Introduction/State of Play

Current problems/challenges

More than a third of the world’s land is currently used for agriculture, and future expansion of this industry threatens the stability and survival of the wider ecosystem. To sustain the world’s expanding population, which is estimated to increase from 7 to 11 billion in the next 50 years, it is essential to generate more food from each hectare of arable land; however, this needs to be done in an environmentally sustainable manner.

Climate change poses a significant additional challenge, resulting in hotter and drier conditions in many parts of the world, as well as the increased likelihood of unpredictable weather extremes. Crop plants and their growing conditions will need to be more plastic; capable of coping with a variety of different environmental conditions. Maintaining and exploiting biodiversity (plants, insects, microorganisms, etc.) using well-organized management practices is essential to sustainably maintain and expand productivity.

Global solutions to stress resilient cropping systems are not exempt from the ‘data storm’ that is overwhelming all of biology. The development of new technologies, such as high-throughput sequencing and phenotyping, has led to an increasing number of large-scale and complex data sets. Integration and effective mining of these data sets is difficult. In particular, field and greenhouse experiments dedicated to plant responses to environmental conditions are by definition not reproducible, because the combination of environmental conditions in one experiment will never be experienced again by a particular set of genotypes. Data sets are generated on different technical platforms and the number of environmental variables even between growth chamber experiments is significant. Reuse of these data sets (for example, reanalysis based on different biological hypotheses and/or new methods) provides countless opportunities for new biological insights and achievements, but this requires data to be easily discoverable, as well as clearly annotated and explained, and in a format to easily enable use and integration with other data sets or analysis methods.

The ability to share data globally, and integrate and utilize such data in a variety of analysis pipelines and
approaches, overcoming nonbiological variation and varying amounts of noise, will greatly influence the efficiency and impact of biological research around the world, and play a key role in attempts to feed the world in the face of changing environments by linking crop genotype with the agro-ecosystem to predict phenotype.

**Agro-ecosystems**

Understanding the wide diversity of ecosystems available for crop production is a key challenge. Manipulation and exploitation of agro-ecosystems should help to expand crop productivity while minimizing the resources used; for example, to maximize the use of available water it should be applied when the plant will make maximum use of it (Tardieu 2012), which requires an understanding of water deficit scenarios for different crops. Alternate wetting and drying, which results in a mild water deficit, has been shown to save water and maintain yields in rice (Yao et al. 2012; Price et al. 2013), and other crops such as maize and cotton (Kang et al. 1998; Tang et al. 2010).

In the Shiyang river project in China, an intensive effort to limit nitrogen and reduce water use using methods such as drip irrigation has resulted in sustained or improved yields with reduced environmental impact (Du et al. 2015).

A genotype that has been selected for favorable performance in one environment does not necessarily perform equally well in a different environment, and the introduced alleles may potentially have a negative effect. It is important to understand the mechanisms by which plants respond to environmental stress; for example, different varieties may be susceptible to water deficit at different times of development (Tardieu and Tuberosa 2010; Tardieu 2012). A better understanding of ecosystem scenarios mapped with suitable genotypes may allow more appropriate varieties to be selected. The timing of planting (e.g., growing autumn instead of spring wheat) and soil manipulation even when not growing crops, can have a significant effect on water use and availability. Removal of summer weed cover in Australia resulted in retained water levels in the soil and significant subsequent increase in crop yield (https://grdc.com.au/Media-Centre/Ground-Cover/Ground-Cover-Issue-106-Sept-Oct-2013/Productivity-gains-there-for-the-taking?). Other agronomic approaches, such as intercropping, are also advantageous in certain conditions (Stoltz and Nadeau 2014), while problems with disease caused by crop monoculture could be addressed by mixing several genotypes of the crop in the same field (Zhu et al. 2000).

**Crop resources**

The genetic diversity of traditional varieties, modern cultivars, and wild relatives is crucial for crop improvement and food production, and also to act as a buffer for adaptation and resilience in the face of climate change. In recent years, however, there has been a strong tendency for farmers worldwide to abandon their multiple local varieties and landraces for genetically uniform, high-yielding varieties (http://www.fao.org/Int/cgrfa/ctthemes/plants/en/; van den Wouw et al. 2010). This means that, currently, approximately 75% of the genetic diversity of crops may have been lost. Breeding for yield under good conditions means that stress resilience genes are not necessarily selected. An example of this problem was clearly shown in Bengal, where modern high-yielding varieties of rice were no match for the traditional varieties following the instant salinization of soil caused by a hurricane in 2009 (https://www.independentsciencenews.org/un-sustainable-farming/valuing-folk-crop-varieties/). For sustainable production under variable conditions, breeding for resilience of yield rather than maximal yield under optimal conditions is needed, which may require the reintroduction of such lost alleles.

Humans use only around 150–300 of the approximately 80,000 known edible plant species. Three of these – rice, maize, and wheat – contribute nearly 60% of our intake of plant-based calories and protein. The exploitation of additional crop species that are more resilient to certain growth environments could be key to expanding the productivity of the most challenging agronomic areas, both in terms of yield but also nutrient quality. Research into orphan crops such as teff, millet, cassava, sweet potato, and bambara groundnut are crucial for strengthening regional agriculture and improving nutrition (Crops for the Future http://www.cropsforthefuture.org).

**Genetic improvement**

Many research projects around the world have generated genetic markers and mapping populations for phenotyping and quantitative trait loci (QTL) identification in the major crops such as rice, maize and wheat. For example, CIMMYT’s extensive international wheat improvement program is developing genomics for precision breeding in wheat using techniques such as high-throughput phenotyping, the collation of genetic resources (half a million wheat lines are available worldwide), and interspecific hybridization. QTL have been identified for heat and drought tolerance in several crops, and commercial drought-tolerant maize varieties have been developed, using marker-assisted breeding (e.g., Artisan hybrid maize from Syngenta and AquaMax from Pioneer (Tollefson 2010)).

There is an increasing gap between science and breeding. Research cannot deliver directly to the farmer without including breeding companies; therefore, large,
collaborative projects are essential. For example, the DROPS project (http://www6.inra.fr/dropsproject, http://cordis.europa.eu/project/rcn/95052_en.html) is a multi-scale, multi-environment project on drought tolerance that exploits natural diversity, phenotyping platforms, and field analysis. It involves collaborators from 11 countries, including 11 public organizations and five large seed companies, which are involved in development of the method, as well as the results. It is strongly interdisciplinary, involving modelers and statisticians, as well as plant breeders, molecular geneticists, and biochemists. It aims to combine precise crop modeling and genome prediction with environmental influences (Tardieu 2012; Tardieu and Tuberosa 2010).

**Gene identification**

Many research projects across the globe aim to genetically improve stress resilience in plants, using both the model plant Arabidopsis and several crop species. Many of these projects are at the fundamental research level. Knowledge of individual genes that can be manipulated to confer tolerance to a single stress is relatively advanced; for example, multiple genes conferring drought tolerance (reviewed for rice in Todaka et al. (2015)) and salt tolerance (e.g. Nax2 in wheat (Munns et al. 2012) and SALT3 in soybean (Guan et al. 2014)) have been identified, and the SUB1 gene was found to control submergence tolerance in rice (Xu et al. 2006)). Also, the ALT1 gene, which confers aluminum tolerance in sorghum, could be highly beneficial to use in Africa (Ryan et al. 2011). So far, few individual genes have been used to develop transgenic, stress-tolerant commercial lines, though DroughtGuard maize from Monsanto (which contains the cold shock protein cspB from the bacterium Bacillus subtilis) is one example.

**Sharing data sets**

High quality data relating to environmental stress responses in plants is extremely expensive to generate, and is obtained from specific species under specific conditions. The complexity and cost of generating good data tend to make organizations and scientists protective of those data, limiting their impact and value by restricting their availability. There is an abundance of less complete data sets, many of which would make valuable contributions to building up more robust collections and conclusions. No-one has the resources to test all genotypes or agronomic techniques in multiple situations, and reusing data provides a means to extend what is possible. Globally, we want to know: in what environments does a particular genotype succeed?

Cultural changes in the availability of data, and the ways in which they are queried, are required if effective outcomes are to be delivered. It is important that there is a shift toward quantitative data, and that data can be easily shared in a reliable and informed manner. The continual growth in data requires the linking of multiple data stores around the world, as well as the development of appropriate best-practice guidelines to adequately describe data so that computational methods can be used to query and/or discover data in different databases/structures. Several initiatives are working toward enhancing description of data and hence its accessibility including NCBI (Barrett et al. 2012), with the Bio Project database and submission portal, and the development of recommendations for metadata in plant phenotyping experiments (Krajewski et al. 2015).

**Utilizing negative results**

Furthermore, most current data release is via publication in scientific journals; this leads to an emphasis on positive results. It is likely that such positive results are less common than negative results, but negative results are never made available. However, we consider as results per se the lack of response of a given set of genotypes in a given environmental scenario and/or with the chosen physiological traits with the chosen protocol. Although some publishers are attempting to establish journals for the publication of negative results, it is debatable whether researchers will spend the time writing up a negative results publication; despite the buzz surrounding Elsevier’s *New Negatives in Plant Science*, the journal was discontinued in September 2016, just 1 year after its first issue. Simple and fast deposition in a database may be more likely to succeed. Open data stores are increasingly popular and could be used in this way, along with data-only journals in which high quality, curated data sets can be described without the requirement of scientific interpretation (Leonelli et al. 2013). The availability of negative results could be a game changer in international plant breeding, as it would stop continuous retesting of genotypes that do not work, and save time and money on a global scale, so speeding the development of stress tolerant local varieties.

**Training the next generation of data scientists**

To fully exploit new technologies for generating genome and phenotypic data, the plant research/breeding community needs to attract (and/or train) a new generation of researchers who are skilled at extracting relevant knowledge from diverse large-scale datasets, and can
combining diverse data sets to provide new insights. This is not just people with bioinformatics skills – most bioinformatics-trained scientists have used a limited number of tools, approaches and data sets. Bioinformaticians usually have skills in looking at genome data, creating novel ways of displaying data, and using computational tools, but not in developing new innovative methods for merging and analyzing data. We need people or teams with an amalgamation of computer science, statistical and biological knowledge to be able to explore and mine these datasets effectively to solve global food production challenges caused by population growth and environmental induced stress (see e.g., DROPS Project. http://www6.inra.fr/dropsproject, http://cordis.europa.eu/project/rcn/95052_en.html). Above all, we need a generation of researchers interested in quantitative analyses based on practical solutions that define when and where specific genotypes will thrive and not only interested in nice stories to satisfy reporting requirements for funding agencies. There is a real need to bring innovative quantitative scientists into the sustainable crop production arena; often too much money is spent on technology, and not enough on creating the innovative scientists who will use these data for the benefit of mankind.

Effective and innovative use of plant data will require dialog between different groups of expertise, and a focusing of effort on key problems and challenges. Such dialog initiatives exist within Big Data (for example, the Alan Turing Institute in the UK or, in the plant science community, projects such as EU DROPS (Millet et al. 2016), CyVerse in the USA (http://www.cyverse.org) and the Agrimetrics Big Data Centre of Excellence in the United Kingdom (http://www.agrimetrics.co.uk/), which brings together capability in data science and smart analytics with agrifood research expertise to drive exploitation of agrifood data thought not necessarily focused on crop science. Additional crop-focused initiatives include the French programmes Amaizing and Breedwheat (http://www.amaizing.fr; http://www.breedwheat.fr).

The ultimate goal of this field would be the ability to design plant varieties/genotypes that would thrive in certain environmental and agronomic conditions. A plant breeder should be able to define the environment in which they wish a plant to grow, and be able to mine the genetic variants that will generate such a crop plant with the optimum characteristics to produce a stable yield (Hammer et al. 2006). The time is ripe to move away from old ways of working, to combine accurately described and shared genetic resources with new data structures together with a generation of scientists to interact with them, which will enable plant breeding in the 21st century.

**Actions**

**Enhance information gathering, access and reuse**

There are multiple individual projects worldwide that clearly contribute to our knowledge of mechanisms to improve crop stress resilience at all scales, from agro-ecosystems to individual genes. Working with other relevant organizations, the Global Plant Council (http://globalplantcouncil.org/) which is described in an Editorial article in this volume, could help to make the outputs of these projects more easily accessible to a global audience. This short-term goal is not straight-forward, especially on a global scale, but initiatives do exist. For example, UK’s Collaborative Open Plant Omics (COPO) project is finding ways to make it easier for researchers to annotate and deposit data (http://copo-project.org), POPCorn in the US is an online resource providing access to distributed and diverse maize data (Cannon et al. 2011, http://dx.doi.org/10.1155/2011/923035) and the Genomes to Field (G2F) initiative that has released its data through CyVerse (http://www.cyverse.org/news/genomes-environment-data-set-now-publicly-accessible; http://www.genomes2fields.org).

Actions that GPC might take include:

1. Current status: Landscaping the different initiatives in this area and facilitating communication between them to develop bridges, enhance outputs and thus enable more effective global data reuse, while respecting and valuing different approaches and their comparative advantages, and showcasing successes to inspire and inform future activities.

2. Environmental conditions: Identifying minimum requirements and standards for the annotation of environments associated with a given experiment. These are not just “metadata” and are intrinsically part of phenotypic datasets. “Minimum datasets” have been defined in projects such as EU DROPS (http://www6.inra.fr/dropsproject, http://cordis.europa.eu/project/rcn/95052_en.html) or EPPN (Tardieu 2013)

3. Collating methods and protocols to be included in information systems, including the methods that have been used for image analyses or sensor calibration

4. Genotypes: what is being used, and what is publically available for testing and breeding?

5. Information on modeling: surveying the methods that are being used, and how these can be made widely accessible.

6. Data on orphan crops and their associated growth characteristics: gather available data to guide recommendations on which crops should be funding priorities.
7. Germplasm resources: increase knowledge of what has been developed and help to enable access. A directory of germplasm resources could be created, listing the resources that each facility has (including contact details), and this could be promoted and distributed via GPC channels and members.

8. Case studies from different communities: facilitating discussion between communities, with the aim of developing a simple, easy-to-use set of criteria that will be more likely to be adopted by experimentalists, while at the same time enabling data scientists to analyze the likelihood of specific outcomes and combine the maximum amount of available data.

9. Data integration and reuse projects: helping to define what is likely to have significant impact, including publicizing success stories in this area and disseminating resources about game-changing projects to prevent duplication, and trigger new ideas, and initiatives in gap areas.

10. Facilitate a cultural change in the way that plant stress resilience research is structured; switching from essentially qualitative to quantitative approaches, thus moving away from the traditional emphasis on telling a nice qualitative “story” that can be published in a high impact factor journal.

11. Training courses: landscaping existing courses, mirroring e-courses, and facilitating local workshops in different languages to help train and build capacity in scientific experts applying quantitative approaches to global datasets.

**Building on available information**

As an organization with links to plant scientists and professional societies from six continents, the GPC is well placed to undertake landscaping surveys and information-gathering activities that help to build on and link up existing silos of knowledge. For example, the GPC might coordinate efforts to:

1. Identify and prioritize the most devastating stresses, and combinations of stress factors, to assess the biggest challenges. Taking drought as an example, what are the most common dynamics of drought (e.g., the time at which plants experience drought, extent of drought)?

2. Collate knowledge on the diverse agro-ecosystems and map their similarities worldwide (Tardieu 2013). What is the limiting factor at each location? How conserved are these factors between different available data sets?

3. Identify suitable cropping systems for the different environments identified. Match cropping systems to environments using local knowledge, identify geographically distinct but agro-ecologically similar areas and propose cropping systems for these additional areas globally. The identification of common environments and stresses around the world will provide opportunities for sharing germplasm.

4. Enable the sharing of germplasm. It would be a very valuable achievement if GPC could improve this major bottleneck, facilitating the testing of germplasm bred for a particular environment in one country in similar environments across the globe.

5. Utilize the modeling performed by the Intergovernmental Panel on Climate Change to predict the most likely stress scenarios for different regions in the future, and where will be most impacted in terms of temperature, water, etc.

6. Integrate international research to generate a set of optimal practices in cropping systems.

**Provision of characterized genetic resources**

A key activity the GPC could undertake is to enable well coordinated comparisons of selected genotypes, tested over a long time scale, and under different combinations of stresses in geographically diverse natural environments. For example, the GPC might coordinate efforts as follows:

1. Pull together a small panel of genetically diverse lines for each target crop and make these available to researchers worldwide to allow phenotyping of known genetic materials under different agronomic environments. Data comparisons will allow germplasm information to be linked to stress resilience (individual or combinations of stress) and other traits (See Borrell and Reynolds 2017, this volume).

2. Model the outcomes and use these data to predict yield potential under variable or known adverse conditions, as well as determine the cost of plasticity and resilience.

3. Define links between genotypes/phenotypes and agricultural practices; help to promote the cultural change that is required to facilitate greater exchange and sharing of expertise, resources and data.

**Improve data availability**

Data are only of any value if they are available and contain sufficient annotation to ensure their utility. The current transition from a traditional, secretive approach to data to working on globally available resources can be influenced by research funders and scientific journals. The community must approach them to demand adherence to minimum data standards, to widely distributed data information systems, and to the...
availability of negative, as well as positive data. Together with European and International initiatives (www.fao.org/it/gaez/en; http://www.plant-phenotyping.org; http://www.plant-phenotyping-network.eu/eppn/home; http://emphasis.plant-phenotyping.eu), ways in which the GPC could help in this regard are highlighting on a global scale the critical issues of:

1. Building, using and interacting with IT infrastructures;
2. Forming appropriate biological questions using this infrastructure;
3. The ability to extract meaningful results and predictions from data; by collecting international experience in what has and has not worked.
4. Identifying where innovation is needed, whether it is already available in other fields (e.g. data-search, finance), and what resources are needed to transfer the technologies, approaches and mindset.

**Securing long-term global funding**

These issues of scientific culture, data availability, and the ability to ask the relevant questions are a major challenge that, although very difficult, need to be addressed urgently. The scale of the issue is such that, although small steps can be made immediately, to make the phase shift that would impact global food security may require a globally funded initiative. GPC could structure such a program aimed at crop stress resilience and propose it to major independent charitable funders.

**Acknowledgements**

This paper is based on outcomes from a Stress Resilience Symposium held in Brazil in October 2015 organized by the Global Plant Council and Society for Experimental Biology. The authors would like to thank the Society for Experimental Biology for funding support for this symposium.

**Conflict of Interest**

None declared.

**References**


