Opinion TRENDS in Ecology and Evolution Vol.xxx No.x

Full text provided by www.sciencedirect.com ScienceDirect

Phenotype ontologies: the bridge between genomics and evolution

Paula M. Mabee[1], Michael Ashburner[2,3], Quentin Cronk[4], Georgios V. Gkoutos[2], Melissa Haendel[5], Erik Segerdell[5,6], Chris Mungall[7] and Monte Westerfield[5,8]

[1] Department of Biology, University of South Dakota, Vermillion, SD 57069, USA
[2] Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK
[3] EMBL-European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK
[4] UBC Botanical Garden and Centre for Plant Research, University of British Columbia, 6804 SW Marine Drive, Vancouver, BC, V6T 1Z4, Canada
[5] Zebrafish Information Network (ZFIN), University of Oregon, Eugene, OR 97403-5291, USA
[6] Current address: Xenbase, Department of Computer Science, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4, Canada
[7] Berkeley Drosophila Genome Project, 240C Building 64, Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA 94720, USA
[8] Institute of Neuroscience, University of Oregon, Eugene, OR 97403-1254, USA

Understanding the developmental and genetic underpinnings of particular evolutionary changes has been hindered by inadequate databases of evolutionary anatomy and by the lack of a computational approach to identify underlying candidate genes and regulators. By contrast, model organism studies have been enhanced by ontologies shared among genomic databases. Here, we suggest that evolutionary and genomics databases can be developed to exchange and use information through shared phenotype and anatomy ontologies. This would facilitate computing on evolutionary questions pertaining to the genetic basis of evolutionary change, the genetic and developmental bases of correlated characters and independent evolution, biomedical parallels to evolutionary change, and the ecological and paleontological correlates of particular types of change in genes, gene networks and developmental pathways.

Introduction

One of the most challenging questions in biology is how the genome and its emergent properties are modified over evolutionary time to produce the diverse anatomical forms seen throughout the natural world. Studying this question requires a systems approach [1] that synthesizes knowledge from various biological levels, including gene structure and function, development, evolutionary and phylogenetic relationships, and ecology. Such synthesis also requires bioinformatics tools; however, global bioinformatics efforts are primarily focused at the genomic level and researchers have made significant progress by using databases to catalog information based on ontologies, that is, the use of constrained, structured vocabularies with well defined relationships among terms. Ontologies represent a knowledge-base of a particular discipline, and provide not only a mechanism for consistent annotation of data, but also greater interoperability among people and machines [2]. The most widely used biological ontology is the Gene Ontology (GO) (http://www.geneontology.org), which is utilized to annotate molecular function, biological processes and subcellular localization to gene products from different organisms. This approach has provided much insight into the molecular nature and evolution of gene products across taxa.

New initiatives to connect the genome to mutant phenotypes of model organisms, such as projects of the National Center for Biomedical Ontology (http://www.bioontology.org), have resulted in an ontology of phenotypic qualities, called the ‘Phenotype And Trait Ontology’ (PATO), which can be used in combination with anatomy ontologies for model organism species to describe phenotypes. For example, researchers in the Zebrafish Information Network (ZFIN; http://www.zfin.org) are annotating mutant phenotypes using the zebrafish anatomy ontology and the PATO ontology. Here, we propose to link phylogenetic and homology data to genetic data using multi-species anatomy ontologies. This method provides a computable connection from evolution to genotype through anatomy ontologies.

The rise of ontologies

The Linnean species ontology

An ontology is a representation of the types of entities that exist, and of the relationships among them [3]. In systematics, for example, a Linnean classification is an ontology. Its classes, also called types, are the taxa at various ranks, each of which have formal definitions and a specific formal subtyping relationship to each other. A specific species (e.g. common carp *Cyprinus carpio*) *is_a*[*] specific genus, *Cyprinus,* which, in turn, *is_a* specific family, Cyprinidae and so on. Each type (e.g. the genus *Cyprinus*) has general properties that it has inherited from its parent family (i.e. synapomorphies) and less general properties (i.e. less inclusive synapomorphies) that distinguish it from other

Corresponding author: Mabee, P.M. (pmabee@usd.edu).
Available online xxxxxx.

[*] By convention, all relationships and ontology types are italicized.

www.sciencedirect.com 0169-5347/$ – see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.tree.2007.03.013

genera of this family. In turn, all *Cyprinus* species inherit ontologically all the properties of the genus *Cyprinus,* with each species having its own distinguishing properties.

The Linnean classification is a single hierarchy because it represents the single ancestor–descendant branching tree of life. Each child type (i.e. descendant) has a single parent type (i.e. ancestor), whose properties it inherits. Other ontologies can be more complex, consisting of multiple parents with the same or different relationships. For example, my index finger *is_a* finger and also *part_of* my hand (i.e. it has two parents). The definition of the child, *index finger*, would further refine the definition of the parent, *finger*, but it would not inherit the definition of the parent, *hand*, with which it has a different type of relationship. The structure of this ontology would not be a strict hierarchy, but would instead be represented by a directed acyclic graph, in which types can have multiple parents and different relationships between them. The richness of this ontological structure is one of the features that makes ontologies distinct from controlled vocabularies, which are simply constrained lists of terms. Strictly defining both the types and the relationships between them enables reasoning across the ontology. For instance, if we declare that my index finger is *part_ of* my hand and my hand *part_of* my arm, we can conclude that my index finger is *part_of* my arm.

Ontologies facilitate interoperability

Ontologies are important because they are formal specifications of some aspect of reality, and both humans and computers can use them. They promote interoperability, that is, communication such as cross-querying among databases. Ontologies are ultimately used for communication between people and machines [2]. One of the reasons that the GO has been so successful and widely used is because it attains this objective. Use of the GO means, for example, that when gene products in FlyBase (http://flybase.bio.indiana.edu/) and the *Saccharomyces* Genome Database (http://www.yeastgenome.org/) are described as having the function ‘protein tyrosine phosphatase activity’, both databases use exactly the same definition. Moreover, a search of these databases for gene products with the more general term ‘protein phosphatase activity’, returns, *inter alia,* all gene products with ‘protein tyrosine phosphatase activity’.

The GO is now in use in all major model organism databases and in many other large databases in the genomics domain, for example, the UniProt protein sequence and function database (http://www.ebi.uniprot.org/) [4] and the Protein Data Bank (PDB) protein structure database (http://www.pdb.org/) [5]. It is also used extensively for literature analysis [6,7], and for the annotation of physical objects, such as probes on microarrays.

The types in an ontology should be rigorously defined and can have synonyms to help database searches and to enable different terms to be used by different communities to denote the same type of entity. For example, the human Foundational Model of Anatomy (http://fme.biostr.washington.edu:8089/FME/index.html) has a type *neuraxis*, a term used by the medical community, and a synonym *central nervous system,* which is more commonly used by the broader biological community. The types must also have unique identifiers (IDs) that never change even if the name or spelling of the term changes, and any ontology must declare the rules governing the persistence and change of identifiers.

Anatomy ontologies

More recently, anatomical ontologies [8] have been developed by model organism research communities (e.g. mouse, zebrafish and *Drosophila*). These consist of standard vocabularies of entities (organs, tissues, cell types and developmental stages) that are related hierarchically. Relationships are typically *is_a* (subtyping*), part_of* or *develops_from* [9]. For example, the zebrafish *naris is_a surface structure* and is *part_of* the *peripheral olfactory organ*, which itself *develops_from* the *olfactory placode* (Figure 1). There is also a Cell Type ontology (http://obo.sourceforge.net/cgi-bin/detail.cgi?cell) that is broadly applicable across plants and animals [10]. Typically, these ontologies are used to annotate gene expression data or phenotypic data within the context of databases.

There are currently ~15 anatomical ontologies registered at the National Center for Biomedical Ontology, many of which are linked to organism databases [8], including the *Drosophila* database (FlyBase at http://flybase.bio.indiana.edu/); Edinburgh Mouse Atlas Project (http://genex.hgu.mrc.ac.uk/; zebrafish database (ZFIN at http://zfin.org) and Plant Ontology Consortium (http://www.plantontology.org/); they can be accessed at Open Biomedical Ontologies (http://obo.sourceforge.net/browse.html). These anatomical ontologies, or ‘anatomics’ [8] have sprung up in many cases without significant input from comparative evolutionary morphologists or systematists. Similar to the types in molecular ontologies, anatomical types can have multiple synonyms so that users can search via different aliases for the same entity.

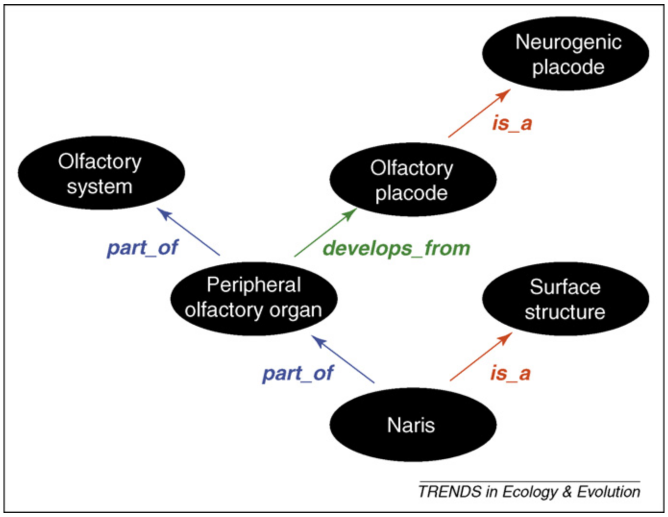
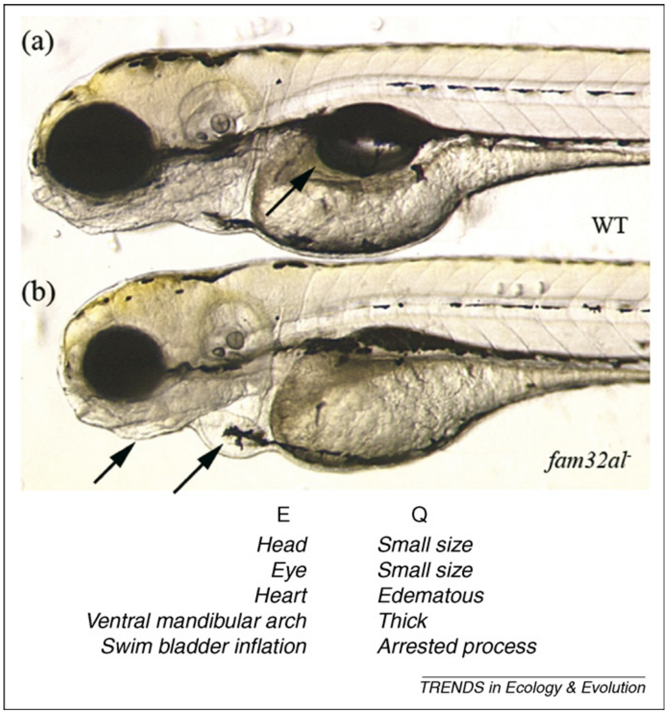


Figure 1. Portion of the zebrafish anatomy ontology in which the different relationships of the *naris* to other anatomical entities are shown. Colors represent different types of relationships: red, *is_a*; blue, *part_of*, and green, *develops_from*. Arrows points toward the parent in each relationship. The naris (child) *is_a surface structure* (parent). The naris is also *part_of* a *peripheral olfactory organ*, which itself *develops_from* an *olfactory placode.* In turn, the *olfactory placode is_a neurogenic placode* and *part_of* the *olfactory system.*

Please cite this article in press as: Mabee, P.M. et al., Phenotype ontologies: the bridge between genomics and evolution, Trends Ecol. Evol. (2007), doi:10.1016/j.tree.2007.03.013

A phenotypic quality ontology

There are many descriptions of disease and mutant phenotypes in the literature. However, searching these descriptions is limited to free text search algorithms, which are problematic because different authors and communities use different terminology and syntax. Therefore, a controlled and consistent method has been developed to describe these phenotypes. With the objective of capturing qualitative and quantitative information about phenotypes in a species-neutral and systematic way, Ashburner and Lewis proposed PATO (http://www.bioontology.org/wiki/index.php/PATO:Main_Page), which is an ontology of phenotypic qualities that can be used to capture the differences between wild-type and mutant phenotypes of all organisms. PATO qualities can be combined with types from entity ontologies, such as the various anatomical ontologies, to describe how an entity is changed by a mutation or other experimental procedure. These types are components of a bipartite syntax, called EQ (Entity–Quality; Figure 2), which provides a biologically relevant means of describing morphology and whose ontological underpinnings make it easily interpreted by computers [11–13]. The EQ syntax is now being used to describe a range of phenotypic changes in different species. For example, a disease or mutant phenotype can be described as the sum of multiple EQ annotations (Figure 2) together with the organism attributes such as genotype [14]. In this way, the phenotypic effects of gene mutation in a model organism can be compared and analyzed together with genetic sequence data to piece together the underlying genetic and molecular mechanisms. This schema is applicable to any organism, as has been shown recently for the mouse [14].



E	Q
Head	*Small size*
Eye	*Small size*
Heart	*Edematous*
Ventral mandibular arch	*Thick*
Swim bladder inflation	*Arrested process*

TRENDS in Ecology & Evolution

Figure 2. The use of EQ syntax by the zebrafish model organism database (ZFIN). The zebrafish in **(a)** is wild type, whereas that in **(b)** carries a mutant in the gene with sequence *fam32al* (family with sequence similarity 32, member A, like) **Q6GQN4** (GenBank accession number). Arrows point to the mutant ventral mandibular arch and edematous heart, and the normal wild-type swim bladder. *Head, eye, heart* and *ventral mandibular arch* are entity types from the zebrafish anatomy ontology, and *swim bladder inflation* is an entity from the GO. The qualities are from the PATO. Combinations of different entities and qualities are used to describe phenotypes.

Status of comparative morphological systematics

Comparative pre-genomic era studies of the phenotypes of organisms have produced a large body of text describing homologous features of evolutionary anatomy. Most of this descriptive text is in museum monographs and other literature and is not comprehensively searchable, let alone computable. Such homologous features, or systematic characters, are drawn from every observable aspect of the organism: molecular, morphological and behavioral. Since the implementation of rigorous phylogenetic methods [15], organisms have been compared with the goal of discovering which characters contain historical information (i.e. are synapomorphic), at a particular hierarchical level in the phylogeny of life. They are coded in a matrix format and analyzed with increasing variety of optimality criteria (e.g. maximum parsimony or maximum likelihood), probabilistic models (e.g. Bayesian) and other assumptions [16–18]. The results of such phylogenetic studies are published mainly in journals and monographs, but attempts have recently been made to describe and store comparative images and text in Web-accessible databanks such as MorphoBank (http://morphobank.informatics.sunysb.edu/) and MorphBank (http://www.morphbank.com/) [19,20].

MorphoBank is a repository of morphological character matrices used in phylogenetic analysis. It can also be used as a workspace for coding such matrices and writing them in the NEXUS format, a file format designed to contain systematic data for use by computer programs [21]. MorphBank is a repository of images of organisms or parts of organisms that can be used as a digital record of structures coded as characters in a phylogenetic analysis, or specimens from which molecular data have been extracted.

Identifying the phylogenetic relationships of life on a large scale requires integration across species [22], data from disparate biological levels (molecular and phenotypic, including morphological, behavioral, paleontological, etc.), and collaborative groups of investigators. The NSF Assembling the Tree of Life program (http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5129andorg=BIOandfrom=home) supports investigative groups who specialize in particular clades of the tree, with the ultimate goal of reconstructing the evolutionary relationships of all of life. Many additional investigators focus independently on small branches of the tree. DNA sequence data can be gathered rapidly and cost effectively, and provide a useful way to reconstruct phylogeny. Morphological data, however, are needed to understand how biological form has changed during evolution. From the standpoint of phenotypic data, there are two key issues of concern: (i) how can the phenotypic characters of systematics, currently embedded as free text in character matrices, be connected to homologous characters in other matrices?; and (ii) how can such phenotypic data be connected to genomic information? To resolve these issues, phenotypic characters need to be rendered more widely computable. For integration to occur, such characters must be computable not

Please cite this article in press as: Mabee, P.M. et al., Phenotype ontologies: the bridge between genomics and evolution, Trends Ecol. Evol. (2007), doi:10.1016/j.tree.2007.03.013

only in phylogenetic analyses, but also in the broader bioinformatics domain. We propose here that using anatomy and PATO ontologies is one way of resolving these issues. This approach can be used to answer a new and revolutionary scale of questions such as: which genes change to produce evolutionary shifts in body form at particular times during phylogeny? Do evolutionary phenotypes mirror human disease states and what is their significance? What is the genetic basis for evolutionary parallelism or for the correlated evolution of body parts?

Phylogenetic characters and ontologies

The EQ syntax, implemented by model organism communities chiefly to connect the phenotype to the genotype, provides an adaptable starting point for evolutionary morphologists. We propose that the syntax can be extended to describe the characters and character states of evolutionary biology. For example, morphological studies of fish evolution typically focus on variation in the presence, absence, shape and number of skeletal parts. These features, or characters, resolve phylogenetic relationships of species at various levels. A typical description of the condition in a single species might read: Character = 'Shape of the second basibranchial' and Character State = 'rod-like'. Another species might be described by: Character = 'Shape of the second basibranchial' and Character State = 'spathulate'. Translating these characters into the EQ vocabulary and syntax using an anatomical ontology for entities and the PATO for qualities would produce this description for the first species: Entity = Basibranchial 2; Quality = rod-like (a child of the parent quality *shape*), and this for the second species: Entity = Basibranchial 2; Quality = spathulate (another child of *shape*). More complex characters, such as 'Connection between basibranchial 1 and basibranchial 2' would involve two entity terms, and other characters might involve additional qualifier terms.

Although the details of a standardized syntax for morphological data remain to be worked out, it is clear that the morphology can be described using ontological terms. The benefits of using a consistent ontology and syntax to describe systematic characters are immense. The most prominent and exciting of these benefits would be the ability of evolutionary morphologists to query genomic databases via anatomical ontologies shared with the model organism communities, thus linking evolutionary changes to genetic changes.

Limiting the proliferation of ontologies

One difficulty in unifying descriptions of phenotype, as used by evolutionary biologists, with those of the genomics community is the complexity and subtlety of differences that evolutionary biologists have found to be important [23,24]. This contrasts with the less detailed descriptions of morphology used by model organism communities, which are concerned with a single organism and specific anatomical parts. In particular, model organism communities are concerned with describing deviation from 'wild type,' whereas evolutionary morphologists compare many species with respect to 'outgroups' [25]. The benefits of applying a standardized syntax to systematic characters provide the means of connecting data among organisms in a standardized way. Conceivably, this could require as many ontologies as species. Here, we propose a mechanism to limit the number of entity ontologies by sharing terms among species wherever possible using a taxonomy ontology to codify species names.

Multi-species anatomy ontologies

Because it would be too cumbersome to have as many anatomical ontologies as species, the question arises as to what the optimal taxonomic coverage of anatomical ontologies should be. Anatomical ontologies should cover as large a range as possible, but should be limited to taxa that share most anatomical terms. However, large ontologies will pose formidable problems for resolving homology relations and terminological usage. Large ontologies will also require implementation of new tools to aid curation and searching of long lists. An alternative is to develop a series of more taxonomically specialized ontologies, within which terminology and homology would be relatively easy to settle. By contrast, organisms with distinctly different body plans, such as echinoderms and chordates, would need separate anatomical ontologies.

Single-species anatomy ontologies can be used as the basis for expansion into multi-species ontologies. These ontologies would contain terms from several species where the ancestors of the lineages are united by common features. Multi-species anatomical ontologies could be developed in conjunction with a taxonomic ontology where anatomical terms could be limited to specific taxa via a new relationship type, such as *in_taxon*. The resultant multi-species anatomical ontology would contain all the entities for a given set of taxa, but the types themselves would be limited to use within their respective taxa. These multi-species anatomy ontologies can then be associated with phylogenetic and homology data, and geological time. Development of multi-species anatomy ontologies will require significant work and commitment by the evolution community and associated species experts at various branches in the tree of life (Figure 3).

Homology

This brings up a crucial issue for evolutionary biology that must be addressed to formalize connections among the anatomical ontologies: how should homology be treated? Evolutionary comparisons operate at a high tier in systems biology, namely at the level of continuity and modification of the phenotype across the tree of life. Similarity of the phenotype owing to the continuity of inherited information (i.e. homology) [26,27] is important for biologists at all levels and central to all comparative biology [28]. Thus, ultimately, connections among terms within and across anatomical ontologies must be defined by homology relationship statements. Developing a mechanism to define and use this relationship is required to address the queries of the evolutionary community. Homology relations can be expressed as a mapping between ontologies in a database, which enables greater flexibility than does encoding homology relations within an anatomical ontology itself. Homology assignments need to be based on evidence for relationship (e.g. position, development or composition), and they need attribution to a literature or

Please cite this article in press as: Mabee, P.M. et al., Phenotype ontologies: the bridge between genomics and evolution, Trends Ecol. Evol. (2007), doi:10.1016/j.tree.2007.03.013

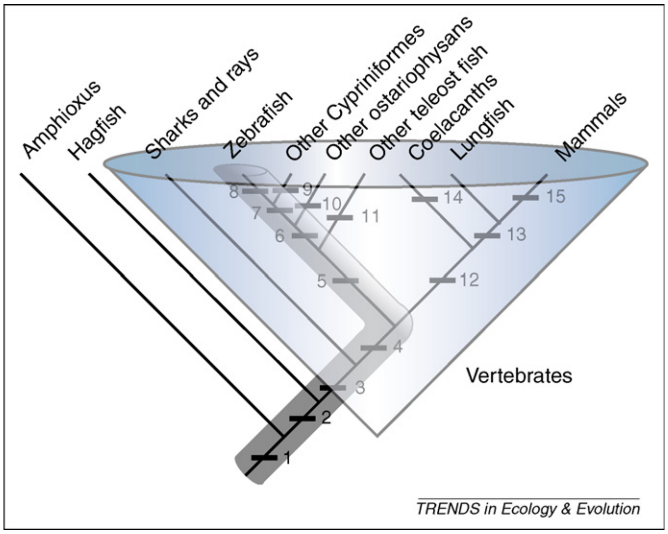


Figure 3. Expansion of the zebrafish anatomical ontology (grey inner cylinder) into a multi-species anatomy ontology that includes features of all vertebrates (blue cone). The zebrafish anatomical ontology (~1500 terms) already includes many features that evolved at various times along the common vertebrate lineage (characters 1–4). For example, the jaw (character 3) is present not only in zebrafish but in all vertebrates. However, anatomical features that are uniquely present in divergent lines of fishes (e.g. characters 9–11) or mammals (character 15) are not in the zebrafish anatomical ontology. To develop a vertebrate multi-species anatomy ontology, the zebrafish ontology must be extended to cover the full range of anatomical characteristics in the other vertebrate species.

other source. Such evidence codes could be used to identify the strength of the homology statement. Moreover, an investigator could specify which types or levels of homology evidence they wish to use when building phylogenies. This would also enable homology statements to be made within a multi-species anatomical ontology or among ontologies, facilitating cross-species comparisons of diverse taxa.

Data formats for analyzing EQ data

Currently, evolutionary biologists collect and analyze phenotypic data in the form of characters and character states (C, CS) rather than entities and qualities (EQ). The advantage of the C, CS method is that it is simple to form a matrix of taxa by characters. However, C, CS studies tend to produce character lists and state lists that are highly specific to a single study, because characters are free text amalgams of both entities and qualities and there are many ways, both in terms of vocabulary and syntax, to represent the same character. An example would be: 'Basibranchial element number, four' or 'Number of basibranchial elements present, four'. It is difficult for a computer to parse those strings identically.

The decomposition of characters into entities and qualities enables standard lists of entities and qualities, all with unique identifiers, to be used with a standard syntax. Thus, under the EQ syntax, the form: (Basibranchial, four) is easy to standardize and has the additional advantage that each part of the character, (i.e. both entity and quality), are independently computable as objects. Moreover, the EQ description can easily be mapped computationally to a C, CS description, whereas the reverse mapping requires human intervention. At present, the C, CS system is 'locked in' to evolutionary biology, in part, because the standard data formats for morphological data currently enable only the C, CS system to be used. The NEXUS format and the DELTA (commonly used in systematics, particularly in plants) are examples of two such formats. Relatively minor modifications to these formats, and others like them, would enable them to include information in EQ format. Descriptions can then be treated as bipartite (C, CS) data for use with phylogenetic software or (EQ) data for integration with information from genomics databases; the formalization should always be independent of the technology.

Conclusions

A computable connection from phenotype to genotype, via a standardized EQ syntax, will support a new scale of research questions. These might be straightforward, such as: which genes are known to be expressed in the development of a particular morphological structure? Is there a model organism mutant that has a phenotype similar to a human disease? But even more interestingly, this strategy will also support studies of complex evolutionary questions such as: what is the set of genes that are associated with a particular type of morphological change that occurred independently in several clades during evolution? Which genes are responsible for evolutionary change in the shape of a particular body part? What evolutionary phenotype mirrors a particular human disease? Analysis of changes in morphology correlated with genetic changes will lead to a greater understanding of gene function as a whole.

The opportunity now exists to unify over 300 years of traditional morphological investigation with the fruits of the genomics revolution to provide a common descriptive platform for the whole of biology. Although the rewards would be immense, there are significant obstacles; databases and tools need to be developed and research communities need to change (Box 1). Linking evolutionary and genomics databases through phenotype provides a path to understanding the levels of complexity that separate the developmental and evolutionary transformation of

Box 1. Requirements for a computable connection between genotype and phenotype across evolution

- Anatomy ontologies must be expanded to encompass all organisms, either by extending those core ontologies of model organisms (Figure 3, main text), or by developing new ontologies for groups in which there are presently no model organisms. These anatomical ontologies must conform to a common set of standards.
- The existing morphological characters of systematics currently embedded as free text in character matrices and literature need to be parsed in the EQ syntax (Figure 2, main text) and referenced to ontologies. Evolutionary biologists need new tools to capture, store and analyze character data using ontologies.
- Because evolutionary homology connects entities across separate ontologies, this historical and genealogical relationship must be accommodated to compute across databases.
- Bioinformatics methods are needed to facilitate visualization [29] of multidimensional genotype and phenotype data in multiple organisms simultaneously.
- Finally, and perhaps most difficult, a change in viewpoint is required not only on the part of evolutionary anatomists, who will need to see the benefits of the EQ syntax and of integration with genomics, but also on the part of the genomics community, who will need to plan for comparative genomics encompassing all life rather than only a few model organisms.

Please cite this article in press as: Mabee, P.M. et al., Phenotype ontologies: the bridge between genomics and evolution, Trends Ecol. Evol. (2007), doi:10.1016/j.tree.2007.03.013

phenotype. Such understanding will be crucial to realizing the full potential of genomics to explore and understand the evolution of life on Earth.

Acknowledgements

We thank the NSF National Evolutionary Synthesis Center (EF-0423641), NSF 0431290, NIH HG002659 and HG004028 for support. G.V.G. is supported by the National Center for Biomedical Ontology (NIH: 1-U54-HG004028-01). We thank Adam Amsterdam and Nancy Hopkins for use of the image in Figure 2.

References

1 Kirschner, M.W. (2005) The meaning of systems biology. *Cell* 121, 503–504

2 Masolo, C. *et al.* (2003) *Ontology Library (final)*, WonderWeb Deliverable D18. http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf

3 Gruber, T.R. (1993) A translation approach to portable ontologies. *Knowledge Acquisition* 5, 199–220

4 Apweiler, R. *et al.* (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119

5 Wolstencroft, K. *et al.* (2005) Constructing ontology-driven protein family databases. *Bioinformatics* 21, 1685–1692

6 Jenssen, T-K. *et al.* (2001) A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.* 28, 21–28

7 Hirschman, L. *et al.* (2005) Overview of BioCreAtIvE: critical assessment of information extraction for biology. *BMC Bioinform.* 6 (Suppl. 1), 1–10

8 Bard, J.B.L. (2005) Anatomics: the intersection of anatomy and bioinformatics. *J. Anat.* 206, 1–16

9 Smith, B. *et al.* (2005) Relations in biomedical ontologies. *Genome Biol.* 6, R46

10 Bard, J. *et al.* (2005) An ontology for cell types. *Genome Biol.* 6, R21

11 Li, J.L. *et al.* (2005) PhD: a web database application for phenotype data management. *Bioinformatics* 21, 3443–3444

12 Nadkarni, P.M. *et al.* (2000) WebEAV: automatic metadata-driven generation of web interfaces to entity-attribute-value databases. *J. Am. Med. Inform. Assoc.* 7, 343–356

13 Nadkarni, P.M. *et al.* (1999) Organization of heterogeneous scientific data using the EAV/CR representation. *J. Am. Med. Inform. Assoc.* 6, 478–493

14 Gkoutos, G.V. *et al.* (2004) Ontologies for the description of mouse phenotypes. *Comp. Funct. Genomics* 5, 545–551

15 Hennig, W. (1966) *Phylogenetic Systematics.* University of Illinois Press

16 Page, R.D.M. and Holmes, E.C. (1998) *Molecular Evolution: A Phylogenetic Approach.* Blackwell Science

17 Felsenstein, J. (2004) *Inferring Phylogenies.* Sinauer Associates

18 Delsuc, F. *et al.* (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375

19 Fontal-Cazalla, F.M. *et al.* (2002) Phylogeny of the Eucoilinae (Hymenoptera: Cynipoidea: Figitidae). *Cladistics* 18, 154–199

20 Hill, R.V. (2005) Integration of morphological data sets for phylogenetic analysis of Amniota: the importance of integumentary characters and increased taxonomic sampling. *Syst. Biol.* 54, 530–547

21 Maddison, D.R. *et al.* (1997) NEXUS: an extensible file format for systematic information. *Syst. Biol.* 46, 590–621

22 Wilson, E.O. (2003) The encyclopedia of life. *Trends Ecol. Evol.* 18, 77–81

23 Diederich, J. (1997) Basic properties for biological databases: character development and support. *Mathl. Comput. Modell.* 25, 109–127

24 Diederich, J. *et al.* (1997) Construction and integration of large character sets for nematode morpho-anatomical data. *Fund. Appl. Nematol.* 20, 409–424

25 Lundberg, J.G. (1972) Wagner networks and ancestors. *Syst. Zool.* 18, 1–32

26 Van Valen, L.M. (1982) Homology and causes. *J. Morphol.* 173, 305–312

27 Roth, V.L. (1984) On homology. *Biol. J. Linnean Soc.* 22, 13–29

28 Bock, G.R. and Cardew, G., eds (1999) *Homology*, John Wiley & Sons

29 Tao, Y. *et al.* (2005) Visualizing information across multidimensional post-genomic structured and textual databases. *Bioinformatics* 21, 1659–1667

Please cite this article in press as: Mabee, P.M. et al., Phenotype ontologies: the bridge between genomics and evolution, Trends Ecol. Evol. (2007), doi:10.1016/j.tree.2007.03.013