The Expression Comparison Tool in Bgee

Marc Robinson-Rechavi, Valentine Rech de Laval, Frédéric B. Bastian, Julien Wollbrett, Team Bgee

HAL Id: hal-02535720
https://hal.archives-ouvertes.fr/hal-02535720

Submitted on 10 Apr 2020

Chapter 4.3 The Expression Comparison Tool in Bgee

Marc Robinson-Rechavi, Valentine Rech de Laval, Frédéric B. Bastian, Julien Wollbrett, Bgee Team

Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
marc.robinson-rechavi@unil.ch

Abstract

We present Expression Comparison, a tool to compare expression patterns between species. It uses curated annotations of homology between anatomical structures, such as organs or tissues. Expression calls are based on the curated transcriptome data integrated within the Bgee database. Gene homology can be of any type, from user input. The results are presented according to conservation of pattern, as well as rank of expression per species. Expression Comparison is freely available on the Bgee website: `https://bgee.org`.

1 Introduction

One of the most challenging aspects of phylogenomics is comparing function between homologous genes (see Chapters 4.1 and 4.2 [Necsulea 2020; Robinson-Rechavi 2020]). While gene expression is probably the aspect of function which is the most amenable to comparison between genes and between species, there has not been any tool to automatically provide such comparisons. This stands in contrast to the situation for sequences, where many databases exist which provide not only orthologous and paralogous genes, but also their sequences, multiple sequence alignments, and trees (Vilella et al., 2009; Scornavacca et al., 2019), as well as genomic regions and comparative synteny (Nguyen et al., 2018). We present here the first version of a new tool, Expression Comparison (Table 1), which leverages the Bgee database to provide such a service.

Bgee is a database of gene expression (Bastian et al., 2008, 2020) which provides manually curated healthy wild-type data for a variety of animal species, annotated to the Uberon ontology of anatomy (Haendel et al., 2014) and to standard ontologies of development and aging. Annotations also capture sex and strain or population when possible. Expression data is integrated from RNA-seq, microarrays, in situ hybridization, and ESTs. Calls of presence and absence of gene expression are made for each gene – condition combination, where a condition is a combination of anatomical structure (e.g., organ, tissue, cell type), developmental stage, sex and strain. These calls integrate all the data types together. The importance of expression of each anatomical structure in the expression of a gene is also integrated over data types, by a weighted mean of expression ranks. Thus Bgee provides a global view of "normal" gene expression in a comparable way between species, despite

differences in anatomy and data availability. These integrated calls and ranks provide a standard source of information to compare expression between genes.

To compare gene expression between species, we need to define comparable conditions. As for genes, where we seek to compare orthologs or paralogs, we need to define homologous anatomical structures to compare. In Bgee, these are manually annotated from the literature, and provided as a supplementary annotation to the Uberon ontology (Table 1). Importantly, they are integrated into the database and can be automatically queried.

Resource	Url
Expression Comparison	`https://bgee.org/?page=expression_comparison`
Anatomical homology annotations	`https://github.com/BgeeDB/anatomical-similarity-annotations`

■ **Table 1** Resources cited in main text.

2   Using Expression Comparison

The Expression Comparison is a webserver tool which queries the Bgee database for gene expression presence or absence, rank, and anatomical homology. The user should insert into the query form a list of gene identifiers. At time of writing, it is up to the user to define this list, e.g. based on another source of information for orthology. It is possible to compare expression patterns of non homologous genes. At present, only Ensembl identifiers are accepted.

From this gene list, the tool will query Bgee and retrieve all homologous anatomical structures which have expression in at least one of the genes of the list, and have homology between the species represented by the list. Thus, a gene list with orthologs in human and zebrafish will only retrieve expression in anatomical structures which have defined homology between mammals and teleost fishes, e.g. not in placenta. The user is then provided with a list of such structures ordered by their presumed relevance; the user can re-order by clicking on column headers. A subset of results for the brain-specific gene SRRM4 (Serine/arginine repetitive matrix protein 4, required for neural cell differentiation) is presented in Table 2. The score goes from +1 for perfect conservation of expression presence, to -1 for perfect conservation of expression absence, with 0 indicating no conservation. Notice that not all 13 orthologs used in the comparison of Table 2 are always present, even for a score of 1, because there is less data in some species than others, thus some orthologs are neither called absent nor present in, e.g., cerebellar cortex. Anatomical structures are ranked by default by this score, then by genes with presence of expression, and finally by "Minimum rank". The latter is the lowest rank of any of the compared genes in the homologous anatomical structure; a lower rank indicates a higher importance of expression. Thus the top anatomical structures reported have high consistency, have expression for many of the genes compared, and have high expression levels for at least some of these genes. Indeed, for SRRM4, the top structures are the brain and sub-parts of the brain.

User can re-order the table online. They can unfold an anatomical structure to see all the genes and the species according to their presence or absence of expression, or lack of data. The anatomical structures are linked to their description in Uberon, for users who would not know what is, e.g., Ammon's horn. The genes are linked to their Bgee gene page, which provides detailed and species-specific expression information. The species names are linked to their Bgee species page, which provides all the data for the species in downloadable files.

Anatomical entities	Score	Minimum rank	Anatomical entity IDs	Gene count with presence of expression
brain	1	5.86E+03	UBERON:0000955	13
central nervous system	1	1.64E+04	UBERON:0001017	13
multi-cellular organism	1	1.69E+04	UBERON:0000468	13
forebrain	1	5.29E+03	UBERON:0001890	10
telencephalon	1	1.54E+04	UBERON:0001893	10
cerebellum	1	3.52E+03	UBERON:0002037	9
male reproductive system	1	2.31E+04	UBERON:0000079	6
female reproductive system	1	2.05E+04	UBERON:0000474	4
cerebellar cortex	1	3.21E+03	UBERON:0002129	3
Ammon's horn	1	1.14E+04	UBERON:0001954	3

■ **Table 2** Top result of Expression Comparison on SRRM4 orthologs (subset of table generated from Bgee 14.1)

The table itself can be easily downloaded as TSV or copied to clipboard, for further use in, e.g., R or MS Excel.

3 Conclusion and perspectives

The Expression Comparison tool is the first available tool to automatically compare gene expression between genes taking into account curated information on anatomical homology. It leverages the data integration, transcriptome annotation, and anatomical homology annotation, in Bgee. It is already proving to be one of the more popular pages of the Bgee website, which shows that there was an unmet need for expression comparison. Future development will include automated recovery of orthologs and in-paralogs, to allow a fully automated expression comparison starting from one gene, similar to what is available in phylogenomic databases for sequences.

References

Bastian, F., Comte, A., Echchiki, A., Escoriza, A., Gharib, W., Gonzales-Porta, M., Jarosz, Y., Laurenczy, B., Mendes de Farias, T., Moret, P., Moretti, S., Niknejad, A., Parmentier, G., Person, E., Rech De Laval, V., Roelli, P., Rosikiewicz, M., Roux, J., Sanjeev, K., Seppey, M., Wollbrett, J., and Robinson-Rechavi, M. (2020). The bgee database: curated reference gene expression data and analytics tools. In preparation.

Bastian, F., Parmentier, G., Roux, J., Moretti, S., Laudet, V., and Robinson-Rechavi, M. (2008). Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species. In *Data Integration in the Life Sciences*, volume 5109, pages 124–131.

Haendel, M., Balhoff, J., Bastian, F., Blackburn, D., Blake, J., Bradford, Y., Comte, A., Dahdul, W., Dececchi, T., Druzinsky, R., Hayamizu, T., Ibrahim, N., Lewis, S., Mabee, P., Niknejad, A., Robinson-Rechavi, M., Sereno, P., and Mungall, C. (2014). Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. *Journal of Biomedical Semantics*, 5(1):21.

Necsulea, A. (2020). Phylogenomics and genome annotation. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 4.1, pages 4.1:1–4.1:26. No commercial publisher | Authors open access book.

Nguyen, N. T. T., Vincens, P., Roest Crollius, H., and Louis, A. (2018). Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Research*, 46(D1):D816–D822.

4.3:4   REFERENCES

Robinson-Rechavi, M. (2020). Molecular evolution and gene function. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 4.2, pages 4.2:1–4.2:20. No commercial publisher | Authors open access book.

Scornavacca, C., Belkhir, K., Lopez, J., Dernat, R., Delsuc, F., Douzery, E. J., and Ranwez, V. (2019). Orthomam v10: scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian genomes. *Molecular Biology and Evolution*, 36(4):861–862.

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, 19(2):327–335.