

Finding Out What's Going On with Phenology

Andre R. Dean^{1*}, Christina E. Lee², Elizabeth A. Berner^{3,4}, Salvatore S. Analdino⁵, Michael A. Berner⁶, James P. Balhoff⁷, David C. Blackburn⁸, Judith A. Blake⁹, J. Gordon Burleigh¹⁰, Bruno Chane¹¹, Laurel D. Cooper¹², Melanie Corcoran¹³, Sandor Cserkes¹⁴, Hong Gao¹⁵, Waila Dahal¹⁶, Sandip Das¹⁷, T. Alexander Dececchi¹⁶, Agnes DeGroot¹¹, Brian Diogo¹⁸, Robert E. Doolittle¹⁹, Michel Ducommunier²⁰, Nico M. Franke⁵, Frank Friedrich²¹, George V. Gkounis²², Melissa Haendel²³, Luke J. Harmon²⁴, Terrance F. Hamrick²⁵, Yongqiang He²⁶, Heather M. Hine¹, Nizar Ibrahim²⁷, Laura M. Jackson¹⁶, Pankaj Jaiswal¹², Christina James-Zorn²⁸, Sebastian Köhler²⁹, Guillaume Lecointre¹¹, Hilmar Lapp⁷, Carolin J. Lawrence³⁰, Nicola LeNoble³¹, John G. Lindberg³², James Macklin³³, Benjamin R. Mauck³⁴, Peter E. Midford³⁵, Iwan Miko¹, Christopher J. Mungall², Anika Oellrich³⁶, David Oumiherland³⁶, Helen Parkington³⁶, Martin J. Ramirez³⁷, Stefan Richter³⁸, Peter N. Robinson³⁹, Alan Rosenberg⁴⁰, Kaaja S. Schill⁴¹, Erik Segerdell⁴², Kaaja C. Selmann⁴³, Michael J. Sharke⁴⁴, Aaron D. Smith⁴⁵, Barr Smith⁴⁶, Chelsea D. Specht⁴⁷, R. Burke Stupple⁴⁸, Robert W. Thacker⁴⁹, Anne Theissen⁵⁰, Joëlle Fernande-Triana⁵¹, Mouna Vihinen⁵², Peter D. Vize⁵³, Larissa Vogt⁵⁴, Christine E. Wall⁵⁵, Ramona L. Wall⁵⁶, Monique Werfeld⁵⁷, Robert A. Wharton⁵⁸, Christian S. Wirkner³⁸, James B. Woolfe⁵⁸, Matthew J. Yoder⁵⁹, Aaron M. Zorn²⁸, Paula M. Mabee¹⁶

¹ Department of Entomology, Pennsylvania State University, University Park, Pennsylvania, United States of America, ² Genome Division, Lawrence Berkeley National Lab, Berkeley, California, United States of America, ³ Department of Plant Biology, Carnegie Institution for Science, Stanford, California, United States of America, ⁴ Phoenix Bioinformatics, Palo Alto, California, United States of America, ⁵ School of Life Sciences, Arizona State University, Tempe, Arizona, United States of America, ⁶ Department of Genetics, University of Cambridge, Cambridge, United Kingdom, ⁷ National Evolutionary Synthesis Center, Durham, North Carolina, United States of America, ⁸ Department of Vertebrate Zoology and Anthropology, California Academy of Sciences, San Francisco, California, United States of America, ⁹ The Jackson Laboratory, Bar Harbor, Maine, United States of America, ¹⁰ Department of Biology, University of Florida, Gainesville, Florida, United States of America, ¹¹ Muséum national d'Histoire naturelle, Département Systématique et Evolution, Paris, France, ¹² Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, United States of America, ¹³ Molecular Biology and Biochemistry Department, Simon Fraser University, Burnaby, British Columbia, Canada, ¹⁴ MTA-ELTE-MTM, Ecology Research Group, Pázmány Péter sétány 1C, Budapest, Hungary, ¹⁵ School of Information Resources and Library Science, University of Arizona, Tucson, Arizona, United States of America, ¹⁶ Department of Biology, University of South Dakota, Vermillion, South Dakota, United States of America, ¹⁷ Department of Botany, University of Delhi, Delhi, India, ¹⁸ Department of Anatomy, Howard University College of Medicine, Washington D.C., United States of America, ¹⁹ Department of Oral Biology, College of Dentistry, University of Illinois, Chicago, Illinois, United States of America, ²⁰ Stanford Center for Biomedical Informatics Research, Stanford, California, United States of America, ²¹ Biocenter Grindel and Zoological Museum, Hamburg University, Hamburg, Germany, ²² Department of Computer Science, Aberystwyth University, Aberystwyth, Ceredigion, United Kingdom, ²³ Department of Medical Informatics & Epidemiology, Oregon Health & Science University, Portland, Oregon, United States of America, ²⁴ Department of Biological Sciences, University of Idaho, Moscow, Idaho, United States of America, ²⁵ Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine, United States of America, ²⁶ Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, Center for Computational Medicine and Bioinformatics, and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, ²⁷ Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois, United States of America, ²⁸ Cincinnati Children's Hospital, Division of Developmental Biology, Cincinnati, Ohio, United States of America, ²⁹ Institute for Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany, ³⁰ Department of Genetics, Development and Cell Biology and Department of Agronomy, Iowa State University, Ames, Iowa, United States of America, ³¹ Signalling ISP, Babraham Institute, Babraham, Cambridgeshire, UK, ³² Department of Ichthyology, The Academy of Natural Science

Abstract: Despite a large and multifaceted effort to understand the vast landscape of phenotypic data, their current form inhibits productive data analysis. The lack of a community-wide, consensus-based, human- and machine-interpretable language for describing phenotypes and their genomic and environmental contexts is perhaps the most pressing scientific bottleneck to integration across many key fields in biology, including genomics, systems biology, development, medicine, evolution, ecology, and systematics. Here we survey the current phenomics landscape, including data resources and handling, and the progress that has been made to accurately capture relevant data descriptions for phenotypes. We present an example of the kind of integration across domains that computable phenotypes would enable, and we call upon the broader biology community, publishers, and relevant funding agencies to support efforts to surmount today's data barriers and facilitate analytical reproducibility.

Introduction

Phenotypic data, such as the vast landscape of phenotypic data, their current form inhibits productive data analysis. The lack of a community-wide, consensus-based, human- and machine-interpretable language for describing phenotypes and their genomic and environmental contexts is perhaps the most pressing scientific bottleneck to integration across many key fields in biology, including genomics, systems biology, development, medicine, evolution, ecology, and systematics. Here we survey the current phenomics landscape, including data resources and handling, and the progress that has been made to accurately capture relevant data descriptions for phenotypes. We present an example of the kind of integration across domains that computable phenotypes would enable, and we call upon the broader biology community, publishers, and relevant funding agencies to support efforts to surmount today's data barriers and facilitate analytical reproducibility.

Phenotypic data, such as the vast landscape of phenotypic data, their current form inhibits productive data analysis. The lack of a community-wide, consensus-based, human- and machine-interpretable language for describing phenotypes and their genomic and environmental contexts is perhaps the most pressing scientific bottleneck to integration across many key fields in biology, including genomics, systems biology, development, medicine, evolution, ecology, and systematics. Here we survey the current phenomics landscape, including data resources and handling, and the progress that has been made to accurately capture relevant data descriptions for phenotypes. We present an example of the kind of integration across domains that computable phenotypes would enable, and we call upon the broader biology community, publishers, and relevant funding agencies to support efforts to surmount today's data barriers and facilitate analytical reproducibility.

Fig. 1. How do we discover branching phenotype? (Bottom panel) Phenotype data exhibiting various forms of branchiness are not easily discerned from diverse natural language descriptions. (A) Bee hairs are different from most other insect hairs in that they are plumose, which facilitates pollen collection. (B) A mutant of *Drosophila* exhibits forked bristles, due to a variation in *scabrous*. (C) In zebrafish larvae (*Danio rerio*), angiogenesis begins with vessels branching. (D) Plant trichomes take on many forms, including trifurcation. (Top) Phenotypes involving some type of “branched” are easily recovered when they are represented with ontologies. In a semantic graph, free text descriptions are converted into phenotype statements involving an anatomy term from animal or plant ontologies [56,118] and a quality term from a quality ontology [106], connected by a logical expression (“inheres_in some”). Anatomy (purple) and quality (green) terms (ontology IDs beneath) relate phenotype statements from different species by virtue of the logic inherent in the ontologies, e.g., plumose, bifurcated, branched, and tripartite are all subtypes

$3 \frac{v}{y} w$ y -
 w y . w v v y y ,
 , y -
 y. ,
 y v , v -
 , v -
 y 40
 v v y j y -
 () 41 .
 y , w . -
 y w -
 v y -
 - v .

Achieving Data Integration

$\frac{w}{y}$ y
 w y
 -
 . y -
 (. 2) w - v y
 y y -

y v y y
y v , ,
. ,4A ,

7. (). //www. . . . y v. 26
2014.

8. . //www. . . / v . 26
2014.
. A (). //www.

77. v y w (185) L A (2007) L y 56 .
78. y v w (2011) L A (2012) A y 158 .
79. 283 2 4 (2011) A y 158 .
80. 1115 112 (2011) A y 158 .
81. L 6 17855 (2014) A y 158 .