Title
Data access for the 1,000 Plants (1KP) project.

Permalink
https://escholarship.org/uc/item/1553083t

Authors
Matasci, Naim
Hung, Ling-Hong
Yan, Zhixiang
et al.

Publication Date
2014

DOI
10.1186/2047-217X-3-17

Peer reviewed
Data access for the 1,000 Plants (1KP) project

Naim Matasci1,2, Ling-Hong Hung3, Zhixiang Yan4, Eric J Carpenter5, Norman J Wickett6,7, Siavash Mirarab8, Nam Nguyen8, Tandy Warnow8, Saravananay Ay Ayyampalayam9, Michael Barker2, J Gordon Burleigh10, Matthew A Gitzendanner10, Eric Wafula11, Joshua P Der11, Claude W dePamphilis11, Béatrice Roure12, Hervé Philippe12,13, Brad R Ruhe10,14, Nicholas W Miles15, Sean W Graham16, Sarah Mathews17, Barbara Surek18, Michael Melkonian18, Douglas E Soltis19,15,16, Pamela S Soltis19,18, Carl Rothfels10,20, Lisa Pokorny21, Jonathan A Shaw20, Lisa DeGironimo23, Dennis W Stevenson23, Juan Carlos Villarreal24, Tao Chen25, Toni M Kutchan26, Megan Rolf26, Regina S Baucom27, Michael K Deyholos5, Ram Samudrala3, Zhijian Tian4, Xiaolei Wu4, Xiao Sun4, Yong Zhang4, Jun Wang4, Jim Leebens-Mack9* and Gane Ka-Shu Wong4,5,28*

Abstract
The 1,000 plants (1KP) project is an international multi-disciplinary consortium that has generated transcriptome data from over 1,000 plant species, with exemplars for all of the major lineages across the Viridiplantae (green plants) clade. Here, we describe how to access the data used in a phylogenomics analysis of the first 85 species, and how to visualize our gene and species trees. Users can develop computational pipelines to analyse these data, in conjunction with data of their own that they can upload. Computationally estimated protein-protein interactions and biochemical pathways can be visualized at another site. Finally, we comment on our future plans and how they fit within this scalable system for the dissemination, visualization, and analysis of large multi-species data sets.

Keywords: Viridiplantae, Biodiversity, Transcriptomes, Phylogenomics, Interactions, Pathways

Introduction
The 1,000 plants (1KP) project is an international multi-disciplinary consortium that has now generated transcriptome data from over 1,000 plant species. One of the goals of our species selection process was to provide exemplars for all of the major lineages across the Viridiplantae (green plants), representing approximately one billion years of evolution, including flowering plants, conifers, ferns, mosses and streptophyte green algae. Whereas genomics has long strived for completeness within species (e.g., every gene in the species), we were focused on completeness across an evolutionary clade – obviously not every species, but one representative species for everything at some phylogenetic level (e.g., one species per family, and perhaps more than one species when the family is especially large). Because many of our species had never been subjected to large-scale sequencing, 2 gigabases (Gb) of data per sample was sufficient to increase the number of plant genes by approximately 100-fold in comparison to the totality of the public databases.

The 1KP project began as a public-private partnership, with 75% of the funding provided by the Government of Alberta and 25% by Musea Ventures. Significant in-kind contributions were provided by BGI-Shenzhen in the form of reduced sequencing costs and by the NSF-funded iPlant collaborative [1] in the form of computational informatics support. Many plant scientists from around the world were involved in the collection of live tissue samples and in the extraction of RNA. Additional computing resources were provided by Compute Canada and by the China National GeneBank. Despite the constraints of this funding model, we released our data (on a collaborative basis) to scientists who approached us with goals that did not compete with ours. For the general community, access was provided through a BLAST portal [2].

We believed that there would be intrinsic value in data of this nature that is beyond our imagination. But for the initial publication, we agreed on two objectives. Firstly, by adopting a phylogenomics approach we hoped to resolve many of the lingering uncertainties in species
relationships, especially in the early lineages of streptophyte green algae and land plants, where previous analyses were based on comparatively sparse taxonomic densities. And secondly, despite the limitations of these data, we hoped to identify some of the gene changes associated with the major innovations in Viridiplantae evolution, such as multicellularity, transitions from marine to freshwater or terrestrial environments, maternal retention of zygotes and embryos, complex life history involving haploid and diploid phases, vascular systems, seeds and flowers. Our RNA extraction protocols [3] and our RNA-Seq transcriptome assembly algorithms [4] have already been published. Here, we are publishing the second of two linked papers. The first is a review of the state-of-knowledge for Viridiplantae species relationships and our initial foray into the phylogenomics on a subset of 1KP [5]. The other is a description of the websites that we created in order to provide access to the data (from raw reads to computed results), visualize the results, and perform custom analyses in conjunction with external data that the users can upload. An initial gene annotation is also provided, which focuses on the functional relationships between proteins and their associated metabolites.

Review
Access to raw and processed data
Our initial phylogenomics effort used sequences from multiple sources. They include transcriptomes from 1KP representing 85 species, transcriptomes from other sources representing 7 species, and genomes representing an additional 11 species. A summary of these data sources is given in Table 1. We submitted all of the unassembled reads from the 1KP transcriptomes to the Short Reads Archive (SRA) under project accession PRJEB4921 “1000 Plant (1KP) Transcriptome: The Pilot Study.” Note that, with the exception of Eschscholzia californica, we sequenced only one sample per species.

To make it easier for others to reproduce our phylogenomics analyses, we are releasing our intermediate computations, not just the final results. Everything is hosted at the iPlant Data Store, a high performance, large capacity, distributed storage system. The contents include transcriptome assemblies, putative coding sequences, orthogroups (i.e., from the 11 reference genomes), as well as gene and species trees with related sequence alignments. There are quite a lot of files and their total sizes are not negligible; so before users begin to download these files, we suggest that they consult Table 2 for a description of what to expect.

At the simplest level, anonymous downloads are permitted from a designated area of the iPlant Data Store [6]. However, much greater functionality is available through the iPlant resources that we describe in the following sections.

Visualization and custom analyses
To take full advantage of the iPlant computational infrastructure, it is necessary to first register at [7]. Accounts are free, and in addition to 1KP data, users will find high performance computing and cloud-based services. Multiple access modalities are supported: anonymous and secure web interfaces, desktop clients and high-speed command lines. However, we feel that for most users the best option is the iPlant discovery environment (DE), a web-based interface that provides users with high-performance computing resources and data storage. Most contemporary web browsers are supported, including Safari v. 6.1, Firefox v. 24, and Chrome v. 34. The caveat is that some of these functionalities (see below) require Java 1.6.

To guide users through its resources, iPlant is constantly producing new tutorials and teaching materials, including live and recorded webinars. The full catalog can be found at [8]. Here, we describe the new resources specifically created for 1KP.

Discovery environment (DE)
For access to the 1KP files, users should visit [9] and search for a folder called Community Data/onekp_pilot Figure 1.

From the data window it is possible to download individual files or perform bulk downloads of multiple files and directories through a Java plugin. Note that for security reasons, some operating systems will not allow users to run Java applets. In this instance, a window will pop up to tell the user that there is a problem, and the user should follow the instructions that are given to configure an iDrop desktop [10] Figure 2.

It is possible to perform analyses directly in the DE using any of the 1KP files as input; for example, users can re-compute the sequence alignments and gene trees using different algorithms and parameters [11] Figure 3. More generally, users can select from a variety of applications in the Apps catalogue, which is constantly growing, and includes many popular bioinformatics tools for large-scale phylogenetics, genome-wide associations and next generation sequence analyses.

Species and gene trees can be explored with the iPlant tree viewer, Phylozoom, a newly developed web-based phylogenetic tree viewer that supports trees with hundreds of thousand leaves and allows for semantic zooming Figure 4. To access the tree viewer, users need only click on a tree file. This will open a preview window with two tabs: one for the tree’s newick string (a format for graph-theoretical trees as defined at [12]) and another for the web link that opens a window to the tree display. Notice that pop-ups must be enabled on the user’s browser.

To zoom in and expand the collapsed clades, click on the node of interest. To zoom out, click and drag the tree
| Species               | Type               | Accession                  | iPlant ID |
|----------------------|--------------------|----------------------------|-----------|
| Arabidopsis thaliana | genome             | n/a                        | n/a       |
| Brachypodium distachyon | genome            | n/a                        | n/a       |
| Carica papaya        | genome             | n/a                        | n/a       |
| Medicago truncatula  | genome             | n/a                        | n/a       |
| Oryza sativa         | genome             | n/a                        | n/a       |
| Physcomitrella patens | genome           | n/a                        | n/a       |
| Populus trichocarpa  | genome             | n/a                        | n/a       |
| Selaginella moellendorffii | genome | n/a                        | n/a       |
| Sorghum bicolor      | genome             | n/a                        | n/a       |
| Vitis vinifera       | genome             | n/a                        | n/a       |
| Zea mays             | genome             | n/a                        | n/a       |
| Aquilegia formosa    | meta-assembly      | PlantGDB                   | AQUI      |
| Cymatoglossum rumphii | meta-assembly    | SRX022306, SRX022215        | CYCA      |
| Liriophyllum tulipifera | meta-assembly | PRJNA46857                 | LIRI      |
| Persea americana     | meta-assembly      | PRJNA46857                 | PERS      |
| Pinus taeda          | meta-assembly      | PRJNA79733                 | PINU      |
| Pteridium aquilinum  | meta-assembly      | PRJNA48473                 | PTER      |
| Zamia vazquezii      | meta-assembly      | PRJNA46857                 | ZAMI      |
| Acorus americanus    | OneKP meta-assembly | ERR364395, PRJNA46857       | ACOR      |
| Amborella trichopoda | OneKP meta-assembly | ERR364329, PRJNA46857       | AMBO      |
| Catharanthus roseus  | OneKP meta-assembly | ERR364390, PRJNA79951, PRJNA236160 | CATH      |
| Eschscholzia californica | OneKP meta-assembly | ERR364338, ERR364335, ERR364336, ERR364337, ERR364334, SRX002988, SRX002987, PlantGDB | ESCH      |
| Ginkgo biloba        | OneKP meta-assembly | ERR364401, PlantGDB         | GINK      |
| Nuphar advena        | OneKP meta-assembly | ERR364330, PRJNA46857       | NUPH      |
| Ophioglossum petiolatum | OneKP meta-assembly | ERR364410, SRX666586        | OPHI      |
| Saruma henryi        | OneKP meta-assembly | ERR364383, PRJNA46857       | SARU      |
| Welwitschia mirabilis | OneKP meta-assembly | ERR364404, PRJNA46857       | WELW      |
| Allamanda cathartica | OneKP              | ERR364399                  | MGVU      |
| Angiopteris exilis   | OneKP              | ERR364409                  | NHCM      |
| Anomodon attenuatus  | OneKP              | ERR364349                  | QMWB      |
| Bazzania trilobata   | OneKP              | ERR364415                  | WZYK      |
| Bohermania nivea     | OneKP              | ERR364387                  | ACFP      |
| Bryum argenteum      | OneKP              | ERR364348                  | JMWX      |
| Cedrus libani        | OneKP              | ERR364342                  | GGEA      |
| Ceratodon purpureus  | OneKP              | ERR364350                  | FFPD      |
| Chaetosphaeridium globosum | OneKP | ERR364369                  | DRGY      |
| Chara vulgaris       | OneKP              | ERR364366                  | CHAR      |
| Chloroerybus atmosphaericus | OneKP | ERR364371                  | AZZW      |
| Colchicum autumnale  | OneKP              | ERR364397                  | NHIX      |
| Coleochaete irregularis | OneKP          | ERR364367                  | QPDY      |
| Coleochaete scutata  | OneKP              | ERR364368                  | VQBJ      |
| Cosmarium ochthodes  | OneKP              | ERR364376                  | STKI      |
| Cunninghamia lanceolata | OneKP            | ERR364340                  | OUOI      |
| Species                                      | Project | Accession | Location |
|----------------------------------------------|---------|-----------|----------|
| Cyathea (Alsophila) spinulosa                | OneKP   | ERR364412 | GANB     |
| Cycas micholitzi                            | OneKP   | ERR364405 | XZUY     |
| Cylindrocystis brebissanii                  | OneKP   | ERR364378 | YOHI     |
| Cylindrocystis cushleckae                    | OneKP   | ERR364373 | JOJQ     |
| Dendrolycopodium obscursum                  | OneKP   | ERR364346 | XNFX     |
| Dioscorea villosa                           | OneKP   | ERR364396 | OCWZ     |
| Diospyros malabarica                        | OneKP   | ERR364339 | KVFU     |
| Entransia fimbriata                         | OneKP   | ERR364372 | BFIK     |
| Ephedra sinica                              | OneKP   | ERR364402 | VDAO     |
| Equisetum diffusum                          | OneKP   | ERR364408 | CAPN     |
| Gnetum montanum                             | OneKP   | ERR364403 | GTHK     |
| Hedwigia ciliata                            | OneKP   | ERR364352 | YWNY     |
| Hibiscus cannabinus                         | OneKP   | ERR364388 | OLFX     |
| Houttuynia cordata                          | OneKP   | ERR364332 | CSSK     |
| Huperzia squarrrosa                         | OneKP   | ERR364407 | GAON     |
| Inula helenium                              | OneKP   | ERR364393 | AFQQ     |
| Ipomoea purpurea                            | OneKP   | ERR364392 | VXKB     |
| Juniperus scopulorum                        | OneKP   | ERR364341 | XMGP     |
| Kadsura heteroclitica                       | OneKP   | ERR364331 | NWMY     |
| Klebsormium subtile                         | OneKP   | ERR364370 | FQLP     |
| Kochia scopania                             | OneKP   | ERR364385 | WGET     |
| Larrea tridentata                           | OneKP   | ERR364386 | UDUT     |
| Leucodon brachypus                          | OneKP   | ERR364353 | ZACW     |
| Marchantia emarginata                       | OneKP   | ERR364417 | TFIY     |
| Marchantia polymorpha                       | OneKP   | ERR364416 | JPYU     |
| Mesostigma viride                           | OneKP   | ERR364365 | KYIO     |
| Mesotaenium endlicherianum                  | OneKP   | ERR364377 | WDCW     |
| Metzgeria crassipilis                       | OneKP   | ERR364359 | NRWZ     |
| Monostis axiphostigma                       | OneKP   | ERR364362 | BTFM     |
| Mougeotia sp.                               | OneKP   | ERR364374 | ZRMT     |
| Nephrseelms pyriformis                      | OneKP   | ERR364363 | ISIM     |
| Nettium digitus                             | OneKP   | ERR364379 | FFGR     |
| Nothoceras oenigmaticus                     | OneKP   | ERR364356 | DXOU     |
| Nothoceras vincentianus                     | OneKP   | ERR364357 | TCBC     |
| Penium margaritaceum                        | OneKP   | ERR364382 | AEKF     |
| Podophyllum peltatum                        | OneKP   | ERR364384 | WFBF     |
| Polytrichum commune                         | OneKP   | ERR364413 | SZYG     |
| Prumnopitys andina                          | OneKP   | ERR364343 | EGLZ     |
| Pseudolycopodiella caroliniana              | OneKP   | ERR364345 | UPMJ     |
| Psilotum nudum                              | OneKP   | ERR364411 | QVMR     |
| Pynaminomas parkeae                         | OneKP   | ERR364361 | TNAW     |
| Rhynchostegium serralatum                   | OneKP   | ERR364355 | JADL     |
| Ricciocarpus natans                         | OneKP   | ERR364358 | WJLO     |
| Rosmarinus officinalis                      | OneKP   | ERR364391 | FDMM     |
| Rosulabryum cf. capillare                   | OneKP   | ERR364351 | XWHK     |
Meta-assemblies refer to a transcriptome assembled from more than one sequenced sample. Some of these were a combination of 1KP and other data; some were entirely non-1KP. Accession numbers (SRA or otherwise) are given for all of the transcriptomes that we used.
| File count | Median size (Mb) | Average size (Mb) | Largest size (Mb) | Total size (Mb) | Similar directories | iPlant directory name |
|------------|-----------------|-------------------|------------------|----------------|--------------------|-----------------------|
| 68,253     | 0.0             | 0.3               | 481.1            | 23,116.6       | onekp_pilot        |                       |
| 48,053     | 0.0             | 0.3               | 481.1            | 14,956.7       | onekp_pilot/orthogroups |                      |
| 19,220     | 0.1             | 0.7               | 243.8            | 13,276.5       | onekp_pilot/orthogroups/alignments |                |
| 9,610      | 0.1             | 0.3               | 79.8             | 3,289.6        | onekp_pilot/orthogroups/alignments/FAA |                |
| 9,610      | 0.2             | 1.0               | 243.8            | 9,986.9        | onekp_pilot/orthogroups/alignments/FNA |                |
| 28,833     | 0.0             | 0.1               | 481.1            | 1,680.2        | onekp_pilot/orthogroups/gene_trees |                |
| 9,611      | 0.0             | 0.1               | 481.1            | 583.3          | onekp_pilot/orthogroups/gene_trees/FAA |                |
| 9,610      | 0.0             | 0.0               | 0.5              | 102.2          | onekp_pilot/orthogroups/gene_trees/FNAA/genes |                |
| 19,222     | 0.0             | 0.1               | 458.0            | 1,096.8        | onekp_pilot/orthogroups/gene_trees/FNA |                |
| 9,611      | 0.0             | 0.1               | 458.0            | 556.6          | onekp_pilot/orthogroups/gene_trees/FNA/12_codon |                |
| 9,610      | 0.0             | 0.0               | 0.5              | 98.5           | onekp_pilot/orthogroups/gene_trees/FNA/12_codon/trees |                |
| 9,611      | 0.0             | 0.1               | 439.1            | 540.3          | onekp_pilot/orthogroups/gene_trees/FNA/all_codon |                |
| 9,610      | 0.0             | 0.0               | 0.5              | 101.2          | onekp_pilot/orthogroups/gene_trees/FNA/all_codon/dna_tree |                |
| 19,919     | 0.0             | 0.2               | 175.2            | 3,468.8        | onekp_pilot/phylogenetic_analysis |                |
| 2,556      | 0.1             | 0.1               | 292.7            | 142.7          | onekp_pilot/phylogenetic_analysis/alignments |                |
| 852        | 0.0             | 0.3               | 41.8             | 41.8           | onekp_pilot/phylogenetic_analysis/alignments/FNA |                |
| 852        | 0.1             | 0.1               | 125.5            | 125.5          | onekp_pilot/phylogenetic_analysis/alignments/FNA2AA |                |
| 17,197     | 0.0             | 0.1               | 1,827.3          | 1,827.3        | onekp_pilot/phylogenetic_analysis/gene_trees |                |
| 1,704      | 0.0             | 0.1               | 238.3            | 238.3          | onekp_pilot/phylogenetic_analysis/gene_trees/FAA |                |
| 2          | 0.3             | 0.1               | 0.3              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/FAA/raxmlboot.### |                |
| 3,408      | 0.0             | 0.1               | 476.7            | 476.7          | onekp_pilot/phylogenetic_analysis/gene_trees/FNA2AA |                |
| 2          | 0.3             | 0.1               | 0.3              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/FNA2AA/raxmlboot.### |                |
| 2          | 0.3             | 0.1               | 0.3              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/FNA2AA/raxmlboot.###c1c2 |                |
| 10,381     | 0.0             | 0.1               | 874.0            | 874.0          | onekp_pilot/phylogenetic_analysis/gene_trees/filtered |                |
| 2,548      | 0.0             | 0.1               | 169.3            | 169.3          | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FAA |                |
| 1          | 0.0             | 0.0               | 0.0              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FAA/raxmlboot.### |                |
| 1          | 0.2             | 0.1               | 0.2              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA |                |
| 852        | 0.0             | 0.0               | 0.0              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA/raxmlboot.### |                |
| 1          | 0.0             | 0.0               | 0.0              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA2AA |                |
| 6,980      | 0.0             | 0.1               | 700.9            | 700.9          | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA2AA/raxmlboot.### GAMMA.2 |                |
| 2          | 0.3             | 0.1               | 0.3              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA2AA/raxmlboot.###c1c2.GAMMA.2 |                |
| 2          | 0.3             | 0.1               | 0.3              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA2AA/raxmlboot.###c1c2.f25 |                |
| 1          | 0.0             | 0.0               | 0.0              | 852            | onekp_pilot/phylogenetic_analysis/gene_trees/filtered/FNA2AA/raxmlboot.###c1c2.f25 |                |
### Table 2 Number and size of data files on websites (Continued)

| File count | Median size (Mb) | Average size (Mb) | Largest size (Mb) | Total size (Mb) | Similar directories |
|------------|------------------|------------------|------------------|----------------|---------------------|
| 178        | 1,915.0          | 2,045.5          | 3,371.0          | 364,100.0      | Contents at SRA (PRJEB4921) |
| 2          | 1,915.0          | 2,045.5          | 3,371.0          | 4,091.0        | expecting per sample – uncompressed, but downloads are compressed to a quarter of these sizes |

In some instances, users will find many directories with similar names, as indicated in this table by hash (#) marks. The total number of directories is given in the preceding column.

![Figure 1 iPlant DE data window.](image_url)
Figure 2 Bulk download window if Java is disabled. Click on the circled link to access the instructions to install and configure an iDrop desktop.

Figure 3 Realigning a group of sequences using Muscle.
Figure 4 Phylozoom display of 1KP species phylogeny.

Figure 5 Phenylpropanoid synthesis pathway for *Colchicum autumnale*. Labelled rectangles are proteins. Small circles are metabolites. Black lines show the KEGG pathway. Red lines show the BioGRID interactions emanating from protein (K12355), which was interactively selected. A right-click on the protein will display the inferred function and a link to the sequence(s).
Abbreviations
1KP: 1,000 Plants project; DE: Discovery Environment; KEEG: Kyoto Encyclopedia of Genes and Genomes; NSF: National Science Foundation; SRA: Short Read Archive.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
CWD, BRR, NW, SWG, S Ma, BS, MM, DES, PSS, CR, LP, JAS, LD, DWS, JCV, TC, TMK, MR, RSB, MKD, and JLM collected the plant samples. NM, NW, S Mi, NN, TW, SA, MB, JBG, MAG, EW, JPD, CWD, BR, HP, BR, and JLM performed the phylogenomic analyses. NM, LHII, ZT, and EJC setup and maintained web-resources used to communicate data. LHII and RS performed the protein and KEGG pathway analyses. EJC, ZT, XW, XS, YZ, JW, and GKW generated the sequence data. GKW and JLM designed and oversaw the research. All authors read and approved the final manuscript.

Acknowledgments
The 1000 Plants (1KP) initiative, led by GKW, is funded by the Alberta Ministry of Agriculture, Innovates Centre of Research Excellence (iCORE), Musea Ventures, The Chinese Academy of Sciences, and the Alberta Ministry of Agriculture, Innovates Centre of Research Excellence (iCORE). The authors thank the many people responsible for sample collection on 1KP and the staff at BGI-Shenzhen for doing our sequencing. Phylogenomic analyses were supported by the US National Science Foundation through the iPlant collaborative. CANDO was funded by an NSF Pioneer Award 1DP1OD006779-01.

Author details
1Plant Collaborative, Tucson 85721, AZ, USA. 2Department of Ecology and Evolutionary Biology, University of Arizona, Tucson 85721, AZ, USA. 3Department of Microbiology, University of Washington, Seattle 98109, WA, USA. 4BGI-Shenzhen and China National GeneBank (CNGB). We thank the many people responsible for sample collection on 1KP and the staff at BGI-Shenzhen for doing our sequencing. Phylogenomic analyses were supported by the US National Science Foundation through the iPlant collaborative. CANDO was funded by an NIH Director’s Pioneer Award 1DP1OD006779-01.

Author details
1Department of Ecology, and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. 2Department of Microbiology, University of Washington, Seattle, WA, USA. 3Department of Botany, University of Alberta, Edmonton, AB, Canada. 4BGI-Shenzhen, Bei Shan Industrial Zone, Shenzhen, China. 5Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada. 6Chicago Botanic Garden, Glencoe, IL, USA. 7Program in Biological Sciences, Northwestern University, Evanston, IL, USA. 8Department of Computer Science, University of Texas, Austin, TX, USA. 9Department of Plant Biology, University of Georgia, Athens, GA, USA. 10Department of Biology, University of Florida, Gainesville, FL, USA. 11Department of Biology, Penn State University, University Park, Pennsylvania, PA, USA. 12Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Succursale Centre-Ville, Montréal, Québec H3C3J7, Canada. 13CNRS, USR 2936, Station d’Écologie Expérimentale du CNRS, Moulis 09200, France. 14Department of Biological Sciences, Eastern Kentucky University, Richmond, KY, USA. 15Florida Museum of Natural History, Gainesville, FL, USA. 16Département de Botany, University of British Columbia, Vancouver, BC, Canada. 17Airlord Arboretum of Harvard University, Cambridge, MA, USA. 18Botanical Institute, Universität zu Köln, Köln D-50674, Germany. 19Genetics Institute, University of Florida, Gainesville, FL, USA. 20Department of Biology, Duke University, Durham, NC, USA. 21Department of Biology, University of British Columbia, Vancouver, BC, Canada. 22Department of Botany, University of Alberta, Edmonton, AB, Canada. 23New York Botanical Garden, Bronx, NY, USA. 24Systematic Botany and Mycology, University of Minnesota, St. Paul, MN, USA. 25Department of Biology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. 26Department of Medicine, University of Alberta, Edmonton, AB, Canada.

Received: 22 May 2014 Accepted: 2 October 2014 Published: 27 October 2014

References
1. Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Geisser D, Matasci N, Wang L, Hanlon M, Lenards A, Muir A, Merchant N, Lowsy S, Mock S, Helmkem K, Kubach A, Narro M, Hopkins N, Micklos D, Hilgert U, Gonzales M, Jordan C, Skidmore E, Dooley R, Cazes J, McLay R, Lu Z, Pasternak S, Koesterke L, Pieli WH, et al. The iPlant Collaborative: Cyberinfrastructure for Plant Biology. Front Plant Sci 2011, 2:254.
2. 1KP BLAST Search Portal [http://www.bioinfodata.org/app/Blast1KPN/]
3. Johnson MT, Carpenter EJ, Tian Z, Biskulichewich R, Burrn JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, dePamphilis CW, Edger PP, Goh F, Graham S, Greiner S, Hibberd JM, Jordan-Thaden I, Kutchum TW, Leebens-Mack J, Melkonian M, Miles N, Myburg H, Patterson J, Pires JC, Ralph P, Rolf M, Sage RF, Solits D, Solits P, Stevenson D, Stewart CN Jr, et al. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One 2012, 7:e50226.
4. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 2014, 30:1660–1666.
5. Wickett NJ, Mirabab N, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayamalalyam S, Barker M, Burleigh JG, Gitzendanner MA, Ruhfel B, Wafu E, Der JP, Graham SW, Mathews S, Melkonian M, Solits DE, Solits ES, Miles NW, Rothfels C, Pokorny L, Shaw AJ, deGlorinio M, Stevenson DW, Sereb B, Villareal JC, Roule EB, Philippe H, dePamphilis CW, Chen T, et al. A phylotranscriptomics analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A IN PRESS.
6. iPlant Data Store for 1KP Pilot. [http://mirrors.iplantcollaborative.org/browse/plant/home/shared/onekp_pilot]
7. iPlant User Registration. [http://user.iplantcollaborative.org]
8. iPlant Learning Center. [http://www.iplantcollaborative.org/learning-center/all-tutorials]
9. iPlant Discovery Environment. [http://de.iplantcollaborative.org]
10. Using the iDrop Desktop. [https://pods.iplantcollaborative.org/wiki/display/DS/Using+iDrop+Desktop]
11. Matasci N, McKay J. Phylogenetic analysis with the iPlant discovery environment. Curr Protoc Bioinformatics 2013, 6:Unit 6.13.
12. Newick Trees Format. [http://evolution.genetics.washington.edu/phylip/newicktree.html]
13. iRODS Data Management Software [http://irods.org]
14. Using Commands (Unix). [https://pods.iplantcollaborative.org/wiki/display/DS/Using+Commands]
15. Minie M, Chopra G, Sethi G, Horst J, White G, Roy A, Hatti K, Samudrala R: CANDO and the infinite drug discovery frontier. Drug Discov Today 2014, 19:1353–1363.
16. BioGRID Interactions. [http://thebiogrid.org]
17. Kyoto Encyclopedia of Genes and Genomes (KEGG). [http://www.genome.jp/kegg]
18. 1KP Protein-Protein Interactions Mapped to Metabolic Pathways. [http://proteinfocus.org/1kp]
19. 1KP Capstone Objective. [https://pods.iplantcollaborative.org/wiki/display/pts/OneKP%3BCapstone%3BWiki]
20. 1KP Papers in Progress. [https://pods.iplantcollaborative.org/wiki/display/pts/OneKP%3Bcompanion%3Bpapers]
21. Sayou C, Monniaux M, Nanou IH, Moyroud E, Brockington SF, Thévenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y, Wong GK, Weigel D, Farcy M, Dumas R: A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. Science 2014, 343:645–648.
22. Klapperette NC, Murata Y, Kim SS, Pulver SR, Birdsey-Benson A, Cho YK, Morimoto TK, Chuong AS, Carpenter EJ, Tian Z, Wang J, Xie Y, Yan Z, Zhang Y, Chow BY, Surek B, Melkonian M, Jayaraman V, Constantine-Paton M, Wong GK, Boyden ES: Independent optical excitation of distinct neuronal populations. Nat Methods 2014, 11:358–364.

Cite this article as: Matasci et al.: Data access for the 1,000 Plants (1KP) project. GigaScience 2014, 3:17.

doi:10.1186/2047-217X-3-17

Cite this article as: Matasci et al: Data access for the 1,000 Plants (1KP) project. GigaScience 2014, 3:17.