

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/105927/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Joanna, Friesner, Sarah, Assmann, Bastow, Ruth, Julia, Bailey-Serres, Jim, Beynon, Volker, Brendel, Robin, Buell, Alexander, Buksch, Wolfgang, Busch, Taku, Demura, R, Dinneny Jose, J, Doherty Colleen, L, Eveland Andrea, Pascal, Falter-Braun, A, Gehan Malia, Michael, Gonzales, Erich, Grotewold, Rodrigo, Gutierrez, Ute, Kramer, Gabriel, Krouk, Shisong, Ma, RJ, Cody Markelz, Molly, Megraw, C, Meyers Blake, Murray, James, J, Provart Nicholas, Sue, Rhee, Roger, Smith, Edgar, Spalding, Crispin, Taylor, Tracy, Teal, U, Torii Keiko, Chris, Town, Matthew, Vaughn, Richard, Vierstra, Doreen, Ware, Olivia, Wilkins, Cranos, Williams and M, Brady Siobhan 2017. The next generation of training for arabidopsis researchers: Bioinformatics and Quantitative Biology. *Plant Physiology* 175 , pp. 1499-1509. 10.1104/pp.17.01490 file

Publishers page: <http://dx.doi.org/10.1104/pp.17.01490> <<http://dx.doi.org/10.1104/pp.17.01490>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



1 **The Next Generation of Training for Arabidopsis Researchers: Bioinformatics and Quantitative Biology**

2

3 Joanna Friesner¹, Sarah Assmann², Ruth Bastow³, Julia Bailey-Serres⁴, Jim Beynon⁵, Volker Brendel⁶, Robin
4 Buell⁷, Alexander Buksch⁸, Wolfgang Busch^{9,10}, Taku Demura^{11,12}, Jose R. Dinneny¹³, Colleen J. Doherty¹⁴,
5 Andrea L. Eveland¹⁵, Pascal Falter-Braun^{16,17}, Malia A. Gehan¹⁵, Michael Gonzales¹⁸, Erich Grotewold¹⁹
6 Rodrigo Gutierrez²⁰, Ute Kramer²¹, Gabriel Krouk²², Shisong Ma²³, R.J. Cody Markelz²⁴, Molly Megraw²⁵,
7 Blake C. Meyers¹⁵, Jim Murray²⁶, Nicholas J. Provart²⁷, Sue Rhee¹³, Roger Smith²⁸, Edgar Spalding²⁹, Crispin
8 Taylor³⁰, Tracy Teal³¹, Keiko U. Torii³², Chris Town³³, Matthew Vaughn³⁴, Richard Vierstra³⁵, Doreen
9 Ware^{36,378}, Olivia Wilkins³⁸, Cranos Williams³⁹, Siobhan M. Brady⁴⁰⁺

10

11 ¹ Agriculture Sustainability Institute and Department of Neurobiology, Physiology and Behavior, University
12 of California, Davis, CA, USA 95616

13 ² Biology Department, Penn State University, University Park, PA, USA, 16802

14 ³ GARNet, School of Biosciences, Cardiff University, Cardiff, UK

15 ⁴ Center for Plant Cell Biology and Botany and Plant Sciences Department, University of
16 California, Riverside, CA, USA 92521

17 ⁵ School of Life Sciences, Gibbet Hill Campus, The University of Warwick, Coventry, CV4 7AL, UK

18 ⁶ Department of Biology and Department of Computer Science, Indiana University, Bloomington, IN, USA
19 47405

20 ⁷ Department of Plant Biology, Michigan State University, East Lansing, MI, USA 48824

21 ⁸ University of Georgia, Department of Plant Biology; Warnell School of Forestry and Natural Resources;
22 and Institute of Bioinformatics, Athens, GA, USA 30602

23 ⁹ Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030
24 Vienna, Austria

25 ¹⁰ Salk Institute for Biological Studies, Plant Molecular and Cellular Biology Laboratory, 10010 N Torrey Pines Rd,
26 La Jolla, CA, USA 92037

27 ¹¹ Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, 630-
28 0192, Japan

29 ¹² RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, 230-0045, Japan

30 ¹³ Carnegie Institution for Science, Department of Plant Biology, Stanford, CA, USA 94305

31 ¹⁴ Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC ,
32 USA 27695

33 ¹⁵ Donald Danforth Plant Science Center, St. Louis, MO, USA, 63132

34 ¹⁶ Institute of Network Biology (INET), Helmholtz Zentrum München (HMGU), German Research Center for
35 Environmental Health, 85764 Neuherberg, Germany

36 ¹⁷ Department of Microbe-Host Interactions, Ludwig-Maximilians-Universität München (LMU Munich),
37 Planegg-Martinsried, Germany

38 ¹⁸ Center for Applied Genetic Technologies (CAGT), 111 Riverbend Road, Athens, GA, USA 30602

39 ¹⁹ Center for Applied Plant Sciences and Dept. of Molecular Genetics, The Ohio State University,
40 Columbus, OH, USA 43220.

41 ²⁰ FONDAP Center for Genome Regulation. Millennium Nucleus Center for Plant Systems and Synthetic
42 Biology. Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia
43 Universidad Católica de Chile, Avenida Libertador Bernardo O'Higgins 340, Santiago, Chile 833115

44 ²¹ Molecular Genetics and Physiology of Plants, Faculty of Biology and Biotechnology, Ruhr University
45 Bochum, 44801 Bochum, Germany

46 ²² Laboratoire de Biochimie et Physiologie Moléculaire des Plantes, UMR CNRS/INRA/SupAgro/UM,
47 Institut de Biologie Intégrative des Plantes "Claude Grignon," Place Viala, 34060 Montpellier Cedex,
48 France.

49 ²³ School of Life Sciences, University of Science and Technology of China, 443 Huangshan Road, Hefei,
50 Anhui, 230027, China
51 ²⁴ Department of Plant Biology, University of California, Davis, CA, USA 95616
52 ²⁵ Department of Botany and Plant Pathology; Department of Computer Science; and Center for Genome
53 Research and Biocomputing, Oregon State University, Corvallis, OR, USA 97331
54 ²⁶ School of Biosciences, Sir Martin Evans Building, Cardiff University, Museum Avenue, Cardiff CF10 3AX,
55 Wales, UK.
56 ²⁷ Department of Cell & Systems Biology / Centre for the Analysis of Genome Evolution and Function,
57 University of Toronto, Toronto, ON. M5S 3B2, Canada
58 ²⁸ Syngenta Biotechnology, Inc. PO Box 122573054 E. Cornwallis Road, Research Triangle Park, NC 27709,
59 USA
60 ²⁹ Department of Botany, University of Wisconsin, 430 Lincoln Drive, Madison, WI, USA 53706
61 ³⁰ American Society of Plant Biologists, Rockville, MD, USA 20855
62 ³¹ Data Carpentry, Davis, CA, USA 95616
63 ³² Howard Hughes Medical Institute and Department of Biology, University of Washington, Seattle, WA,
64 USA 98195
65 ³³ J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD, USA 20850
66 ³⁴ Life Sciences Computing, Texas Advanced Computing Center, 10100 Burnet Rd, Austin, TX, USA 78758
67 ³⁵ Department of Biology, Washington University in St. Louis, St. Louis, MO, USA 63130
68 ³⁶ Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA 11724
69 ³⁷ US Department of Agriculture, Agricultural Research Service, Ithaca, New York, USA 14853
70 ³⁸ Department of Plant Science, McGill University, Montreal, QC, Canada H9X 3V9
71 ³⁹ Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA
72 27695
73 ⁴⁰ Department of Plant Biology and Genome Center, University of California, Davis, CA, USA 95616
74 +Corresponding author – sbrady@ucdavis.edu

75 **Abstract**

76 Researchers studying plants using the model organism, *Arabidopsis thaliana*, can easily generate or access
77 massive datasets using modern technologies. However, in order to best analyze such datasets to
78 elucidate novel biological mechanisms, many individuals face critical deficiencies in their training. Ideally,
79 these scientists will be able to, individually or in a team, integrate foundational concepts from biological
80 science, chemistry, mathematics, statistics, computer science, bioinformatics and data science. Here, we
81 provide examples of guidelines, skill sets, and core competencies that should be considered when
82 developing curricula or training efforts at the undergraduate, graduate, postdoctoral and faculty levels.
83 Discussion of specific training needs from the perspective of the agricultural biotechnology industry are
84 also provided. Critical to “large-scale biology” is the formation of productive collaborations. Methods to
85 identify the best collaborator, to define an effective collaboration on the part of all partners, and
86 pedagogical methods to train students in the art of collaboration are also discussed. Finally, these
87 challenges and potential solutions are addressed in a selected case study on high-throughput
88 phenotyping.

89

90 **Introduction**

91 It has been over 50 years since *Arabidopsis thaliana* was first introduced as a model organism to
92 understand basic processes in plant biology. A well-organized scientific community has used this small
93 “reference” plant species to make numerous fundamental plant biology discoveries (Provart et al., 2016).
94 Due to an extremely well annotated genome and advances in high-throughput sequencing, our
95 understanding of this organism and other plant species has become ever more intricate and complex.
96 Computational resources including CyVerse¹, Araport², TAIR³ and BAR⁴ have further facilitated novel
97 findings with just the click of a mouse. As we move towards understanding biological systems,
98 *Arabidopsis* researchers will need to use more quantitative and computational approaches in order to
99 extract novel biological findings from these data. Here, we discuss guidelines, skill sets, and core
100 competencies that should be considered when developing curricula or training undergraduate or
101 graduate students, postdoctoral researchers, and faculty. A selected case study provides more specificity
102 as to the concrete issues that plant biologists face and how best to address such challenges.

103

104 **Transforming Education and Training – from Undergraduates to Faculty**

105 An overhaul in training is necessary for plant biologists to make use of massive data sets and enabling
106 technologies. This is not a novel idea in the life sciences. In fact, Bialek and Botstein (2004) articulated a
107 concept for an integrated introductory quantitative science curriculum, primarily for undergraduates, in
108 order to address this specific issue. Their publication has been highly cited and used as a foundational
109 resource. They noted that biologists have too little education and experience in quantitative thinking and
110 computation relative to what is needed for full participation in this new era of genomics research. Both
111 then, and still now, many upper-level undergraduates in the life sciences versus quantitative sciences
112 already speak “noticeably different languages”. Bialek and Botstein (2004) proposed that instead of
113 prerequisite courses in mathematics, physics, chemistry and computation, the fundamental ideas of each
114 of these disciplines should be introduced at a high level of sophistication. Their point is that these ideas
115 should be presented in context and with relevant biological problems for a “seamless” educational
116 experience. This would also avoid the delivery of these quantitative science courses as a “service” for the
117 life sciences students. In a “service course”, students often exhibit a lack of enthusiasm due to the fact
118 that they are required to take these courses. An additional issue is that many of the quantitative concepts
119 presented are devoid of a biological perspective. Training at the graduate level must also necessarily

¹ <http://www.cyverse.org/>

² <https://www.araport.org/>

³ <http://www.arabidopsis.org/>

⁴ <http://bar.utoronto.ca>

120 integrate foundational concepts from biological science, chemistry, mathematics, statistics, computer
121 science, bioinformatics and data science. We stress that this is more than simply an understanding of
122 bioinformatics - that is, more than just using computation to extract knowledge from biological data.
123 Instead, education in plant biology should be truly interdisciplinary, perhaps as exemplified by (i)
124 theoretical biology whereby theoretical perspectives (often mathematical) are used to give insights into
125 biological processes, (ii) quantitative biology whereby quantitative approaches and technologies are used
126 to analyze and integrate biological systems or to construct and model engineered life systems, or (iii)
127 computational biology whereby biological data are used to develop algorithms or models to understand
128 relationships amongst various biological systems.

129

130 **Implementation of Quantitative Training in the Life Sciences**

131 Significant administrative, content, and logistical challenges often exist to impede the creation of new
132 academic programs. Despite this, a growing number of institutions are developing undergraduate and
133 graduate curricula in bioinformatics and computational biology for the life sciences, many of which
134 incorporate the vision of Bialek and Botstein⁵. Practical strategies to overcome many of these challenges
135 have been described for an overhaul in the graduate training program at Harvard Medical School
136 (Gutlerner and Van Vactor, 2013). Our primary recommendation is to include in life sciences curricula the
137 teaching of the skills and competencies described above, with the aim to develop students and future
138 scientists that are adept at using transdisciplinary approaches to solve challenges in biology, and thus,
139 well adapted to addressing current and future needs in modern plant biology research.

140

141 **Minimal Skill Sets and Core Competencies**

142 Over the last 15+ years a variety of meetings and task forces have been convened to determine the
143 nature, extent, content, and available delivery tools for degree and training programs utilizing
144 bioinformatics or computational biology in life sciences programs. Tan et al. (2009) proposed a
145 generalized minimum set of competencies that the next generation of biologists will need to effectively
146 cope with ever-increasing amounts of information and datasets, and the growth of importance in
147 informatics in this genomics era; the following competencies have increased in relevance since they were
148 first published, and thus could guide curricula development (or revisions of existing curricula):

- 149 1. Basic knowledge in the specific domains of computer science, statistics and mathematics that
150 intersect with modern biology.

⁵ http://www.bioinformatics.org/wiki/education_in_the_united_states

- 151 2. Expertise in communicating and representing biological knowledge and processes in
 152 mathematical, statistical and computing terms and concepts.
- 153 3. The ability to use or develop efficient bioinformatics and biocomputational tools and techniques
 154 for the acquisition, interpretation, analysis, prediction, modeling, simulation and visualization of
 155 experimental and other biological data.
- 156 4. Proficiency in the search, retrieval, processing, curation, organization, classification,
 157 management and dissemination of biological data and information in databases for deriving
 158 biological insights and knowledge discovery.
- 159 5. Critical thinking and problem-solving skills in quantitative aspects of biology.

160 As a community with expertise in quantitative and computational plant biology, and using these
 161 competencies as a guideline, we further propose a suite of minimal skill sets [adapted from (Rubinstein
 162 and Chor, 2014; Welch et al., 2014)] which will enable a plant biologist to generate and utilize multi-
 163 dimensional and scaled plant biology data in order to answer central biological questions (**Table 1**).

164 **Table 1:** Minimal Skill Sets Recommended for Plant Biology Students

Category	Specific Skills
Unix/Linux	Comfort/familiarity with using command line
Scripting Language	Perl or Python, for advanced students – C++, CUDA
Database creation and query	Mongo or MySQL, data mining
Software carpentry	Best practices, proper commenting, version control
Computation	Machine learning, algorithm design and analysis, distributed and high-performance computing
Statistical methods	Descriptive and inferential statistics, hypothesis testing, parameter estimation, power analysis, data transformations, meta-analysis, hierarchical clustering
Mathematical	Probability theory, differential equations, graph theory. Linear algebra, information theory
Statistical programming	R/Bioconductor (particularly for analysis of next-generation sequencing data)
Biological databases and resources	NCBI, EBI, Araport, TAIR, MaizeGDB, Gene Ontology etc.
Network Analysis	Cytoscape plugins
Data Visualization	Could include ggplot, visualization of genome-scale data in genome browsers, volcano plots, heat maps etc.

165

166 We suggest two possibilities to implement across diverse institutions this integrated paradigm for training
 167 in this suite of minimal skill sets and core competencies. So as not to reinvent the wheel, it may be fairly
 168 straightforward for a plant biology program to participate in an extant integrative biology/quantitative
 169 sciences program within their respective institution, if those programs fulfill this suite of core
 170 recommended competencies/skill sets, simply by augmenting existing programs with elective plant

171 courses. Alternatively, a program could implement course curricula (both undergraduate and graduate)
172 that have been described in the literature and for which resources are available. These include the
173 Course Source Bioinformatics Learning framework, which has been developed and reviewed by members
174 of the Genomics Education Partnership, the Network for Integrating Bioinformatics into Life Science
175 Education, the Genome Consortium for Active Teaching of Next Gen Sequencing, and the Howard Hughes
176 Medical Institute-sponsored Bioinformatics Workshop for Student/Scientist Partnerships (Rosenwald et
177 al., 2016). Other curricula include a basic bioinformatics curriculum offered at the Free University of
178 Berlin which emphasizes fundamentals in biology, mathematics and computer science (Koch et al., 2008),
179 or a first-year graduate course in quantitative biology which emphasizes the integrated curriculum
180 proposed by Bialek and Botstein (2004). The latter example uses breakthrough papers in diverse areas of
181 biology, and that emphasize quantitative reasoning, theory, and experimentation, to convey the
182 importance of quantitative knowledge to understand basic biological processes (Wingreen and Botstein,
183 2006). Similar curricula have been implemented in the UK and are considered requisite training for
184 graduate students in plant biology⁶. A course entitled “Computational Approaches for Life Scientists⁸”
185 has also been described which focuses on enriching the curriculum of life science students with abstract,
186 algorithmic and logical thinking and exposes them to “computational culture” (Rubinstein and Chor,
187 2014). Such curricula should be followed by a more focused track in plant biology, again emphasizing the
188 quantitative premises underlying plant biology. Finally, a capstone problem-solving course that integrates
189 teamwork could provide practical examples of how to integrate these diverse and interdisciplinary subject
190 materials to address unsolved questions in plant biology.

191

192 **Bridge Programs, Bootcamps and Supportive Environments for Quantitative-Based Plant Biology** 193 **Education**

194 Even without creating new programs, supportive environments for students interested in both
195 plant and computational biology could help lower the “intimidation” barrier. For example, this could
196 involve creating quantitative biology interest groups. Additional vehicles to encourage peer-to-peer
197 learning could include hackathons (events that bring people together in teams for collaborative computer
198 programming efforts to creatively solve a problem) that would provide training, while encouraging
199 interactions between plant biology and computational students.

200 Recently, organizations such as Software Carpentry⁷ and Data Carpentry⁸ (which are merging into
201 one organization) and Amelieff⁹ have been created to fill in some of the gaps in education for

⁶ <https://sysmic.ac.uk>

⁷ <http://ca4ls.wikidot.com>

⁸ <http://www.datacarpentry.org/>

202 programming and data science skills. Since 2015, these organizations have held workshops at institutions
203 across the world. Other short courses also exist globally which focus on training experimental biologists in
204 bioinformatics, statistical genetics and mathematical modeling including the Summer Institute of
205 Statistical Genetics (USA)¹⁰, the Summer School for Statistical Genetics (Japan)¹¹, the Santa Barbara
206 Advanced Summer School of Quantitative Biology (USA)¹², the BioComp training series (Austria), the
207 Summer School (Germany)¹³, the Saclay Plant Sciences summer schools (France)¹⁴, the Integrative
208 Database training course (Japan)¹⁵, the Large Biological Data Analysis Course (Japan)¹⁶, and the Cold
209 Spring Harbor Laboratory courses¹⁷ (USA) in “Frontiers and Techniques in Plant Science” and
210 “Programming for Biology”. However, access to these courses is limited, and the course fees and travel
211 necessary to participate may present significant barriers. In order to enhance the flexibility and to
212 minimize financial input, curricula could be complemented with short-courses or with certificates from
213 online Massive Open Online Courses (MOOCs). As a community, developing a portal that provides reviews
214 and ratings of these programs would be a valuable resource (Searls, 2012). It should be noted, however,
215 that a recent report assessing boot camp programs (from 2 days to 2 weeks in length) typically designed
216 to expose graduate students to data analysis techniques (amongst others) found a null difference when
217 assessing research skill development, despite a statistically significant increase in perceived skill
218 advancement (Feldon et al., 2017).

219

220 **Funding**

221 While many academic institutions recognize the importance of these training efforts, they need
222 funding to come into existence. The United States National Science Foundation (NSF) Research
223 Traineeship (NRT) Program¹⁸ Traineeship Track specifically fosters interdisciplinary training. The German
224 Research Foundation provides funding for International Research Training Groups dedicated to a focused
225 “study abroad” research program and a structured training strategy. In France, local funding agencies
226 named LABEX (for “Laboratoire d’Excellence”) fund interdisciplinary interactions between local partners,

⁹ <http://amelieff.jp/english/>

¹⁰ <https://www.biostat.washington.edu/suminst/sisg>

¹¹ <http://www.sg.med.osaka-u.ac.jp/school.html>

¹² <https://www.kitp.ucsb.edu/qbio>

¹³ GCBN/de.NBI

¹⁴ https://www6.inra.fr/saclay-plant-sciences_eng/Teaching-and-training/Summer-schools/Summer-School-2016

¹⁵ <https://biosciencedbc.jp/en/>

¹⁶ <https://biosciencedbc.jp/en/>

¹⁷ <http://meetings.cshl.edu/courseshome.aspx>

¹⁸ https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505015

227 an example being Numev¹⁹, which promote interactions between computer and mathematical scientists
228 and biologists with strong support of plant scientists. The CNRS (Centre National de la Recherche
229 Scientifique) regularly promotes biology and math interactions through specific grant calls led by its Office
230 for Interdisciplinary Research (PEPS). In many of these cases, however, proposals are granted only for
231 specific areas deemed to be a 'high priority' to each funding organization, which may lower the success of
232 proposals that do not fit easily into the chosen scope.

233

234 **Additional Recommendations for Postdoctoral Scholars and Faculty**

235 At the moment, there are no standardized modes of quantitative or interdisciplinary training for
236 postdoctoral fellows in plant biology. Thus, postdoctoral scholars often need to identify their own
237 opportunities for additional training, if they have not received such training during their undergraduate or
238 graduate training. Many competitive postdoctoral scholar fellowships offer funds for additional training
239 including NSF's Plant Genome Research Program Postdoctoral Research Fellowships in Biology (PGRP
240 PRFB)²⁰, the USDA's AFRI Food, Agriculture, Natural Resources, and Human Sciences Education and
241 Literacy Initiative Fellowship program (AFRI ELI)²¹ and the National Institute of Health K99 grant
242 program²². The Human Frontiers Science Program offers postdoctoral fellowships for citizens of many
243 countries with a special category for cross-disciplinary fellowships to support training those in
244 quantitative sciences in experimental biology²³. Moreover, the European Union's Marie Skłodowska-Curie
245 Actions Individual Fellowships offer funds for additional training and for short 3 to 6 month visits. The
246 Plant Biology section of the General Program and the Young Scientists Fund of the National Natural
247 Science Foundation of China (NSFC) encourages interdisciplinary research that combine methods from
248 plant biology and other areas, such as mathematics, physics, and computer sciences²⁴.

249 However, these fellowships are quite competitive and can be restricted to postdoctoral scholars
250 trained in certain disciplines. What if a postdoctoral scholar is unsuccessful at receiving such funds but still
251 wishes to undergo interdisciplinary training? In Germany, there is a growing number of structured
252 postdoctoral fellowship programs funded by individual research institutions that offer institutional
253 support in identifying interdisciplinary training opportunities. The Postdoctoral Fellowship Program (PFP)
254 by the HelmholtzZentrum Munich ensures that fellows are integrated into international and

¹⁹ <http://www.lirmm.fr/numev/>

²⁰ https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503622

²¹ <https://nifa.usda.gov/program/afri-education-and-literacy-initiative>

²² <https://www.nhlbi.nih.gov/research/training/programs/postdoc/pathway-parent-k99-r00>

²³ <http://www.hfsp.org/funding/postdoctoral-fellowships>

²⁴ <http://www.nsf.gov.cn/nsfc/cen/xmzn/2017xmzn/01/03sm/001.htm>

255 interdisciplinary research groups, while the University Foundation Fellowship Program by the Technical
256 University of Munich assists with the identification of interdisciplinary and collaborative research
257 programs. Additional institutional solutions could provide the resources for postdoctoral participation
258 (and instruction) in short courses that provide training in a particular competency, or could integrate
259 postdoctoral scholars in existing courses provided for graduate students. At the mid- and senior-
260 postdoctoral scholar level, perhaps the best way is to provide opportunities for senior “biologically-
261 oriented” postdoctoral scholars to engage in dedicated training via short-term “residencies” (3 to 6
262 months) in a laboratory that specializes in quantitative, computational, or modeling analyses. Such
263 longer-term dedicated learning programs would have the advantage of carrying out a distributed practice
264 of learning, which has proven more beneficial in long-term retention of concepts, relative to the shorter
265 mass “boot-camp”-type strategy. (Feldon et al., 2017).

266 Short-term or long-term sabbaticals in a computational lab are also a good solution for faculty
267 members to acquire computational skills. The USA National Science Foundation’s Mid-Career Investigator
268 Awards in Integrative Organismal Biology (MCA-IOB)²⁵ could be a source of funding for associated travel
269 costs. The German Academic Exchange Service (DAAD) and the French AGreskills federal programs, as well
270 as the local LabEx programs (mentioned previously) financially support sabbaticals for this purpose.

271 Alternatively, it may be better for faculty to focus on how they can better assess and support research
272 activities in their own lab and be able to better understand how to review papers or grants that contain
273 research of an interdisciplinary nature. Short workshops could be developed to provide training to faculty
274 on quantitative and computational methods and how to conduct high-quality computational/quantitative
275 research.

276

277 **Computational Training for Industry**

278 The key attributes for researchers in industry with respect to projects involving computational approaches
279 are strong interpersonal skills in teamwork, collaboration, communication, and project management.
280 Industry requires individuals who are expert in one specific area but have the breadth of understanding
281 that allows them to appreciate and respect the input of other disciplines to the overall project. This
282 includes familiarity with biological databases and quantitative biology approaches. In addition, employees
283 in industry benefit from training programs which expose workers in academia and industry to each other’s
284 ways of working. The European funding model encourages partnerships between researchers and
285 industry (e.g., the bread wheat initiative led by INRA²⁶). Another model is to embed master’s or doctoral
286 students in industry placements for three to six months. Two UK-specific examples of this are the

²⁵ <https://www.nsf.gov/pubs/2017/nsf17508/nsf17508.htm>

²⁶ <http://www.wheatinitiative.org>

287 compulsory program of the UK Biotechnology and Biological Sciences Research Council (BBSRC), called
288 “PIPS” (Professional Internships for PhD Students²⁷), and the Flexible Interchange Program (FLIP²⁸) which
289 operates at the postdoctoral scholar and faculty level to promote training and exchange between industry
290 and academic partners. An additional twist on this theme is provided by the Chilean scientific funding
291 agency CONICYT that offers a post-graduate thesis in industry²⁹. At the institutional level, research
292 institutions dedicated to applied sciences and industrial cooperation, like the Fraunhofer Institutes in
293 Germany, traditionally work in close cooperation with industry including master's and doctoral students.
294

295 **Arabidopsis Training for Plant-Curious Data Scientists**

296 A rapidly growing world population and a changing climate demand development of improved
297 crop varieties that yield more with fewer inputs, as well as advances in renewable fuels, and biomaterials.
298 Moving forward, a community-wide effort to promote the value of plant science research to data
299 scientists is needed. Arabidopsis training for “plant-curious” data scientists should emphasize 1) how
300 knowledge gained from Arabidopsis research is relevant to crop improvement, and 2) how to utilize
301 Arabidopsis as a tool to rapidly test gene function and optimize emerging technologies prior to delivery to
302 a crop system. The advent of gene editing technologies, such as CRISPR/Cas9-based approaches to
303 specifically target loci for site-directed mutagenesis or sequence replacement, introduces a new
304 paradigm. While these technologies create opportunities for targeted mutagenesis directly in crop
305 species, significant bottlenecks in the transformation process limit the extent to which these experiments
306 can be performed in crops. Therefore, Arabidopsis can be used to more quickly and efficiently test
307 functional hypotheses and prioritize experiments for the more labor-, cost-, and time-intensive studies in
308 crops.

309 The outcome of an active community of Arabidopsis researchers is the detailed curation of genes
310 and pathways in the Arabidopsis genome, perfect for mining by data scientists. This curation data has
311 been leveraged for annotating orthologous genetic components in other species, and thus is an invaluable
312 resource. It is likely that many fundamental biological processes are conserved across plant species
313 (McGary et al., 2010; Oellrich et al., 2015). As an example, agricultural biotechnology industries make use
314 of this information through large-scale text mining algorithms combined with comparative genomics
315 approaches to project annotations and associations onto crop models (Holtan et al., 2011; Preuss et al.,
316 2012). The depth and breadth of these resources in Arabidopsis also positions this organism at the

²⁷ <http://www.bbsrc.ac.uk/funding/filter/professional-internships/>

²⁸ <http://www.bbsrc.ac.uk/funding/filter/flexible-interchange-programme/>

²⁹ <http://www.conicyt.cl/wp-content/uploads/2012/07/Brochure-Institucional-2011-Ingles.pdf>

317 forefront of predictive modeling in plants through systems biology approaches. Moving forward, there is
318 an immediate need to make better use of existing data from Arabidopsis studies by developing new data
319 integration paradigms aimed at predictive modeling and subsequent discovery. Using Arabidopsis as a
320 framework for how to integrate diverse datasets should facilitate similar analyses in species with less
321 developed resources.

322 On the other hand, Arabidopsis may not be the most appropriate model to understand traits
323 related to domestication, physiology such as C4 photosynthesis, or other aspects of plant biology such as
324 secondary metabolism. To address such questions alternative model systems are being established; this
325 includes *Setaria viridis* and *Brachypodium distachyon* as model grass species (Brutnell, 2015; Brutnell et
326 al., 2015) and *Camelina sativa* for metabolic engineering of co-products (Bansal and Durrett, 2016; Zhu et
327 al., 2016). We recommend that the communities developing these new systems leverage best practices
328 from the Arabidopsis community, particularly with reference to genome annotation and data curation for
329 these species. Fostering such interactions between scientists could occur through cross-species
330 conferences in plant science; for example, a Keystone Meeting focused on “Translational Plant Biology”.
331 Inclusion of data scientists in these forums will be critical to ensure maximal usefulness of these emerging
332 model systems.

333

334 **Collaborations**

335 Taking advantage of large-scale datasets and technologies in order to reveal novel biological conclusions
336 will require groups of people with diverse expertise, skill sets, and at different career levels to work well
337 together. Thus, in order to train the next generation of Arabidopsis biologists in quantitative and
338 computational biology, we also need to train scientists on how to initiate, define, manage and maintain
339 effective collaborations.

340

341 **Identifying collaborators**

342 It is often difficult for biologists to develop their research questions to include tangible opportunities for
343 quantitative experts, or to effectively articulate their specific needs in a vocabulary that is accessible to
344 experts in those fields. Face-to-face communication is particularly important and thus we attribute the
345 highest priority to the identification of regional collaborators. Inclusive, regular, cross-faculty and cross-
346 institute interactions at all career levels, with the clear objective to also empower early-career
347 researchers to take active roles, are required to initiate local collaborations. In order to implement role
348 models for such collaborative efforts, hiring or recruiting researchers who already work across biological
349 science and statistical, computational, or mathematical departments can be beneficial due to their ability
350 to expose biological problems to theoreticians who might not typically see such data as valuable to

351 analyze. However, the infrastructure for promotion and merit within most academic institutions has
352 generally not advanced sufficiently to effectively hire and maintain theoreticians at the tenure-track level
353 in biology departments.

354 Collaborations between disciplinary experts can be accelerated through intensive trainings and
355 activities that promote networking and knowledge sharing. In-depth, week-long immersion sessions have
356 proven effective at providing both the biologist and the quantitative expert with the proper, shared,
357 vocabulary, resulting in productive collaborations. For example, the “Maths in the Plant Sciences” Study
358 Group in the United Kingdom³⁰ has been successful in generating in short timeframes both new
359 collaborations and funded grant applications.

360 Co-supervision of graduate students by a biologist and theoretician is another effective strategy
361 to develop a collaboration. Initiating cross-disciplinary cohorts of graduate students is another approach.
362 Complementing collaborative interactions, or in the absence of local cross-disciplinary opportunities, the
363 availability of more high-quality online video material outlining advances in current plant biology, for
364 example, in a jargon-free format would be useful for quantitative experts. In the long term, graduate
365 students and postdoctoral scholars who have been trained in an interdisciplinary environment will likely
366 generate the best collaborations. By working together from an early career stage, a deep appreciation of
367 diverse abilities will be engendered and the ability to communicate freely will enable new research
368 avenues to be pursued.

369

370 **Defining collaborations**

371 An effective multi-disciplinary collaboration must go beyond the mere provision of a service by a
372 collaborator. As such, before initiating a project, all partners should jointly articulate and agree on the
373 scientifically interesting research questions and discuss experimental design and data analysis. A
374 management plan should involve contributors at all career levels and consider the benefit for each
375 contributing individual. It is important for collaborators to recognize differences in cost bases for
376 biological versus theoretical research (e.g. experimental laboratory-associated costs are quite high
377 whereas in the theoretical sciences, experts command higher salaries than experimental biologists). A
378 realistic assessment of project timelines and deliverables is critical. Furthermore, a plan to include
379 periodic assessment of progress with respect to the defined timelines and deliverables should be
380 implemented to allow for adaptation, with the understanding that things do not always proceed
381 expectedly. Contingency plans are also ideal to establish at the start, as are plans for publications, since
382 biological and theoretical fields have fundamentally differing authorship rules and norms, publication
383 strategies, and career recognition criteria. It is important to discuss and specify the timeframes that are

³⁰ <https://www.cpib.ac.uk/outreach/mpssg/>

384 likely for the publication of biological data and how the development of novel theory or analysis tools
385 could be published prior to their use in biological data analysis. To ensure recognition, CRedit³¹ (through
386 ORCID) comprises structured vocabulary for assigning author credit. It is also critical to put in place an
387 explicit plan for the possibility of managing disagreements that may arise as well as the conditions under
388 which a collaborator might exit a project.

389 In practice, project meetings between collaborators should be held at more frequent intervals
390 than may normally occur in within-discipline collaborations. This is especially true at the beginning of the
391 project where the development of mutual understanding and the building of close working relationships
392 among the researchers are essential. If the collaboration is between local groups, regular, e.g. monthly,
393 joint meetings would be ideal. If the collaboration involves partners at a considerable geographic
394 distance, then monthly web-based meetings are necessary and the collaboration would benefit from face-
395 to-face meetings with all team members, ideally once every six months at a minimum. Budgeting for
396 necessary travel should be considered at the time of project design. Furthermore, the physical movement
397 of postdoctoral scholars or graduate students between groups for reciprocal training or joint work
398 contributes highly to the effective integration of projects. Appreciation of differences in language or
399 culture should be conveyed, as should reciprocal trust and respect, interest in the mutual fields, and the
400 willingness to learn from the expertise of a partner.

401

402 **Case Study: Training Arabidopsis Biologists for High-Throughput Phenotyping**

403 As a concrete example for how scientists can be trained and educated in an interdisciplinary, collaborative
404 fashion using experimental biology and quantitative approaches, we consider phenomics as a case study.
405 Phenomics is an emerging field at the intersection of plant biology, engineering, computer science and
406 mathematics which has led to a deeper understanding of mechanisms for acclimation to environmental
407 variation (Miller et al., 2007; Slovak et al., 2014; Campbell et al., 2015; Fahlgren et al., 2015; Rellán-
408 Álvarez et al., 2015). These studies evolved from the need to characterize phenotypic traits across large
409 numbers of genotypes (Chen et al., 2014; Cruz et al., 2016; Ge et al., 2016).

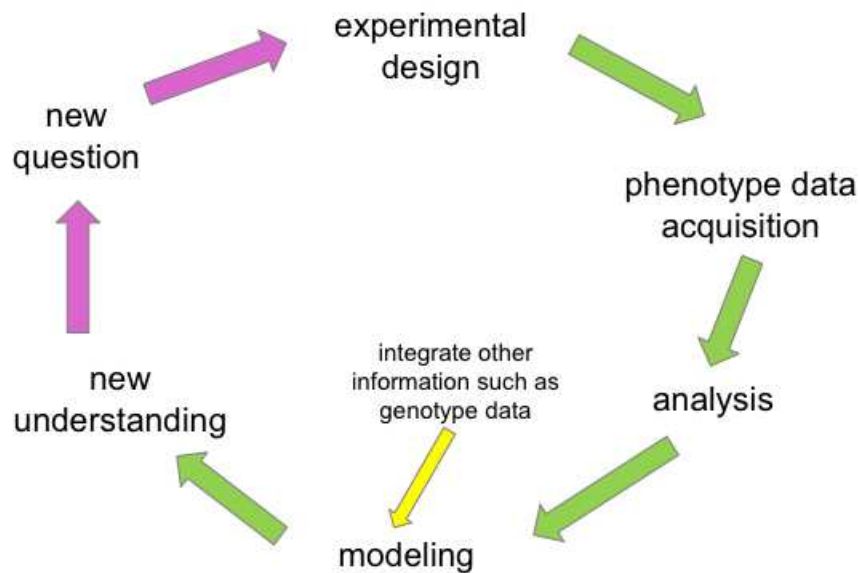
410 A project using phenomics can be considered as a pipeline with three identifiable stages: data
411 acquisition, data analysis, and data modeling (Figure 1), all to answer a clear biological question.
412 Generally, this question is: what genes or genetic regulation underlie a trait of interest? Generally, a
413 consortium of scientists is needed to carry out a phenomic-scale project. Consortium members should
414 have diverse skills, be able to interact collaboratively, and each researcher's role should be well defined.
415 Prior to data acquisition, consortium members should collectively discuss and agree upon experimental

³¹ <http://casrai.org/credit>

416 design, biological replicates, statistical power, the type of data to be acquired and appropriate models
417 used for data analysis. The data acquisition stage includes the use of sensors such as cameras, fluorescent
418 measurement devices, or any tool that can make a measurement when connected to a computer to
419 measure a phenotype associated with a trait. This stage often leverages expertise in the engineering
420 disciplines and may involve robotics. Input from biologists is needed in order to ensure that a
421 physiologically relevant aspect of plant growth or response to the environment is being captured. The
422 output of this stage is the generation of raw data files. The analysis stage includes the computer code
423 needed to extract features from the raw data files - image analysis is a good example - to produce
424 “measurements”. This stage also includes ‘workflow’ software, which brings the raw data from the
425 sensors to the analysis algorithms. The analysis phase passes processed data, or results, to the next stage.
426 Again, input from an “experimental” biologist is needed to ensure that these data are within the expected
427 range of values. The modeling stage involves synthesizing results for the purpose of generating new
428 biological conclusions. A typical example would be integrating phenotype results with genotype
429 information to complete a statistical genetic analysis. However, the modeling stage can also be
430 conceptually general enough to include any sort of analysis that converts phenotype measurements into a
431 new biological understanding.

432

433



434

435

436 **Figure 1.** A computation-based phenotyping project requires a software continuum that takes raw data
437 generated by acquisition activities, analyzes the raw data, integrates them with different data such as

438 genotype or environmental information, and then produces new understanding through modeling
 439 activities such as statistical associations. The new understanding leads to new questions.

440

441 Phenomic projects using Arabidopsis are ideal for training students in collaborative, innovative,
 442 and interdisciplinary approaches. Outreach and training modules on plant phenotyping naturally bridge
 443 multiple disciplines including plant biology, computer science, mathematics, and engineering, and provide
 444 alternative ways of attracting students to the plant sciences. Single-board computers like Raspberry Pi,
 445 Hummingboard, or Cubieboard are low-cost microcomputers originally built for educators, hobbyists and
 446 researchers, and are currently being incorporated into plant phenotyping research and teaching modules.
 447 Online resources provide tutorials to set up imaging systems (Table 2), however next-generation
 448 resources should be designed in collaboration with educational experts.

449

450 **Table 2: Online resources providing tutorials to setup imaging systems.**

Resource	Description	Website	Reference
Scikit-image examples and tutorials	Comprehensive list of imaging tasks with example code. Scikit-image is an imaging library for python	http://scikit-image.org/docs/dev/auto_examples/	(Van der Walt et al., 2014)
OpenCV tutorial	A collection of tutorial for openCV in C++. Open CV is a standard computer vision library available in C++, python and other languages	http://docs.opencv.org/2.4/doc/tutorials/tutorials.html	(Bradski and Kaehler, 2008)
Mahotas documentation	Mahotas is a python library written in C++. The documentation provides many examples for standard imaging tasks	http://mahotas.readthedocs.io/en/latest/	(Coelho, 2013)
DIRT tutorials and videos	DIRT (Digital Imaging of Root Traits) is an online root phenotyping platform that allows users to submit root images for phenotyping. The website contains tutorial and videos for non-technical users as well as documentation for developers. It's source code is freely available.	Online interface: http://dirt.iplantcollaborative.org/get-started Source Code: https://github.com/Computational-Plant-Science/DIRT	(Bucksch et al., 2014) (Das et al., 2015)

Phenotiki	Hardware (Raspberry Pi) and software for analyzing growth chamber collected phenotyping data	http://phenotiki.com/getting_started.html	(Minervini et al., 2014) (Giuffrida et al., 2015)
-----------	--	---	--

451

452 **Executive Summary**

453 Historically, the Arabidopsis research community has been able to effectively combine efforts
 454 internationally and to provide a collective voice regarding our needs to facilitate fundamental biological
 455 discoveries. We propose that such synergism be employed, using the specific recommendations in this
 456 commentary as a guide, in training this next generation of plant biologists to be able to understand and
 457 implement, in a rigorous manner, quantitative approaches in their research.

458 Specifically – for undergraduate and graduate training we recommend an overhaul in curriculum
 459 design for plant biology majors or plant biology graduate students that involves a seamless integration of
 460 concepts in math, physics, statistics and computation within courses that illustrate biological processes.
 461 This could be done according the recommendations of Bialek and Botsein (2004). We have adapted a set
 462 of core competencies and minimal skill sets, adapted from those of Tan et al. (2009), Rubinstein and Chor
 463 (2014), and Welch et al. (2014), and we strongly recommend that, when designing or revising curricula for
 464 this next generation of plant biologists, that these core competencies and skills are kept in mind. We
 465 have highlighted above a set of curricula based on these and which are publicly available either within the
 466 US or internationally; these may serve as a further resource. While there is no existing training standard
 467 for postdoctoral scholars in plant biology, we have identified a suite of fellowships for which postdocs
 468 may apply and which facilitate independent interdisciplinary training. We also advocate for programs
 469 which offer institutional support in identifying interdisciplinary and quantitative training for postdocs who
 470 wish to pursue such opportunities. Additional opportunities are outlined for faculty members who wish
 471 to undergo this training. Collaborations are often the cornerstone of successful quantitative projects and
 472 we provide concrete recommendations to promote effective and meaningful collaborations that we hope
 473 will guide institutional and cross-institutional interdisciplinary efforts. We collectively advocate for the
 474 continued use of Arabidopsis as an ideal organism for use in quantitative training efforts. For cases in
 475 which other organisms are more appropriate, we recommend leveraging best practices from the
 476 Arabidopsis community (e.g. efforts in genome annotation and data curation). Our case study in high-
 477 throughput Arabidopsis phenotyping provides an example of effective interdisciplinary and quantitative
 478 training and of the merging of quantitative and biological science integral for plant breeding in the future.

479

480 **Acknowledgments**

481 We would like to thank NSF-RCN 1518280 and NSF-RCN 1062348 for funding the workshop that led to
482 generation of this commentary.

483

484 **Literature Cited**

- 485 **Bansal S, Durrett TP** (2016) Camelina sativa: An ideal platform for the metabolic engineering
486 and field production of industrial lipids. *Biochimie* **120**: 9–16
- 487 **Bialek W, Botstein D** (2004) Introductory Science and Mathematics Education for 21st-Century
488 Biologists. *Science* **303**: 788–790
- 489 **Bradski G, Kaehler A** (2008) *Learning OpenCV*. O'Reilly
- 490 **Brutnell TP** (2015) Model grasses hold key to crop improvement. *Nat Plants* **1**: 15062
- 491 **Brutnell TP, Bennetzen JL, Vogel JP** (2015) Brachypodium distachyon and Setaria viridis:
492 Model Genetic Systems for the Grasses. *Annu Rev Plant Biol* **66**: 465–485
- 493 **Bucksch A, Burrridge J, York LM, Das A, Nord E, Weitz JS, Lynch JP** (2014) Image-Based
494 High-Throughput Field Phenotyping of Crop Roots. *Plant Physiol* **166**: 470–486
- 495 **Campbell MT, Knecht AC, Berger B, Brien CJ, Wang D, Walia H** (2015) Integrating Image-
496 Based Phenomics and Association Analysis to Dissect the Genetic Architecture of
497 Temporal Salinity Responses in Rice1[OPEN]. *Plant Physiol* **168**: 1476–1489
- 498 **Chen D, Neumann K, Friedel S, Kilian B, Chen M, Altmann T, Klukas C** (2014) Dissecting the
499 Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-
500 Throughput Image Analysis. *Plant Cell* **26**: 4636–4655
- 501 **Coelho LP** (2013) Mahotas: Open source software for scriptable computer vision. *J Open Res*
502 *Softw.* doi: 10.5334/jors.ac
- 503 **Cruz JA, Savage LJ, Zegarac R, Hall CC, Satoh-Cruz M, Davis GA, Kovac WK, Chen J,**
504 **Kramer DM** (2016) Dynamic Environmental Photosynthetic Imaging Reveals Emergent
505 Phenotypes. *Cell Syst* **2**: 365–377
- 506 **Das A, Schneider H, Burrridge J, Ascanio AKM, Wojciechowski T, Topp CN, Lynch JP,**
507 **Weitz JS, Bucksch A** (2015) Digital imaging of root traits (DIRT): a high-throughput
508 computing and collaboration platform for field-based root phenomics. *Plant Methods*. doi:
509 10.1186/s13007-015-0093-3
- 510 **Fahlgren N, Gehan MA, Baxter I** (2015) Lights, camera, action: high-throughput plant
511 phenotyping is ready for a close-up. *Curr Opin Plant Biol* **24**: 93–99
- 512 **Feldon DF, Jeong S, Peugh J, Roksa J, Maahs-Fladung C, Shenoy A, Oliva M** (2017) Null
513 effects of boot camps and short-format training for PhD students in life sciences. *Proc*
514 *Natl Acad Sci* **114**: 9854–9858
- 515 **Ge Y, Bai G, Stoerger V, Schnable JC** (2016) Temporal dynamics of maize plant growth, water
516 use, and leaf water content using automated high throughput RGB and hyperspectral
517 imaging. *Comput Electron Agric* **127**: 625–632
- 518 **Giuffrida MV, Minervini M, Tsafaris S** (2015) Learning to Count Leaves in Rosette Plants.
519 *British Machine Vision Association*, p 1.1-1.13
- 520 **Gutlerner JL, Van Vector D** (2013) Catalyzing Curriculum Evolution in Graduate Science
521 Education. *Cell* **153**: 731–736

- 522 **Holtan HE, Bandong S, Marion CM, Adam L, Tiwari S, Shen Y, Maloof JN, Maszle DR, Ohto**
523 **M, Preuss S, et al** (2011) BBX32, An Arabidopsis B-box Protein, Functions in Light
524 Signaling by Suppressing HY5-Regulated Gene Expression And Interacting with STH2.
525 *Plant Physiol* **156**: 2109-2123
- 526 **McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM** (2010) Systematic
527 discovery of nonobvious human disease models through orthologous phenotypes. *Proc*
528 *Natl Acad Sci U S A* **107**: 6544–6549
- 529 **Miller ND, Parks BM, Spalding EP** (2007) Computer-vision analysis of seedling responses to
530 light and gravity. *Plant J* **52**: 374–381
- 531 **Minervini M, Abdelsamea MM, Tsafaris SA** (2014) Image-based plant phenotyping with
532 incremental learning and active contours. *Ecol Inform* **23**: 35–48
- 533 **Oellrich A, Walls RL, Cannon EK, Cannon SB, Cooper L, Gardiner J, Gkoutos GV, Harper L,**
534 **He M, Hoehndorf R, et al** (2015) An ontology approach to comparative phenomics in
535 plants. *Plant Methods* **11**: 10
- 536 **Preuss SB, Meister R, Xu Q, Urwin CP, Tripodi FA, Screen SE, Anil VS, Zhu S, Morrell JA,**
537 **Liu G, et al** (2012) Expression of the Arabidopsis thaliana BBX32 Gene in Soybean
538 Increases Grain Yield. *PLOS ONE* **7**: e30717
- 539 **Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J,**
540 **Chapple C, Colot V, Cutler S, et al** (2016) 50 years of Arabidopsis research: highlights
541 and future directions. *New Phytol* **209**: 921–944
- 542 **Rellán-Álvarez R, Lobet G, Lindner H, Pradier P-L, Sebastian J, Yee M-C, Geng Y, Trontin**
543 **C, LaRue T, Schragger-Lavelle A, et al** (2015) GLO-Roots: an imaging platform enabling
544 multidimensional characterization of soil-grown root systems. *eLife*. doi:
545 10.7554/eLife.07597
- 546 **Rosenwald AG, Pauley MA, Welch L, Elgin SCR, Wright R, Blum J** (2016) The CourseSource
547 Bioinformatics Learning Framework. *CBE Life Sci Educ*. doi: 10.1187/cbe.15-10-0217
- 548 **Rubinstein A, Chor B** (2014) Computational Thinking in Life Science Education. *PLOS Comput*
549 *Biol* **10**: e1003897
- 550 **Searls DB** (2012) An Online Bioinformatics Curriculum. *PLoS Comput Biol*. doi:
551 10.1371/journal.pcbi.1002632
- 552 **Slovak R, Göschl C, Su X, Shimotani K, Shiina T, Busch W** (2014) A Scalable Open-Source
553 Pipeline for Large-Scale Root Phenotyping of Arabidopsis. *Plant Cell* **26**: 2390–2403
- 554 **Tan TW, Lim SJ, Khan AM, Ranganathan S** (2009) A proposed minimum skill set for university
555 graduates to meet the informatics needs and challenges of the “-omics” era. *BMC*
556 *Genomics* **10**: S36
- 557 **Walt S van der, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N,**
558 **Gouillart E, Yu T** (2014) scikit-image: image processing in Python. *PeerJ* **2**: e453
- 559 **Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, Schneider MV**
560 (2014) Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies.
561 *PLOS Comput Biol* **10**: e1003496

562 **Wingreen N, Botstein D** (2006) Back to the future: education for systems-level biologists. *Nat*
563 *Rev Mol Cell Biol* **7**: 829–832

564 **Zhu L-H, Krens F, Smith MA, Li X, Qi W, Loo EN van, Iven T, Feussner I, Nazareus TJ,**
565 **Huai D, et al** (2016) Dedicated Industrial Oilseed Crops as Metabolic Engineering
566 Platforms for Sustainable Industrial Feedstock Production. *Sci Rep* **6**: 22181

567

568