Finding Our Way through Phenotypes

Andrew R Deans, Suzanna E Lewis, Eva Huala, Salvatore S Anzaldo, Michael Ashburner, James P Balhoff, David C Blackburn, Judith A Blake, J Gordon Burleigh, Bruno Chanet, et al.

OPEN ACCESS Freely available online

PLOS | BIOLOGY

Perspective

Finding Our Way through Phenotypes

Andrew R. Deans[1]*, Suzanna E. Lewis[2], Eva Huala[3,4], Salvatore S. Anzaldo[5], Michael Ashburner[6], James P. Balhoff[7], David C. Blackburn[8], Judith A. Blake[9], J. Gordon Burleigh[10], Bruno Chanet[11], Laurel D. Cooper[12], Mélanie Courtot[13], Sándor Csösz[14], Hong Cui[15], Wasila Dahdul[16], Sandip Das[17], T. Alexander Dececchi[16], Agnes Dettai[11], Rui Diogo[18], Robert E. Druzinsky[19], Michel Dumontier[20], Nico M. Franz[5], Frank Friedrich[21], George V. Gkoutos[22], Melissa Haendel[23], Luke J. Harmon[24], Terry F. Hayamizu[25], Yongqun He[26], Heather M. Hines[1], Nizar Ibrahim[27], Laura M. Jackson[16], Pankaj Jaiswal[12], Christina James-Zorn[28], Sebastian Köhler[29], Guillaume Lecointre[11], Hilmar Lapp[7], Carolyn J. Lawrence[30], Nicolas Le Novère[31], John G. Lundberg[32], James Macklin[33], Austin R. Mast[34], Peter E. Midford[35], István Mikó[1], Christopher J. Mungall[2], Anika Oellrich[36], David Osumi-Sutherland[36], Helen Parkinson[36], Martín J. Ramírez[37], Stefan Richter[38], Peter N. Robinson[39], Alan Ruttenberg[40], Katja S. Schulz[41], Erik Segerdell[42], Katja C. Seltmann[43], Michael J. Sharkey[44], Aaron D. Smith[45], Barry Smith[46], Chelsea D. Specht[47], R. Burke Squires[48], Robert W. Thacker[49], Anne Thessen[50], Jose Fernandez-Triana[51], Mauno Vihinen[52], Peter D. Vize[53], Lars Vogt[54], Christine E. Wall[55], Ramona L. Walls[56], Monte Westerfeld[57], Robert A. Wharton[58], Christian S. Wirkner[38], James B. Woolley[58], Matthew J. Yoder[59], Aaron M. Zorn[28], Paula M. Mabee[16]

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 Sciences, Philadelphia, Pennsylvania, United States of America, **33** Eastern Cereal and Oilseed Research Centre, Ottawa, Ontario, Canada, **34** Department of Biological Science, Florida State University, Tallahassee, Florida, United States of America, **35** Richmond, Virginia, United States of America, **36** European Molecular Biology Laboratory - European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **37** Division of Arachnology, Museo Argentino de Ciencias Naturales - CONICET, Buenos Aires, Argentina, **38** Allgemeine & Spezielle Zoologie, Institut für Biowissenschaften, Universität Rostock, Universitätsplatz 2, Rostock, Germany, **39** Institut für Medizinische Genetik und Humangenetik Charité – Universitätsmedizin Berlin, Berlin, Germany, **40** School of Dental Medicine, University at Buffalo, Buffalo, New York, United States of America, **41** Smithsonian Institution, National Museum of Natural History, Washington, D.C., United States of America, **42** Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon, United States of America, **43** Division of Invertebrate Zoology, American Museum of Natural History, New York, New York, United States of America, **44** Department of Entomology, University of Kentucky, Lexington, Kentucky, United States of America, **45** Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, United States of America, **46** Department of Philosophy, University at Buffalo, Buffalo, New York, United States of America, **47** Department of Plant and Microbial Biology, Integrative Biology, and the University and Jepson Herbaria, University of California, Berkeley, California, United States of America, **48** Bioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, **49** Department of Biology, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **50** The Data Detektiv, 1412 Stearns Hill Road, Waltham, Massachusetts, United States of America, **51** Canadian National Collection of Insects, Ottawa, Ontario, Canada, **52** Department of Experimental Medical Science, Lund University, Lund, Sweden, **53** Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada, **54** Universität Bonn, Institut für Evolutionsbiologie und Ökologie, Bonn, Germany, **55** Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, United States of America, **56** iPlant Collaborative University of Arizona, Thomas J. Keating Bioresearch Building, Tucson, Arizona, United States of America, **57** Institute of Neuroscience, University of Oregon, Eugene, Oregon, United States of America, **58** Department of Entomology, Texas A & M University, College, Station, Texas, United States of America, **59** Illinois Natural History Survey, University of Illinois, Champaign, Illinois, United States of America

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

Phenotypes, i.e., observable traits above the molecular level, such as anatomy and behavior, underlie, and indeed drive, much of the research in the life sciences. For example, they remain the primary data we use to define most species and to understand their phylogenetic history. Phenotype data are also used to recognize, define, and diagnose pathological conditions in plants, animals, and other organisms. As such, these data represent much of what we know of life and are, in fact, necessary for building a comprehensive tree of life [1]. Our observations of organismal phenotypes also inspire science aimed at understanding their development, functions, evolution, and interactions with the environment. Research in these realms, for example, has uncovered phenotypes that could be used to create antimicrobial materials [2] and efficient microrobots [3], yield novel approaches for drug delivery [4], treat the adverse effects of aging [5], and improve crop traits [6], among many other applications.

The Perspective section provides experts with a forum to comment on topical or controversial issues of broad interest.

Disease phenotypes, likewise, provoke us to research their genomic and environmental origins, often through manipulations of model organisms and/or by exploring the wild populations and ancestors, especially in the case of plants. The gamut of research on phenotype is very broad, but given the lack of computability across phenotype data (Fig. 1, bottom panel), there exists minimal cross-domain interaction. By not investing in the infrastructure needed to share phenotype data, we are missing opportunities for extraordinary discoveries.

Annotation strategies for genomes, in contrast to phenomes, are well advanced, with common methodologies, tools, syntaxes, and standards for articulating a precise description of nearly every type of genomic element [7–12]. Genomic data are also aggregated into large datasets, e.g., NCBI [7], EBI [8], DDBJ [9], and others [10–13]. Researchers lack these similarly well-established, linked, and consolidated resources for describing phenotypes and the contexts in which they arise, despite previous calls for more investment in this area [14–17]. Phenotype data (Table 1), although abundant and accumulating rapidly—e.g., species descriptions, image databases, analyses of induced variation, physiological measurements, whole genome knockout studies, high-throughput assays, electronic health records—are extremely heterogeneous, largely decentralized, and exist predominantly as free text. Thus, phenotype data are difficult to locate and impractical to interpret. In some areas of research, such as crop genetics and patient care, a great majority of the phenotype data underlying published research is not publicly available [18]. There also exists a divide between quantitative data and qualitative phenotype data, requiring reference measures or populations and statistical cutoffs to support interoperability (for example, "large head" versus a head circumference measurement). Finally, phenotypes change over time—be it evolutionary time, disease-course time, or developmental time—and the timing and ordering of phenotypic presentation is specific in any given context yet is rarely communicated. In short, while phenotype data are as complex, diverse, and nuanced as genomic data, they have not seen data standardization and analyses applied with the same broad strokes as we have seen for genomics.

Nevertheless, a small quantity of phenotype data, for a handful of species, is indeed formalized, such that it can be reliably searched, compared, and analyzed computationally (see below). However, with many disparate approaches to formalizing phenotypes, including different annotation strategies, the use of unrelated vocabularies, and the use of incomparable models and formats—these data are not fully unified or interoperable between taxa.

Given the latent potential of phenotype data and the emerging approaches to representing and computing across phenotypes, we members of the Phenotype Research Coordination Network (Phenotype RCN) [19], feel that the time is ripe for system-wide investment in the development of the needed tools and standards. As described in Box 1, many projects, sometimes working together but often independently, have begun building the foundation. There is now an opportunity for the large cross-domain phenomics research community to take advantage of new technologies for analyzing and managing the vast and diverse landscape of phenotype data, if attention and resources are applied to build in a consistent fashion on the current foundation.

Building a Phenomics Discovery Environment

How do we develop an environment in which researchers can readily make discoveries concerning the intimate connections among phenotypes, environment, and genetics? Three requirements must

Citation: Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, et al. (2015) Finding Our Way through Phenotypes. PLoS Biol 13(1): e1002033. doi:10.1371/journal.pbio.1002033

Published January 6, 2015

Funding: This effort was funded by the US National Science Foundation, grant number DEB-0956049. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: adeans@psu.edu

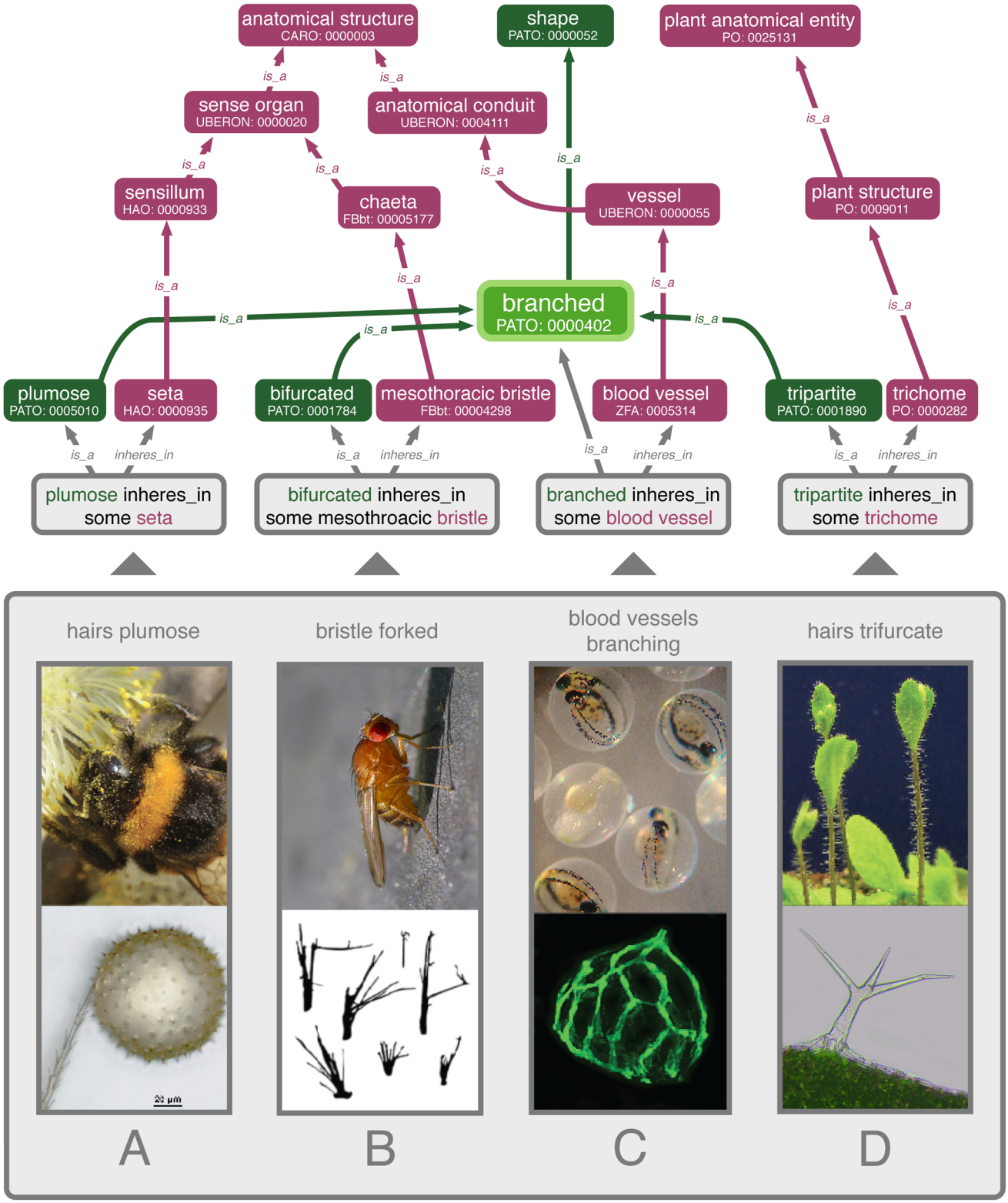

Fig. 1. How to discover branching phenotypes? (Bottom panel) Phenotype data exhibiting various forms of branchiness are not easily discerned from diverse natural language descriptions. (A) Bee hairs are different from most other insect hairs in that they are plumose, which facilitates pollen collection. (B) A mutant of *Drosophila melanogaster* exhibits forked bristles, due to a variation in *mical*. (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

of "branched." Image credits: bumble bee with pollen by Thomas Bresson, seta with pollen by István Mikó, *Arabidopsis* plants with hair-like structures (trichomes) by Annkatrin Rose, *Drosophila* photo by John Tann, *Drosophila* bristles redrawn from [119], scanning electron micrograph of *Arabidopsis* trichome by István Mikó, zebrafish embryos by MichianaSTEM, zebrafish blood vessels from [120]. Figure assembled by Anya Broverman-Wray.
doi:10.1371/journal.pbio.1002033.g001

be met for this vision to become a reality across large-scale data. First, phenotype descriptions must be rendered in a computable format, which usually involves the use of appropriate ontology terms (via Uniform Resource Identifiers [URIs]) to represent the phenotypic descriptions found in narrative text or data sources. Each bit of text is thereby imbued with properties and relationships to other terms (Fig. 1, top panel). Second, these semantically represented phenotype data, which integrate the phenotypes (Fig. 1, top panel) across species and also with their genetic and environmental contexts, must be stored in a way that is broadly accessible on the Internet in a nonproprietary format, e.g., in a Resource Description Framework (RDF). The third requirement is to grow a set of algorithms that enable users to analyze the data. That is, these algorithms combine the logical connections inherent in the ontologies with statistical analyses to, for example, identify similar phenotypes and their correlations with specific genetic or environmental factors.

Examples of systems that have the potential to transform their fields come from several domains. For instance, by computing from natural species phenotypes to the phenotypes resulting from gene disruption in model organisms, the Phenoscape project [20] demonstrated that genes underlying evolutionarily novel phenotypes can be proposed for experimental testing [21–23]. Uniting these previously unlinked data from evolutionary and biomedical domains provided a way to virtually automate the formulation of evolutionary developmental (evo-devo) hypotheses. The reinvention of descriptive taxonomy as a 21st century information science, likewise, requires computable phenotypic data and resources [24], including those for taxonomy [25] and for evolutionary biology [26–28]. This process is an active research focus of the Hymenopteran Anatomy Ontology project [29], which is developing computational methods to allow descriptions of species' phenotypes to be made in explicit and searchable forms [30,31]. Other successes have come from linking human disease phenotypes to annotated genetic data from model organisms, thus yielding insights into the genes involved in human disease [32,33]. Similarly, the Gramene project [34] developed the plant Trait Ontology (TO) to annotate the Quantitative Trait Locus (QTL) [35] for several crop plants, including rice, maize, and wheat.

Remarkably, and despite their significantly different aims, much of the phenotypic data that have been amassed through these projects can be made comparable—an outcome that until recently would have been impossible—because each of these groups shared common ontologies (i.e., semantics) and data annotation strategies. The systems they used are thus logically interoperable, and the bodies of phenotypic data emerging from their work can be compared and aggregated without further intervention. For these limited and domain-specific successes to be brought to bear more generally, approaches to ontology development and data annotation must be scaled up.

Several hurdles must be overcome. First, only a small fraction of the phenotypic diversity of life is currently represented in phenotype ontologies. Ontology development is time-consuming, requires expert knowledge and community buy-in, and is ideally paired with data-driven research that iteratively checks the soundness of the ontology as it simultaneously seeks discovery. New approaches are needed to expedite ontology development. Second, current methods of phenotypic data annotation are largely manual, thus requiring substantial resources for personnel to translate data from the published literature into a computable format. Semi-automated approaches for extracting phenotypes and other data from text [36–38] must be further developed. Though time-consuming, the transformation of legacy data in relation to these resources should be a one-time investment. It is only possible, however, if current and future projects co-develop and adopt common standards, and actively contribute to their ongoing development and maintenance,

Table 1. Finding phenotypes.

Phenotype data source	Characteristics	Example/Reference
published literature from biological and biomedical domains	highly dispersed corpus, mainly digitized, but still in natural language; contains abundant phenotypes	publisher websites, reviews that summarize important reference phenotype datasets [79,80]
supplementary data	spreadsheets, text files	publisher repositories, open repositories (e.g., Dryad [81])
trait databases and large corpora	relational databases containing free text phenotype descriptions	phenotype repositories specific to a particular field of study [82], Biodiversity Heritage Library [83], Encyclopedia of Life [62], Plant Trait Database [84], morphology databases [85–87]
images	annotated with keywords (free text); dispersed across many databases and repositories; phenotype or genotype data contained in these images are not computationally accessible [78].	biodiversity image stores [85–89], patient MRI images, X-rays, bright-field micrographs, image-bases of plant phenotypes [90]
natural history collections	>3,000,000,000 biological specimens worldwide, some with free text descriptions and associated images	iDigBio [91]
auto-generated data	quantitative data from satellite tracking devices, environmental sensors, and high-throughput phenotyping processes	National Ecological Observatory Network (NEON) [92], high throughput [26–28], tracking sensors [93]

The rich legacy of research in the life sciences includes a wealth of phenotype data contained in many sources, for millions of extinct and extant species. Some important sources of phenotypes date from more than 250 years ago [74–77]. With very few exceptions, phenotype data are not computationally accessible [78].
doi:10.1371/journal.pbio.1002033.t001

Box 1. Methodologies to Make Phenotypes Computable

The prospects of computable phenotype data have slowly improved over the past several years, with several domain-specific initiatives yielding results [21,30,32,94,95] and a larger framework of data integration resources [96–100]. These pioneering projects have achieved several goals: (i) more standardized measurements of complex phenotypes (e.g., PhenX [101]); (ii) an integrative phenotype semantic representation (in Web Ontology Language [OWL] [102]) and its use [103–105] to capture the genetic and environmental context of an observed phenotype [106]; (iii) an ontology of classes defining the anatomical, behavioral, and biological function terms and the relevant phenotypic qualities needed to describe phenotypes effectively in detail; and (iv) algorithms, such as OWLSim [107,108], combining the logical connections inherent in the ontologies with statistical analyses to identify phenotypes that are correlated with specific genetic makeups.

These tools have been used effectively in both the model organism biomedical and biodiversity domains, for example to discover new genes involved in gene networks underlying human disease [95,109–111], to prospect for candidate genes associated with crop improvement using Genome-Wide Association Studies (GWAS) experiments [112,113], to propose candidate genes for evolutionary novelties [21], to integrate and organize diverse functional data [114], to understand the characteristics used to diagnose species [30,31] and, when combined with systems biology data such as protein–protein interactions or pathway resources, to augment the analysis used in a clinical setting for diagnostics [95,115–117]. The use of computable phenotypes is expected to be a powerful approach to discovery of the genetic contribution to phenotypes, and it applies across all categories of genetic elements.

and if researchers avoid practices that may create errors [39] by writing their descriptions in ambiguous or locally idiosyncratic ways. Thus we must involve authors, editors, publishers, and funding agencies in the entire scholarly communication process in establishing the needed resources needed for data interoperability.

Predicting an individual organism's phenotypic characteristics based on the combination of its genetic heritage, development, and environmental context is a challenge for research at the intersection of the physical and life sciences [40] and is a driving force behind a major cyberinfrastructure investment by the United States National Science Foundation (NSF) [41]. With focused attention on the requirements for a phenomics-based system, we can expedite this goal. Integrating species phenotypes with data across all levels of the biological hierarchy is possible if strategies for data management are co-developed and coordinated.

Achieving Data Integration

Researchers who attempt to explore biological data using a multidisciplinary approach are aware that it is nearly impossible to integrate comparable data from multiple species and multiple publications. We manually assemble an example (Fig. 2) of how large-scale availability of logically structured phenotype descriptions could inform and relate disparate fields of research and help address this significant problem. Past efforts, however, have largely involved manual integration of limited datasets. In the future, the study of phenotypic causality will be increasingly reliant on large and rapidly growing data stores that can only be effectively searched with automated or semi-automated methods. At this juncture, discoveries in many areas of biology rely on integrating genomic data with phenotypic data, and such integration is at an impasse because of the lack of computable and accessible phenotypic data within and across species [42].

Linking Phenotypes to Genomic and Genetic Variation Data

Given that genomic data are now relatively inexpensive to collect (approximately US$5,000 per individual genome and rapidly approaching US$100 [43]), a growing number of independent projects are explicitly linking genetic variants to related phenotypes at costs upwards of US$1 million per species genome. For example, the NCBI databases [7,44] capture data concerning human variants related to disease using semantic terms [45–47]. Large-scale integration of such variants, including computable descriptions of disease phenotypes in humans, model and non-model organisms, are collected and semantically integrated to help support disease diagnosis and mechanism discovery by the Monarch Initiative [33]. The National Institutes of Health (NIH) Undiagnosed Disease Program [48] captures individual patient phenotype profiles using the Human Phenotype Ontology (HPO) and submits these phenotype data to the database of Genotypes and Phenotypes (dbGaP) [49] and to PhenomeCentral [50] to aid patient matching based on semantic comparisons. Multiple projects and institutions have collaborated to develop an approach for the capture of standardized human pathogen and vector sequencing metadata designed to support epidemiologic and genotype–phenotype association studies [51]. The NIH Knockout Mouse Phenotyping Program ($KOMP^2$) [52] and the International Mouse Phenotype Consortium (IMPC) [53] provide both their quantitative and qualitative phenotype assay data for the mouse using the Mammalian Phenotype Ontology (MP) [54]. Both HP and MP classes (i.e., descriptive terms) are linked to upper-level classes in the UBERON anatomy ontology [55,56]. Thus, the phenotypes and associated variations from these autonomous projects can be compared automatically, as evident in cross-species resources such as PhenomeNET [57] and others [58,59]. Similarly, the Gramene project [34] developed the plant Trait Ontology (TO) to annotate the Quantitative Trait Locus (QTL) [35] for several crop plants, including rice, maize, and wheat. As noted above, however, the paths between genotype and phenotype are not one-to-one. Any successful strategy must also account for environmental contributions, and, as with phenotypes and genotypes, a well-structured, consistent means of describing environmental differences is essential.

Linking Phenotypes to Environment

An organism's phenotypes result from the interplay of environment with genetics and developmental processes. The meaning of "environment" differs according to biological context. For biodiversity, environment refers to the specific conditions and geographical location in which any given organism is found. For model organisms, environment comprises the experimental perturbations relative to what is "normal" for an organism of that time, for example, changes in exposure to a drug or in the concentration of salt in the water that serves as an organism's home. For epidemiological studies, environment may refer to features in the physical proximity, such as to a nuclear plant, or relate to prior personal behavior, such as a history of smoking. Although the pheno-

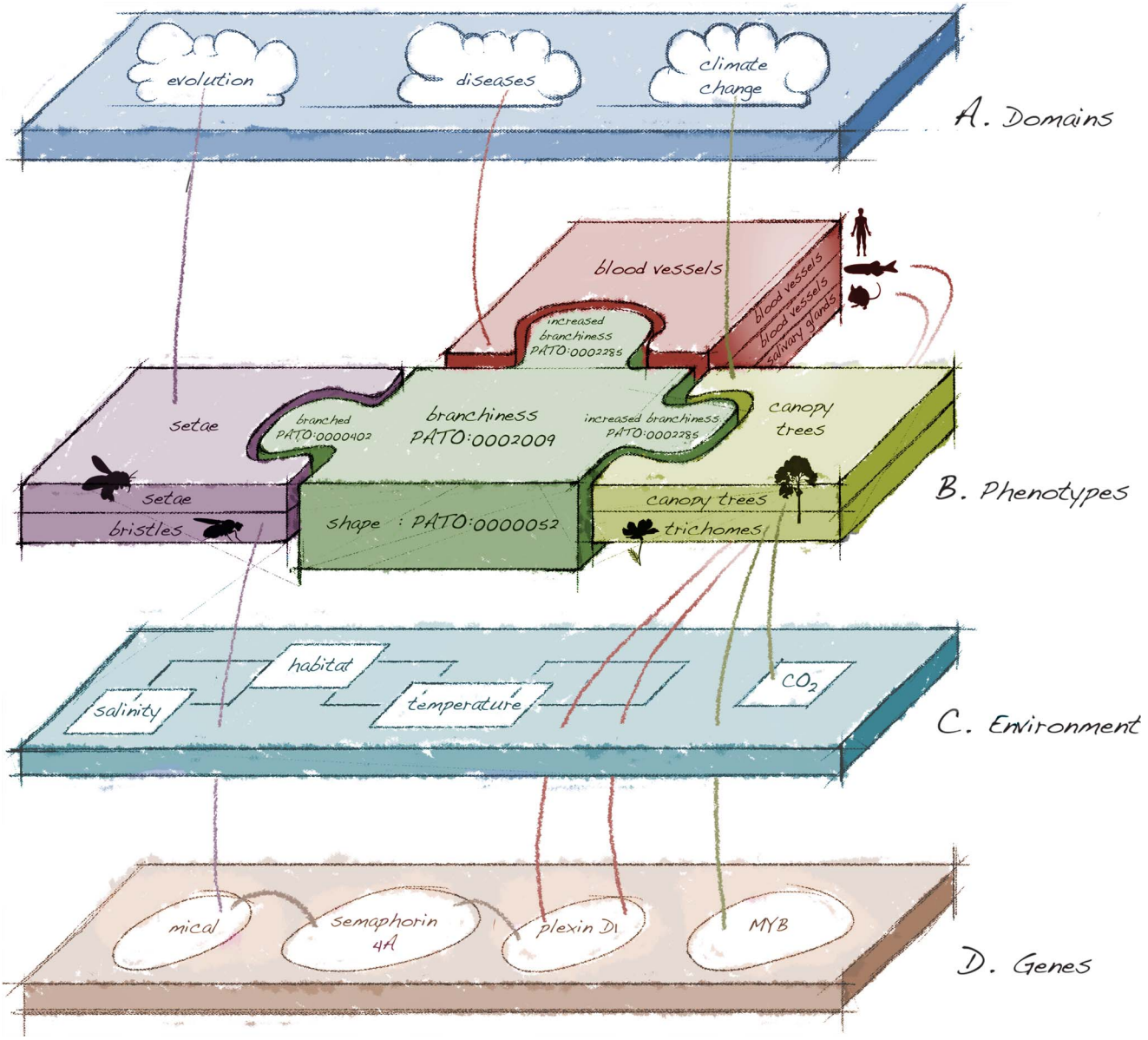

Fig. 2. Phenotypes shared across biology. Phenotype data are relevant to many different domains, but they are currently isolated in data "silos." Research from a broad array of seemingly disconnected domains, as outlined here, can be dramatically accelerated with a computable data store. (**A**) **Domains**: Diverse fields such as evolutionary biology, human disease and medicine, and climate change relate to phenotypes. (**B**) **Phenotypes**: insects, vertebrates, plants, and even forests all have features that are branched in some way, but they are described using different terms. For a computer to discover this, the phenotypes must be annotated with unique identifiers from ontologies that are logically linked. Under "shape" in the PATO quality ontology [106], "branchiness" is an encompassing parent term with subtypes "branched" and "increased branchiness." From left to right, top layer, insects, vertebrates and plants have species that demonstrate phenotypes for which the genetic basis is not known. Often their companion model species, however, have experimental genetic work that is relevant to proposing candidate genes and gene networks. Insects (1): An evolutionary novelty in bees (top layer) is the presence of branched setae used for pollen collection. Nothing is known about the genetic basis of this feature. One clue to the origin of this evolutionary feature comes from studies of *Drosophila* (bottom layer), where *Mical* overexpression in unbranched wild-type bristles generates a branched morphology [119]. Mical directly links semaphorins and their plexin receptors to the precise control of actin filament dynamics [119]. Vertebrates (2): In humans, aberrant angiogenesis, including excessive blood vessel branching (top layer), is one of the six central hallmarks of cancer [121]. Candidate genes have been identified using data from model organisms. In zebrafish (middle layer), studies of the control of sprouting in blood vessel development show that signaling via semaphorins [122] and their plexin receptors is required for proper abundance and distribution [123]; disruption of *plxnd1* results in increased branching [120,124,125]. In mouse (bottom layer), branching of salivary glands is dependent on semaphorin signaling [126], as is the branching of various other epithelial organs [127]. Plants (3): The uppermost canopy of trees of the rainforest (top layer) undergo a marked increase in branching associated with climate change [128]. Nothing is known about the genetic basis of this feature. The branching of plant trichomes (bottom layer), tiny outgrowths with a variety of functions including seed dispersal, has been studied in the model *Arabidopsis thaliana.* Branching occurs in association with many MYB-domain genes [129], transcription factors that are found in both plants and animals [130]. (**C**) **Environment**: Diverse input from the environment influences organismal phenotype. (**D**) **Genes**: At the genetic level, previously unknown associations with various types of "branchiness" between insects and vertebrates are here made to possibly a common core or network of genes (the semaphorin-plexin signaling network). No association between genes associated with plant branching (Myb transcription factors) and animal branching is obvious from the literature. Image credit: Anya Broverman-Wray.
doi:10.1371/journal.pbio.1002033.g002

type data collected in these different types of environments may at first glance seem mutually irrelevant, there is, in fact, often a need to combine them. Exposure to an environmental toxin, for example, could similarly affect the phenotype of local flora and fauna populations and of human patients, and it could be related to phenotypic outcomes identified via experiments involving perturbation of the environments of model organisms. Neither environment nor phenotype is a static entity; both change over developmental and evolutionary time [15,16]. Very few efforts have attempted to relate phenotypic data captured in these varied contexts, in part due to the vastly different mechanisms by which the environmental variables and measures are described.

Building blocks to capture these pieces include the Environment Ontology (EnvO) [60] and the Exposure Science Ontology (ExO) [61], which provide controlled, structured vocabularies designed to enable representation of the relationships between organisms and biological samples to their environment. EnvO has been used by projects as disparate as the Encyclopedia of Life [62] and the International Census of Marine Microbes [63]. It is also one of the ontologies incorporated into the Experimental Factor Ontology (EFO) [64] used for systematic description of experimental variables available in European Bioinformatics Institute (EBI) databases [8] and for National Human Genome Research Institute's catalog of published GWAS [65]. Ontologies and associated tools provide a powerful, rational means for discovering connections between data from multiple projects. This potential can only be realized by reusing and combining classes from core primary ontologies. This is the strategy used by numerous successful cases, such as the EFO's incorporation of EnvO and other ontologies, and has dual benefits. It allows projects to tailor their ontology to suit their own particular needs, while retaining the powerful capability to semantically integrate their data with data from multiple other projects. This approach brings convergence, avoids duplication of effort and enables joint analysis of combined data.

Remarkable advances are being made in measuring environmental data, ranging from fine-scale measurements across the surface of a leaf to variation across a planted field to high-resolution environmental layers at a global scale (e.g., [66,67]). As environmental data rapidly accumulate as a result of these new technologies, now is an opportune moment to ensure the usability and longevity of these data by adopting systematic standards. Towards this end, recent workshops funded by NSF [68] and National Institute of Environmental Health Sciences (NIEHS) [69] brought together diverse sets of experts to aid in developing vocabularies and standards for describing environment.

Recommendations

Recommendation 1

We urge all biologists, data managers, and clinicians to actively support the development, evaluation, refinement, and adoption of methodologies, tools, syntaxes, and standards for capturing and computing over phenotypic data and to collaborate in bringing about a coordinated approach. And we urge university lecturers to introduce their students to these tools and concepts and integrate them into the standard basic curriculum in all relevant fields. The resultant increase in interoperability will enhance broad access to large stores of phenotypic data required or already existing across many areas of biology. It will accelerate discoveries across biological domains and increase significantly the return on the huge past and present investment made to generate the data. Although there are daunting challenges to this critical and enormous undertaking, its success will increase efficiency, greatly reduce the loss of data and duplication of effort, and facilitate reuse of phenotype data [70].

Recommendation 2

We urge publishers to require contribution of structured phenotype data in semantic-enabled ways as the technology is developed, to enable us to compute beyond the impasse of the free-text narrative. Moreover, funding agencies should request appropriate metadata for phenotypic descriptions, and they should require that all phenotypic screening made with their funds result in open and interoperable data.

Recommendation 3

With the community, conceptual, and methodological framework falling into place, the next steps require a new set of resources for phenotypes, including tools for the conversion of important legacy phenotype datasets to the newly established computable formats, putting into place mechanisms to scale up acquisition of new phenotypes, methods that ensure appropriate mark-up and deposition of phenotypic data upon publication [71], organization of the data into accessible online resources, new tools to visualize and analyze the data, and the development of a comprehensive cross-species and cross-domain phenotypic resource.

These needs are urgent and reach across the research spectrum, from understanding biodiversity loss and decline, to interpreting genomes of the new "non-model" systems that are coming online, to elevating the health of the expanding human population. The use of computable phenotypes is expected to be a powerful approach to discovery of the genetic contribution to phenotypes [72,73], and it applies across all categories of genetic elements.

Science revolves around gathering facts and making theories, a repeating cycle of improvement and increasing knowledge. In the history of science, the iterative accumulation of well-integrated facts—starting with the creation of a common system of units—has over and over again determined accelerated growth in scientific understanding. As our base of phenotypic knowledge grows ever larger, it will only become ever more difficult to navigate and comprehend, without the coordinated improvements in infrastructure and culture that will expedite scientific discovery.

Acknowledgments

We thank Anya Broverman-Wray for her expert preparation of Figs. 1 and 2 and the photographers who availed their images for Fig. 1.

References

1. Burleigh JG, Alphonse K, Alverson AJ, Bik HM, Blank C, et al. (2013) Next-generation phenomics for the Tree of Life. PLoS Currents 5.
2. Pogodin S, Hasan J, Baulin VA, Webb HK, Truong VK, et al. (2013) Biophysical model of bacterial cell interactions with nanopatterned cicada wing surfaces. Biophys J 104: 835–840.
3. Ma KY, Chirarattananon P, Fuller SB, Wood RJ (2013) Controlled flight of a biologically inspired, insect-scale robot. Science 340: 603–607.
4. Aboelkassem Y, Staples AE (2013) Selective pumping in a network: insect-style microscale flow transport. Bioinspir Biomim 8: 026004.
5. Salmon AB, Leonard S, Masamsetti V, Pierce A, Podlutsky AJ, et al. (2009) The long lifespan of two bat species is correlated with resistance to protein oxidation and enhanced protein homeostasis. FASEB J 23: 2317–2326.
6. National Plant Genome Initiative: 2009–2013. http://www.nsf.gov/bio/pubs/reports/npgi_five_year_plan_2009_2013.pdf. 26 June 2014.

7. National Center for Biotechnology Information (NCBI). http://www.ncbi.nlm.nih.gov. 26 June 2014.
8. European Bioinformatics Institute (EBI) databases. http://www.ebi.ac.uk/services. 26 June 2014.
9. DNA Data Bank of Japan (DDBJ). http://www.ddbj.nig.ac.jp/. 26 June 2014.
10. Ensembl Plants. http://plants.ensembl.org. 26 June 2014.
11. Phytozome. http://www.phytozome.net. 26 June 2014.
12. European Nurserystock Association (ENA). http://www.enaplants.eu/. 26 June 2014.
13. GigaDB. http://gigadb.org. 26 June 2014.
14. NSF (2011) Phenomics: Genotype to Phenotype. A report of the Phenomics workshop sponsored by the USDA and NSF 2011 National Science Foundation. http://www.nsf.gov/bio/pubs/reports/phenomics_workshop_report.pdf
15. Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. Nat Rev Genet 11: 855–866.
16. Houle D (2010) Colloquium papers: Numbering the hairs on our heads: the shared challenge and promise of phenomics. Proc Natl Acad Sci U S A 107 Suppl 1: 1793–1799.
17. Trelease RB (2006) Anatomical reasoning in the informatics age: Principles, ontologies, and agendas. The Anatomical Record Part B: The New Anatomist 289B: 72–84.
18. Zamir D (2013) Where have all the crop phenotypes gone? PLoS Biol 11: e1001595.
19. Phenotype Research Coordination Network (Phenotype RCN). http://www.phenotypercn.org. 26 June 2014.
20. Phenoscape Knowledgebase. kb.phenoscape.org. 12 Aug 2014.
21. Mabee P, Balhoff JP, Dahdul WM, Lapp H, Midford PE, et al. (2012) 500,000 fish phenotypes: The new informatics landscape for evolutionary and developmental biology of the vertebrate skeleton. J Appl Ichthyol 28: 300–305.
22. Dahdul WM, Balhoff JP, Engeman J, Grande T, Hilton EJ, et al. (2010) Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. PLoS One 5: e10708.
23. Balhoff JP, Dahdul WM, Kothari CR, Lapp H, Lundberg JG, et al. (2010) Phenex: ontological annotation of phenotypic diversity. PLoS One 5: e10500.
24. Deans AR, Yoder MJ, Balhoff JP (2012) Time to change how we describe biodiversity. Trends Ecol Evol 27: 78–84.
25. Franz NM, Thau D (2010) Biological taxonomy and ontology development: scope and limitations. Biodiversity Informatics 7: 45–66.
26. Ramírez MJ, Michalik P (2014) Calculating structural complexity in phylogenies using ancestral ontologies. Cladistics. doi: 10.1111/cla.12075.
27. Richter S, Wirkner CS (2014) A research program for Evolutionary Morphology. Journal of Zoological Systematics and Evolutionary Research 52: 338–350.
28. Wirkner C, Richter S (2010) Evolutionary morphology of the circulatory system in Peracarida (Malacostraca; Crustacea). Cladistics 26: 143–167.
29. Yoder MJ, Miko I, Seltmann KC, Bertone MA, Deans AR (2010) A gross anatomy ontology for Hymenoptera. PLoS One 5: e15991.
30. Balhoff JP, Mikó I, Yoder MJ, Mullins PL, Deans AR (2013) A semantic model for species description, applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia. Syst Biol 62: 639–659.
31. Mikó I, Copeland R, Balhoff J, Yoder M, Deans A (2014) Folding wings like a cockroach: a review of transverse wing folding ensign wasps (Hymenoptera: Evaniidae: *Afrevania* and *Trissevania*). PLoS ONE 9: e94056.
32. Washington NL, Haendel Ma, Mungall CJ, Ashburner M, Westerfield M, et al. (2009) Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol 7: e1000247.
33. Monarch Initiative. [http://monarch.monarchinitiative.org] 11 Aug 2014.
34. Youens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, et al. (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39: D1085–1094.
35. Ni J, Pujar A, Youens-Clark K, Yap I, Jaiswal P, et al. (2009) Gramene QTL database: development, content and applications. Database (Oxford) 2009: bap005.
36. Arighi CN, Carterette B, Cohen KB, Krallinger M, Wilbur WJ, et al. (2013) An overview of the BioCreative 2012 Workshop Track III: interactive text mining task. Database (Oxford) 2013: bas056.
37. Cui H (2012) CharaParser for fine-grained semantic annotation of organism morphological descriptions. Journal of American Society of Information Science and Technology 63: 738–754.
38. Thessen AE, Parr CS (2014) Knowledge extraction and semantic annotation of text from the encyclopedia of life. PLoS One 9: e89550.
39. Markov G, Lecointre G, Demeneix B, Laudet V (2008) The "street light syndrome", or how protein taxonomy can bias experimental manipulations. Bioessays 30: 349–357.
40. NRC (National Research Council U (2010) Research at the Intersection of the Physical and Life Sciences: Grand Challenges. Washington (DC): National Academies Press (US)
41. Genomes - Phenomes Grand Challenge. 26 June 2014. https://extwiki.nsf.gov/display/gpgc/Genomes+-+Phenomes+Grand+Challenge+Home
42. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, et al. (2012) Toward interoperable bioscience data. Nat Genet 44: 121–126.
43. Cost per Genome. http://www.genome.gov/images/content/cost_per_genome.jpg. 26 June 2014.
44. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42: D980–985.
45. MedGen. http://www.ncbi.nlm.nih.gov/medgen. 26 June 2014.
46. Unified Medical Language System (UMLS) http://www.nlm.nih.gov/research/umls/26 June 2014.
47. Robinson PN, Mundlos S (2010) The Human Phenotype Ontology. Clin Genet 77: 525–534.
48. National Institutes of Health (NIH) Undiagnosed Disease Program. http://rarediseases.info.nih.gov/research/pages/27/undiagnosed-diseases-program 26 June 2014.
49. The database of Genotypes and Phenotypes (dbGaP). http://www.ncbi.nlm.nih.gov/gap 26 June 2014.
50. PhenomeCentral. https://phenomecentral.org. 26 June 2014.
51. Dugan VG, Emrich SJ, Giraldo-Calderon GI, Harb OS, Newman RM, et al. (2014) Standardized metadata for human pathogen/vector genomic sequences. PLoS One 9: e99979.
52. Knockout Mouse Phenotyping Project (KOMP2) http://jaxmice.jax.org/news/2013/KOMP_article_3.html 26 June 2014.
53. International Mouse Phenotyping Consortium (IMPC) http://www.mousephenotype.org. 26 June 2014.
54. Smith CL, Eppig JT (2009) The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip Rev Syst Biol Med 1: 390–399.
55. Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, Haendel MA (2012) Uberon, an integrative multi-species anatomy ontology. Genome Biology 13: R5.
56. Haendel MA, Ballhoff JP, Bastian FB, Blackburn DC, Blake JA, et al. (2014) Uberon: Unification of multi-species vertebrate anatomy ontologies for comparative biology. J Biomed Semantics 5: 21.
57. Hoehndorf R, Schofield PN, Gkoutos GV (2011) PhenomeNET: a whole-phenome approach to disease gene discovery. Nucleic Acids Res 39: e119.
58. Köhler S, Doelken SC, Ruef BJ, Bauer S, Washington N, et al. (2013) Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Res 2: 30.
59. Monarch Initiative. http://monarchinitiative.org/26 June 2014.
60. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, et al. (2013) The environment ontology: contextualising biological and biomedical entities. J Biomed Semantics 4: 43.
61. Mattingly CJ, McKone TE, Callahan MA, Blake JA, Hubal EA (2012) Providing the missing link: the exposure science ontology ExO. Environ Sci Technol 46: 3046–3053.
62. Encyclopedia of Life. http://eol.org/26 June 2014.
63. International Census of Marine Microbes http://icomm.mbl.edu/26 June 2014.
64. Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, Shojatalab M, et al. (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res 37: D868–872.
65. National Human Genome Research Institute's catalog of published Genome-Wide Association Studies http://www.genome.gov/gwastudies/ 26 June 2014.
66. Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, et al. (2013) High-resolution global maps of 21st-century forest cover change. Science 342: 850–853.
67. National Ecological Observatory Network (NEON). http://www.neoninc.org/31 October 2014.
68. Phenotype RCN: Environment and Phenotype meeting. https://www.niehs.nih.gov/homePage/slideshow/september_15_2014_workshop_of_the_development_of_a_framework_for_environmental_health_science_language_508.pdf http://www.phenotypercn.org/?page_id=2287. 15 October 2014.
69. National Institute of Environmental Health Sciences: Workshop for the Development of a Framework for Environmental Health Science Language. https://http://www.niehs.nih.gov/homePage/slideshow september_15_2014_workshop_of_the_development_of_a_framework_for_environmental_health_science_language_508.pdf. 19 October 2014.
70. Vogt L, Nickel M, Jenner RA, Deans AR (2013) The need for data standards in zoomorphology. J Morphol 274: 793–808.
71. Piwowar HA, Vision TJ, Whitlock MC (2011) Data archiving is a good investment. Nature 473: 285.
72. Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, et al. (2012) A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Reports 2: 817–823.
73. Rowan Ba, Weigel D, Koenig D (2011) Developmental genetics and new sequencing technologies: the rise of nonmodel organisms. Dev Cell 21: 65–76.
74. Aristotle, Balme DM, Gotthelf A (2002) Aristotle: 'Historia Animalium': Volume 1, Books I-X: Text: Cambridge University Press.
75. von Baer KE (1828) Über die Entwicklungsgeschichte der Thiere. Königsberg: Bornträger
76. Owen R (1849) On the Nature of Limbs: A Discourse. In: Amundson R, editor. On the

Nature of Limbs: A Discourse. Chicago: University of Chicago Press.
77. Darwin C (1859) On the origin of species. Cambridge: Harvard University Press.
78. Ramírez M, Coddington J, Maddison W, Midford P, Prendini L, et al. (2007) Linking of digital images to phylogenetic data matrices using a morphological ontology. Syst Biol 56: 283–294.
79. Lloyd J, Meinke D (2012) A comprehensive dataset of genes with a loss-of-function mutant phenotype in *Arabidopsis*. Plant Physiol 158: 1115–1129.
80. Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One 6: e17855.
81. Dryad Digital Repository. http://datadryad.org/. 26 June 2014.
82. Ephesis: Environment and Phenotype Information System. https://urgi.versailles.inra.fr/Projects/URGI-softwares/Ephesis. 26 June 2014.
83. Biodiversity Heritage Library (BHL). http://www.biodiversitylibrary.org. 26 June 2014.
84. Plant Trait Database. http://www.try-db.org/. 26 June 2014.
85. Morphbank:: Biological Imaging Florida State University, Department of Scientific Computing, Tallahassee, FL 32306-4026 USA. http://www.morphbank.net/. 26 June 2014.
86. O'Leary MA, Kaufman S (2011) MorphoBank: phylophenomics in the "cloud". Cladistics 27: 529–537.
87. Morph.D.Base 2.0: A public data base for morphological data, metadata, and phylogenetic matrices. http://www.morphdbase.de. 26 June 2014.
88. Berquist RM, Gledhill KM, Peterson MW, Doan AH, Baxter GT, et al. (2012) The Digital Fish Library: using MRI to digitize, database, and document the morphological diversity of fish. PLoS One 7: e34499.
89. DigiMorph: Digital Morphology at the University of Texas at Austin. http://www.digimorph.org. 26 June 2014.
90. Australian Phenomics Facility. http://apf.anu.edu.au. 26 June 2014.
91. Integrated Digitized Biocollections (iDigBio). http://www.idigbio.org/. 27 June 2014.
92. National Ecological Observatory Network (NEON) http://www.neoninc.org. 20 October 2014.
93. Greene CH, Block BA, Welch D, Jackson G, Lawson GL (2009) Advances in conservation oceanography: new tagging and tracking technologies and their potential for transforming the science underlying fisheries management. Oceanography 22 210–223.
94. Oellrich A, Hoehndorf R, Gkoutos GV, Rebholz-Schuhmann D (2012) Improving disease gene prioritization by comparing the semantic similarity of phenotypes in mice with those of human diseases. PloS One 7: e38937.
95. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, et al. (2014) The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 42: D966–974.
96. Haendel MA, Neuhaus F, Osumi-Sutherland DS, Mabee PM, Mejino JLV, et al. (2008) CARO – The Common Anatomy Reference Ontology. In: Burger A, Davidson D, Baldock R, editors. Anatomy Ontologies for Bioinformatics: Principles and Practice. Springer. pp. 327–349.
97. The Ontology for Biomedical Investigations (OBI). http://obi-ontology.org/. 26 June 2014.
98. Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, et al. (2010) Modeling biomedical experimental processes with OBI. J Biomed Semantics 1 Suppl 1: S7.
99. Smith B, Ashburner M, Rosse C, Bard J, Bug W, et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25: 1251–1255.
100. Ontology Alignment Evaluation Initiative. http://oaei.ontologymatching.org. 26 June 2014.
101. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, et al. (2011) The PhenX Toolkit: get the most from your measures. Am J Epidemiol 174: 253–260.
102. Web Ontology Language (OWL) http://www.w3.org/TR/owl-features/26 June 2014.
103. Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, et al. (2007) Advancing translational research with the Semantic Web. BMC Bioinformatics 8 Suppl 3: S2.
104. Ruttenberg A, Rees JA, Samwald M, Marshall MS (2009) Life sciences on the Semantic Web: the Neurocommons and beyond. Brief Bioinform 10: 193–204.
105. OBO Foundry Identifier Policy. http://www.obofoundry.org/id-policy.shtml. 26 June 2014.
106. Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D (2005) Using ontologies to describe mouse phenotypes. Genome Biol 6: R8.
107. Chen CK, Mungall CJ, Gkoutos GV, Doelken SC, Kohler S, et al. (2012) MouseFinder: Candidate disease genes from mouse phenotype data. Hum Mutat 33: 858–866.
108. Smedley D, Oellrich A, Kohler S, Ruef B, Sanger Mouse Genetics P, et al. (2013) PhenoDigm: analyzing curated annotations to associate animal models with human diseases. Database (Oxford) 2013: bat025.
109. Gkoutos GV, Mungall C, Dolken S, Ashburner M, Lewis S, et al. (2009) Entity/quality-based logical definitions for the human skeletal phenome using PATO. Conf Proc IEEE Eng Med Biol Soc 2009: 7069–7072.
110. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, et al. (2010) Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc Natl Acad Sci U S A 107: 6544–6549.
111. Robinson PN, Webber C (2014) Phenotype Ontologies and Cross-Species Analysis for Translational Research. PLoS Genet 10: e1004268.
112. Huang X, Wei X, Sang T, Zhao Q, Feng Q, et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42: 961–967.
113. Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, et al. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158: 824–834.
114. Wall CE, Vinyard CJ, Williams SH, Gapeyev V, Liu X, et al. (2011) Overview of FEED, the feeding experiments end-user database. Integr Comp Biol 51: 215–223.
115. Robinson PN, Köhler S, Oellrich A, Project SMG, Wang K, et al. (2014) Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res 24: 340–348.
116. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, et al. (2008) The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 83: 610–615.
117. Zemojtel T, Kohler S, Mackenroth L, Jager M, Hecht J, et al. (2014) Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med 6: 252ra123.
118. Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, et al. (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. Plant Cell Physiol 54: e1.
119. Hung RJ, Yazdani U, Yoon J, Wu H, Yang T, et al. (2010) Mical links semaphorins to F-actin disassembly. Nature 463: 823-827.
120. Alvarez Y, Astudillo O, Jensen L, Reynolds AL, Waghorne N, et al. (2009) Selective inhibition of retinal angiogenesis by targeting PI3 kinase. PLoS One 4: e7867.
121. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646-674.
122. Yazdani U, Terman JR (2006) The semaphorins. Genome Biol 7: 211.
123. Gu C, Giraudo E (2013) The role of semaphorins and their receptors in vascular development and cancer. Exp Cell Res 319: 1306–1316.
124. Zygmunt T, Gay CM, Blondelle J, Singh MK, Flaherty KM, et al. (2011) Semaphorin-PlexinD1 signaling limits angiogenic potential via the VEGF decoy receptor sFlt1. Dev Cell 21: 301–314.
125. Torres-Vazquez J, Gitler AD, Fraser SD, Berk JD, Van NP, et al. (2004) Semaphorin-plexin signaling guides patterning of the developing vasculature. Dev Cell 7: 117–123.
126. Chung L, Yang TL, Huang HR, Hsu SM, Cheng HJ, et al. (2007) Semaphorin signaling facilitates cleft formation in the developing salivary gland. Development 134: 2935–2945.
127. Korostylev A, Worzfeld T, Deng S, Friedel RH, Swiercz JM, et al. (2008) A functional role for semaphorin 4D/plexin B1 interactions in epithelial branching morphogenesis during organogenesis. Development 135: 3333–3343.
128. Niinemets U (2010) A review of light interception in plant stands from leaf to canopy in different plant functional types and in species with varying shade tolerance. Ecol Res 25: 693–714.
129. Serna L, Martin C (2006) Trichomes: different regulatory networks lead to convergent structures. Trends Plant Sci 11: 274–280.
130. Rosinski JA, Atchley WR (1998) Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. J Mol Evol 46: 74–83.