redacted]
www.croptrust.org

Conserving Crop Diversity, Forever

Hutchison, Stasia

From: Bretting, Peter
Sent: Tuesday, July 01, 2014 9:55 AM
To: Richards, Chris
Subject: RE: Visit by Hannes Dempewolf

Hi Chris—good to hear from you. Hannes will visit here tomorrow afternoon. The primary topic for discussion will be DivSeek.

I've not yet seen any landscaping study from the ITPGRFA Secretariat, and will ask Hannes regarding progress with that. Similarly, I've not yet seen any letter-of-intent yet—perhaps tomorrow? Of course we in ARS employ collaboration/partnership mechanisms such as non-funded and "regular" cooperative agreements—so if we cooperate formally with DivSeek hopefully some sort of an acceptable understanding can be devised.

Thanks for the encouraging news regarding the post-doc proposal co-authored with Stephanie. Colin mentioned visiting Ft. C. as part of a US sojourn that will also include the Crop Science Society of America meetings in Long Beach in early November.

All the best,

Peter

Peter Bretting
USDA/ARS Office of National Programs
Room 4-2212, Mailstop 5139
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Phone 1.301.504.5541
Fax 1.301.504.6191
Mobile Phone [redacted]
E-mail peter.bretting@ars.usda.gov
Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

From: Richards, Chris
Sent: Monday, June 30, 2014 6:58 PM
To: Bretting, Peter
Subject: Visit by Hannes Dempewolf

Hi Peter-

Hope you're doing great.

We're getting a visit here from Hannes next Thursday. Is he stopping by headquarters as well? No doubt he is interested in advancing DivSeek. Having worked on the apple GBS data sets with Sean Myles, I can offer up my experience in turning a bioinformatic exercise into data that might be put to use in managing a collection. How do we figure into their landscape/scoping analysis timeline? Do they have a letter of intent crafted?

On a different front, I'm working on a post-doc proposal with Stephanie Greene on methods to extend gap analysis to guide sampling of adaptive genetic lineages. We've got Colin Khoury coming out in October along with another postdoc at NIMBioS, which is a math-biological synthesis center in Tennessee. Glad she's working at the NCGRP. It's fun to work with her.

Cheers,
Chris

Christopher Richards, Ph.D.
Population Geneticist
United States Department of Agriculture
National Center for Genetic Resources Preservation
Colorado State University
1111 South Mason Street
Fort Collins, CO 80521
USA
Chris.richards@colostate.edu
970 495 3201
http://www.ars.usda.gov/pandp/people/people.htm?personid=42033

Hutchison, Stasia

From: Richards, Chris
Sent: Monday, June 30, 2014 6:58 PM
To: Bretting, Peter
Subject: Visit by Hannes Dempewolf

Hi Peter-

Hope you're doing great.

We're getting a visit here from Hannes next Thursday. Is he stopping by headquarters as well? No doubt he is interested in advancing DivSeek. Having worked on the apple GBS data sets with Sean Myles, I can offer up my experience in turning a bioinformatic exercise into data that might be put to use in managing a collection. How do we figure into their landscape/scoping analysis timeline? Do they have a letter of intent crafted?

On a different front, I'm working on a post-doc proposal with Stephanie Greene on methods to extend gap analysis to guide sampling of adaptive genetic lineages. We've got Colin Khoury coming out in October along with another postdoc at NIMBioS, which is a math-biological synthesis center in Tennessee. Glad she's working at the NCGRP. It's fun to work with her.

Cheers,
Chris

Christopher Richards, Ph.D.
Population Geneticist
United States Department of Agriculture
National Center for Genetic Resources Preservation
Colorado State University
1111 South Mason Street
Fort Collins, CO 80521
USA
Chris.richards@colostate.edu
970 495 3201
http://www.ars.usda.gov/pandp/people/people.htm?personid=42033

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Thursday, June 05, 2014 1:01 PM
To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

Hi Stephanie,

Yes 7am sounds great!

Our Genesys program manager, Matija, is curious to learn more and get in touch with Josef Pohl directly. Do you think you could send me his email address?

I would also be interested to learn a bit more about progress with Grin Global but would mainly want to discuss crop wild relatives and DivSeek while I am in Ft. Collins.

Thanks and I am very much looking forward to my visit!
Hannes

On Mon, Jun 2, 2014 at 11:01 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

> Hi Hannes:
>
> Let's meet at the Wild Boar (1510 S. College Ave)- it's close to the Hilton. Will 7:00 am work? Chris , you are welcome to join us as well. Looking forward to talking about alfalfa CWR with you!
>
> Stephanie
>
> ---
>
> **From:** Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
> **Sent:** Sunday, June 01, 2014 1:07 PM
> **To:** Greene, Stephanie
> **Cc:** Richards, Chris
> **Subject:** Re: visit on July 3
>
> Hi Stephanie,
>
> A breakfast meeting sounds great! Do you know a good spot somewhere in Ft. Collins? I have a rental car so I'd be happy to go wherever suits you!
>
> Let me check with Matija on whether he wants me to discuss Grin Global with folks while I am up there. I'll get back to you about this asap. I also have a particular interest to discuss alfalfa wild relatives with you!

duplicate email trailed deleted

Hutchison, Stasia

From: Richards, Chris
Sent: Wednesday, July 02, 2014 6:57 PM
To: Hannes Dempewolf
Subject: RE: visit on July 3

Sure you can walk...it really close.

See you tomorrow. Safe travels!!

C.

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Wednesday, July 02, 2014 3:45 PM
To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

OK, sounds good. Though I guess I could also stay parked at the Hilton, since it seems like an easy walk? Or do I need my car to head to the genebank after?

Cheers,
Hannes

On Wed, Jul 2, 2014 at 5:43 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

Sounds great Hannes- I'll see you there at 7:00. BTW parking is around back (turn on Lake street and go down alley). We will catch up with Chris around 8:00 at the genebank. See you soon!

Stephanie

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Wednesday, July 02, 2014 3:38 PM

To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

Hi Stephanie and Chris,

Am about to leave Dulles for my flight to Denver tonight and just wanted to reconfirm that we'll meet tomorrow at 7am at the Wild Boar?

duplicate email trail deleted

Hutchison, Stasia

From: Hannes Dempewolf <hannes.dempewolf@croptrust.org>
Sent: Monday, June 02, 2014 5:09 PM
To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

Sounds great!

Thanks,
Hannes

On Mon, Jun 2, 2014 at 11:01 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

Hi Hannes:

Let's meet at the Wild Boar (1510 S. College Ave)- it's close to the Hilton. Will 7:00 am work? Chris , you are welcome to join us as well. Looking forward to talking about alfalfa CWR with you!

Stephanie

From: Hannes Dempewolf [mailto:hannes.dempewolf@croptrust.org]
Sent: Sunday, June 01, 2014 1:07 PM
To: Greene, Stephanie
Cc: Richards, Chris
Subject: Re: visit on July 3

Hi Stephanie,

A breakfast meeting sounds great! Do you know a good spot somewhere in Ft. Collins? I have a rental car so I'd be happy to go wherever suits you!

Let me check with Matija on whether he wants me to discuss Grin Global with folks while I am up there. I'll get back to you about this asap. I also have a particular interest to discuss alfalfa wild relatives with you!

Thanks!

Hannes

On Fri, May 30, 2014 at 6:50 PM, Greene, Stephanie <Stephanie.Greene@ars.usda.gov> wrote:

duplicate email trail deleted

Hutchison, Stasia

From: Richards, Chris
Sent: Monday, June 30, 2014 6:58 PM
To: Bretting, Peter
Subject: Visit by Hannes Dempewolf

Hi Peter-

Hope you're doing great.

We're getting a visit here from Hannes next Thursday. Is he stopping by headquarters as well? No doubt he is interested in advancing DivSeek. Having worked on the apple GBS data sets with Sean Myles, I can offer up my experience in turning a bioinformatic exercise into data that might be put to use in managing a collection. How do we figure into their landscape/scoping analysis timeline? Do they have a letter of intent crafted?

On a different front, I'm working on a post-doc proposal with Stephanie Greene on methods to extend gap analysis to guide sampling of adaptive genetic lineages. We've got Colin Khoury coming out in October along with another postdoc at NIMBioS, which is a math-biological synthesis center in Tennessee. Glad she's working at the NCGRP. It's fun to work with her.

Cheers,
Chris

Christopher Richards, Ph.D.
Population Geneticist
United States Department of Agriculture
National Center for Genetic Resources Preservation
Colorado State University
1111 South Mason Street
Fort Collins, CO 80521
USA
Chris.richards@colostate.edu
970 495 3201
http://www.ars.usda.gov/pandp/people/people.htm?personid=42033