**Application Notes**

**HOMECAT: consensus homologs mapping for interspecific knowledge transfer and functional genomic data integration.**

Zorzan Simone\(^1\), Lorenzetto Erika\(^1\), Ettorre Michele\(^1\), Pontelli Valeria\(^1\), Laudanna Carlo\(^2\), Buffelli Mario\(^2\)\(^3\)

\(^1\)Dept. of Neurological, Neuropsychological, Morphological and Motor Sciences, Section of Physiology, University of Verona, Strada le Grazie 8, 37134 Verona-Italy.

\(^2\)Centre for Biomedical Computing, University of Verona, Strada le Grazie 8, 37134 Verona-Italy

\(^3\)National Institute of Neuroscience—Italy

---

**ABSTRACT**

**Motivation:** Comparative studies are encouraged by the fast increase of data availability from the latest high-throughput techniques, in particular from functional genomic studies. Yet the size of datasets, the challenge of complete orthologs findings and not last, the variety of identification formats, make information integration challenging. With HOMECAT we aim to facilitate cross-species relationship identification and data mapping, by combining orthology predictions from several publicly available sources, a convenient interface for high-throughput data download and automatic identifier conversion into a Cytoscape plug-in, that provides both an integration with a large set of bioinformatics tools as well as a user friendly interface.

**Availability:** HOMECAT and the Supplementary Materials are freely available at http://www.cbmc.it/homecat/

**Contact:** simone.zorzan@univr.it

---

**1 INTRODUCTION**

The interpretation of large data sets, in particular when obtained from experiments using model systems, greatly benefits from the transfer of knowledge pertaining to phylogenetically related species. The studies on widely used model organisms such as mouse, fly or rat, yielded a large wealth of information, essential for the understanding of life complexity (Aitman et al., 2011, Loman et al., 2012, Schuster, 2008). The integration of functional genomic and proteomic expression profiles in the modelling of regulatory networks is a valuable approach in biology (Romero et al., 2012).

HOMECAT (HOmology Mapper for Enrichment and Comparative Analysis with Translation) is a plug-in for Cytoscape (Shannon et al., 2003) that allows cross-species data comparison and integration of high-throughput data with automatic identifier conversion. Orthology relationships can be difficult to identify and several approaches exhibit different sensitivity and specificity (Altenhoff and Dessimoz, 2009, Chen et al., 2007, Hulsen et al., 2006). HOMECAT, at present, can combine data from 4 homology data sources, to attain better specificity and increased sensitivity. BridgeDB (Van Iersel et al., 2010) usage allows to support 30 species and nearly 100 identifiers formats (71 from microarrays platforms). HOMECAT also interfaces Array Express ATLAS (Kapushesky et al., 2012), to download and integrate high throughput curated data.

---

**2 COMPARATIVE ANALYSES AND INTEGRATION WITH HIGH-THROUGHPUT DATA**

In HOMECAT, input species identifiers can be chosen between nodes attributes and are then converted to query the available homology data sources for the identification of orthologs in one or more destination species. Finally, the identifiers of the orthologs are converted to the selected output format. By default Homologene (Wheeler et al., 2007), OMA (Roth et al., 2008), Compara (Vilella et al., 2009) and OrthoMCL (Li et al., 2003) are supported; the former two are more specific and the other two are more sensitive (Altenhoff et al., 2009). OMA and Compara servers are queried directly, while Homologene (rel. 67) and OrthoMCL (ver. 5) data are accessed through our server in CBMC. When possible, a direct access to the data was used, to always guarantee the most updated results. All necessary conversions are performed by default through a remote BridgeDB server. BridgeDB database can be installed on any machine supporting Java (see BridgeDB website), and used by HOMECAT to reduce conversion times, particularly when hundreds of identifiers are processed. After the search phase, the user can decide whether to use only the orthologs confirmed by all the sources or those indicated by any of them, hence increasing the reliability or the coverage of the results. Orthologs can be used to enrich the input network, assigning their identifiers to specific attributes, or to create a novel network of metanodes, preserving original network connectivity.

---

\(^*\)To whom correspondence should be addressed.
Each metanode will contain an input species node, along with its orthologs, and can be expanded or collapsed. In the networks of metanodes, the colour of the border of each node always represents the input species data, while the internal colour of the node is always related to the ortholog data. A pie chart summarizes the orthologs data representation on collapsed metanodes (see Fig. 1A for details). When more than one network are created, node selections can be extended between networks, facilitating comparisons.

For each ortholog, an attribute summarizes the sources that support its identification. This attribute is useful to adjust the sensitivity attained by combining the results from different orthology sources. The user can filter less strongly supported orthologs and eventually remove them from the network.

External data can be loaded to the resulting network, either from a local file or from ATLAS, by selecting the experiment code and an experimental factor. When data from different orthologs are present, the average is added as an attribute to each metanode. A general scheme of this workflow is depicted in figure 1B

As an example, HOMECAT was used to compare microarray data after optic nerve crush in three publicly available datasets from zebrafish, mouse and rat samples. The optic nerve crush is a common model to study the regeneration in the central nervous system. After crush, the optic nerve regenerates in zebrafish, while the axonal recovery is absent in mammals. The results highlighted differently regulated genes in fish and mammals, consistent with literature, and showed how the combination of functional genomic and regulatory data analysis can contribute to the identification of putative key factors in comparative biological studies (see the manual in the Supp. Materials).

3 FEATURES AND EXTENSIBILITY
Orthologs resulting from HOMECAT multiple queries can be saved and loaded in “.hcd” files. Data mapped to attributes can be saved along with Cytoscape networks. A programming interface allows the development of Java classes called components, to combine additional homology data sources, that are automatically loaded upon HOMECAT launch (See Suppl. Materials).