 doi:10.1093/bib/bbr057

Editorial: Orthology and applications

The accurate inference of orthologous genes underpins almost all biological studies that consider more than a single genome. Indeed, orthology formalizes the intuitive notion of corresponding genes in different species. As such, orthology finds applications in a broad range of research areas, such as functional genomics, comparative genomics, phylogenetics or pharmacology. Accordingly, well over 30 orthology databases have been developed (http://q4o.org/orthology_databases) and many thousands of scientific papers containing the keyword 'ortholog' are published each year.

But success also comes with new challenges. In particular, each area of orthology applications entails its own constraints and trade-offs. This has given rise to multiple and at times conflicting definitions of orthology and associated relations—a common source of confusion even among long-time practitioners. Hence, to effectively call, interpret or apply ortholog predictions, knowledge of the problem in question is indispensable.

The aim of this special issue is to provide a survey of the current state of orthology through multiple lenses, in form of reviews and original research papers.

We start with a tribute to Walter M. Fitch, who passed away earlier this year. In his note of remembrance, Eugene Koonin provides a retrospective on Fitch's founding role in orthology.

The rest of the first part focuses on definitions and methods for orthology inference. Kristensen *et al.* review the numerous computational methods that have been developed in recent years. They discuss the relative merits of the various approaches, both in theory and in practice.

Doyon *et al.* consider orthology inference in the context of the more general problem of gene and species tree reconciliation. Indeed, orthology can be viewed as a byproduct of tree reconciliation. The authors review latest developments in parsimony and likelihood approaches. In particular, they report on models accounting not only for speciations and gene duplications, but also for lateral gene transfers.

The contribution by Colin Dewey addresses the notion of positional orthology, which he formally defines in terms of past evolutionary events—not in terms of conserved gene neighborhood in present genomes (as in e.g. [1]).

Sjölander *et al.* discuss the challenges of orthology inference when the underlying genes have heterogeneous domain architectures. They discuss a protocol for phylogenetic orthology inference based on domains instead of full-length protein sequences. They argue that the denser taxon sampling afforded by domain-level analyses counterbalances the phylogenetic uncertainty caused by shorter domain sequences.

Boeckmann *et al.* compare seven well-established phylogenomic databases from the perspective of the user. We describe conceptual differences, and what they mean in terms of the orthology, paralogy and tree structure conveyed by each database. The paper shows how measuring these three aspects can allow for effective benchmarking based on reference gene trees.

In the second part of this special issue, our attention shifts to applications of orthology. The perhaps greatest impact of orthology studies lies with gene function characterization. Though it might be tempting to systematically ascribe the same function to orthologous genes, Gharib and Robinson-Rechavi remind us that even for the relatively short human–mouse evolutionary distance, there are numerous instances of orthologs that have diverged functionally. Overall, however, Huerta-Cepas *et al.* report that human–mouse orthologs are significantly more conserved in expression pattern than their paralogous counterparts.

These observations underscore the need for differentiated and prudent approaches to propagating function annotations. In their manuscript, Gaudet *et al.* describe the method used by curators of the Gene Ontology consortium to integrate and transfer function annotations based on the evolutionary history of gene families.

In medical research, the focus is not so much on gene function as on gene dysfunction. Schreiber *et al.* present a major update of OrthoDisease, a database of human disease-associated genes and their orthologs in nearly 100 other species. Based on their data, they observe that disease-associated genes tend to

© The Author 2011. Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

Downloaded from bib.oxfordjournals.org at ETH-Bibliothek on September 29, 2011

have fewer close paralogs than other human genes, thereby supporting the notion that close paralogs can compensate each other functionally [2, 3].

In another orthology application, Dessimoz *et al.* examine the problem of split genes, endemic in low-coverage genome assemblies. We present a comparative genomics approach aimed at detecting these gene fragments present on multiple, unassembled contigs. This is of particular relevance here, because such pseudo-paralogs can confound orthology prediction and other phylogenetic analyses.

The special issue closes with a letter by Schmitt *et al.*, in which they define and motivate new XML formats for protein sequences and orthology predictions. This initiative epitomizes recent community efforts toward better interoperability and joint standards [4], and its outcome should facilitate the interpretation of results provided by the various orthology databases.

FUNDING

CD is supported by an SNSF advanced researcher fellowship (#136461).

Christophe Dessimoz
European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

References

1. Koski LB, Morton RA, Golding GB. Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol* 2001;**18**:404–12.
2. Lopez-Bigas N. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res* 2004;**32**:3108–14.
3. Hsiao T-L, Vitkup D. Role of duplicate genes in robustness against deleterious human mutations. *PLoS Genet* 2008;**4**: e1000014.
4. Gabaldón T, Dessimoz C, Huxley-Jones J, *et al.* Joining forces in the quest for orthologs. *Genome Biol* 2009;**10**:403.

Downloaded from bib.oxfordjournals.org at ETH-Bibliothek on September 29, 2011