BMC Bioinformatics

RESEARCH

Open Access

CrossMark

Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles

Zakhar Sergeevich Mustafin[1†], Sergey Alexandrovich Lashin[1,2*†], Yury Georgievich Matushkin[1], Konstantin Vladimirovich Gunbin[1] and Dmitry Arkadievich Afonnikov[1,2]

Abstract

Background: There are many available software tools for visualization and analysis of biological networks. Among them, Cytoscape (http://cytoscape.org/) is one of the most comprehensive packages, with many plugins and applications which extends its functionality by providing analysis of protein-protein interaction, gene regulatory and gene co-expression networks, metabolic, signaling, neural as well as ecological-type networks including food webs, communities networks etc. Nevertheless, only three plugins tagged 'network evolution' found in Cytoscape official app store and in literature. We have developed a new Cytoscape 3.0 application Orthoscape aimed to facilitate evolutionary analysis of gene networks and visualize the results.

Results: Orthoscape aids in analysis of evolutionary information available for gene sets and networks by highlighting: (1) the orthology relationships between genes; (2) the evolutionary origin of gene network components; (3) the evolutionary pressure mode (diversifying or stabilizing, negative or positive selection) of orthologous groups in general and/or branch-oriented mode. The distinctive feature of Orthoscape is the ability to control all data analysis steps via user-friendly interface.

Conclusion: Orthoscape allows its users to analyze gene networks or separated gene sets in the context of evolution. At each step of data analysis, Orthoscape also provides for convenient visualization and data manipulation.

Keywords: Cytoscape plugin, Ortholog, Paralog, Metabolic pathway, Gene regulatory network, Evolution, Phylostratigraphy, Evolution

Background

Biological networks arise in completely all fields of modern biology gathering both 'real' (experimental data etc.) and virtual (modeling and simulation data) biological information [1–6]. There are software packages to work with biological networks with less or more biological specialization, availability and interactivity [7–11]. Among them, Cytoscape [7, 12] is one of the most comprehensive tools for performing all-round analysis of biological networks. There are many plugins which extend the functionality of Cytoscape by providing visualization and analysis of protein-protein interaction networks [13, 14], including PINA4MS (http://apps.cytoscape.org/apps/pina4ms), Strongest Path (http://apps.cytoscape.org/apps/strongestpath), gene regulatory [15, 16] and gene co-expression [17, 18], metabolic

* Correspondence: lashin@bionet.nsc.ru
†Equal contributors
[1]Institute of Cytology and Genetics SB RAS, Lavrentiev Avenue 10, Novosibirsk 630090, Russia
[2]Novosibirsk State University, Pirogova st. 2, Novosibirsk 630090, Russia

[18–20], signaling [21, 22] as well as ecological-type networks including food webs, communities network and others. In spite of such great diversity, evolution-oriented plugins are in short supply: just three plugins tagged 'network evolution' (even not 'biological evolution') in Cytoscape official app store and in literature. These include ANIMO [23], TieDIE, NetworkEvolution [24]) and a couple of plugins concerning orthology analysis (HOMECAT [25] and OrthoNets [26]).

KEGG (Kyoto Encyclopedia of Genes and Genomes, www.kegg.jp) is a set of databases containing biological information of various types, such as genes, genomes, protein interaction networks, pathway maps and many others. KEGG Pathway is a collection of manually curated maps representing the molecular interaction and reaction networks for various biological processes, such as metabolism, genetic information processing, human diseases, etc. Number of tools have been developed to operate with KEGG pathway maps including CyKEGG-Parser [27], ANIMO [28] and SIREN [16].

In this work, we have developed a Cytoscape application (plugin) Orthoscape aimed to analyze evolutionary information in the gene sets and networks: (1) the orthology relationships between genes; (2) the evolutionary origin of gene network components; (3) the evolutionary regime (diversifying or stabilizing, negative or positive selection) of orthologous groups in general and/ or branch-oriented mode. See Additional file 1 for the Orthoscape jar-file and Additional file 2 for manual.

Clear evolutionary ideas underlying the Orthoscape application and its ability to control all data analysis steps via user-friendly interface will aid biologists in better understanding the factors of evolution and functioning of gene network.

Methods

Orthoscape implemented in Java 1.8 language to use with Cytoscape 3.0 version or higher. Homology, taxonomy, protein domains data, as well as nucleotide and amino acid sequences are extracted from KEGG databases, hence, requiring Internet connection. All downloaded data could be stored in local Orthoscape database, which may require up to several GB hard drive space and later may both speed up the work and decrease dependence on connection to the KEGG.

The orthoscape workflow

Figure 1 depicts a workflow of Orthoscape, with a gene network (or just a set of genes) used as an input which may be presented in several ways:

1) Using the CyKEGGParser plugin (http://apps.cytoscape.org/apps/cykeggparser), a network may be imported directly from the KEGG database. Once it is imported, it is ready to work with the Orthoscape.
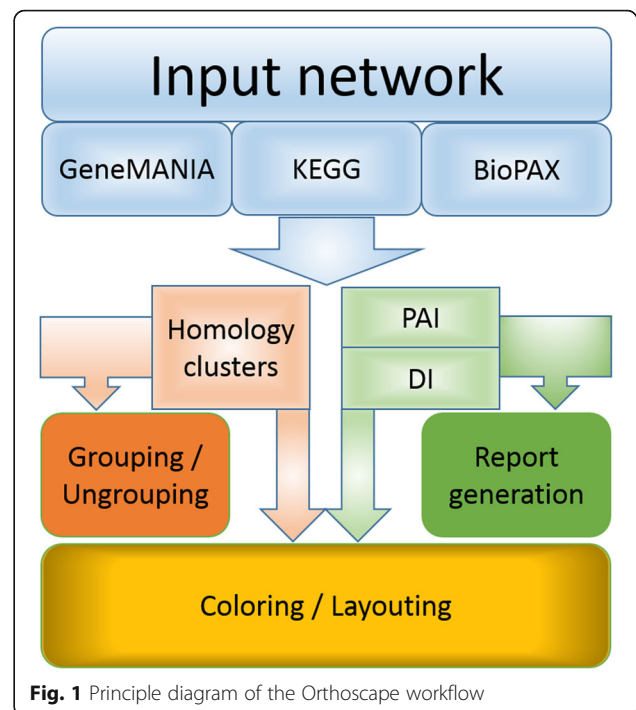

Fig. 1 Principle diagram of the Orthoscape workflow

2) Using the GeneMANIA plugin (http://apps.cytoscape.org/apps/genemania), a network may be reconstructed on the set of necessary genes. The obtained network should be converted using "Orthoscape - > convert GeneMANIA Network" option.

3) Using the CyPath2 plugin (http://apps.cytoscape.org/apps/cypath2) (CyPathwayCommons), the user may choose a network from the list of filtered networks (by presence of necessary genes) in BioPAX format. The obtained network should be converted using "Orthoscape - > convert BioPAX Network" option.

Then gene network is analyzed by the Orthoscape core, which performs the following tasks: (1) reconfiguration (layouting and grouping) of gene networks on the base of sequences homology; (2) coloring and grouping of genes on the base of evolutionary characteristics such as Phylostratigraphic Age Index (PAI) and Divergence Index (DI) (3) report generation. As a result, the information on genes, their interactions and evolutionary characteristics is represented in compact and convenient manner. It provides users a better understanding of gene networks structure and enables to perform analysis of network evolution.

Implementation

Analysis of homologous sequences

For each gene from a gene network (or gene set), the Orthoscape generates lists of potential homologs (paralogs in case of gene network and/or orthologs in case of

set of genes from various genomes) according to SW-Score (Smith-Waterman score [29]) and identity values set up by the user. These operations are performed using requests to KEGG database via REST API protocol (http://www.kegg.jp/kegg/docs/keggapi.html). After that, Orthoscape joins genes into clusters using one of three possible types of similarity measures for the sequence comparison: (1) SW-Score; (2) identity value; (3) domain composition from the proteins previously filtered by SW-Score and/or identity.
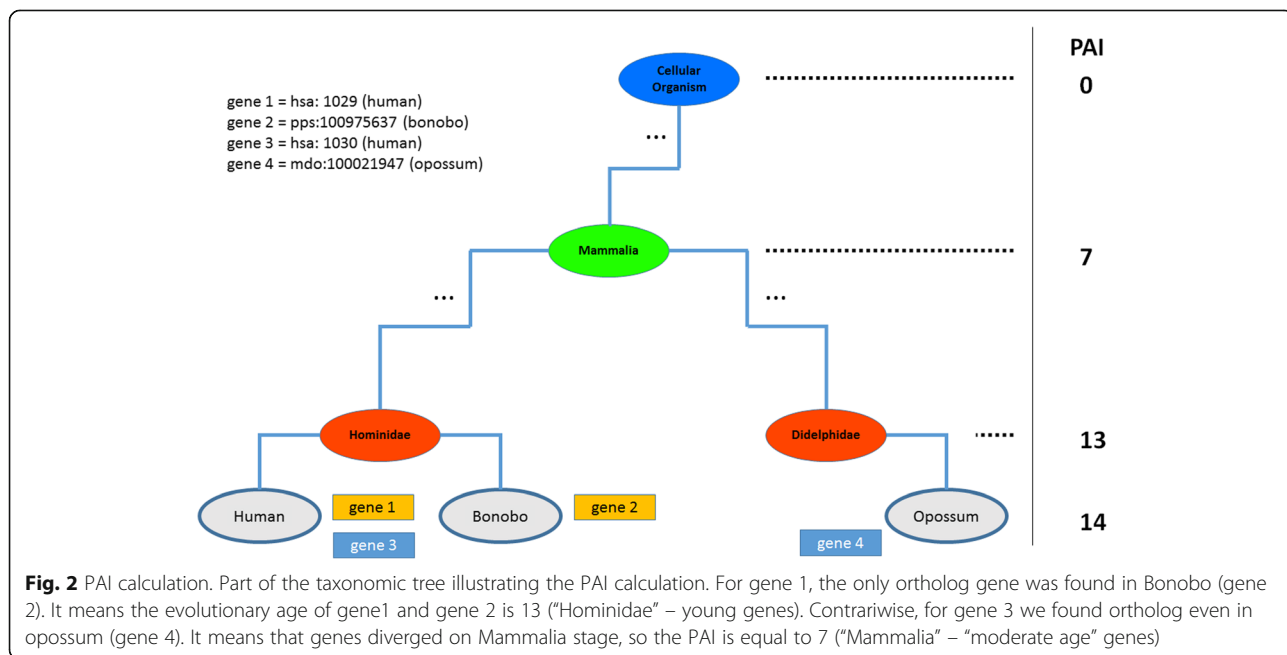
The similarity according to domain composition could be calculated in either simple or detailed way. Simple one implies the analysis of domain frequencies among all potential filtered homologs. The user may set the 'threshold' number T of domains to be the same in gene and its potential homolog. At first, we construct the ordered list of most presented domains among all homologs found. If a homolog contains first T domains from this list, then it is accepted and the homologous genes join. The detailed analysis additionally allows user to choose the particular domains essential for him/her. If a potential homolog does not contain any of those domains, it is deleted from the list.

Once genes are joined into clusters according to their similarity, the Orthoscape allows the following additional analysis of the gene network: (1) group/ungroup each cluster of homologous genes (Group/Ungroup the homologs options); (2) color network nodes using either heatmap or blue-red gradient schemes according to evolutionary characteristics of genes (cluster memberships or evolutionary indices).

Analysis of gene evolutionary indices

The Orthoscape calculates two evolutionary characteristics of genes. The first characteristic is the phylostratigraphic age index (PAI). The PAI indicates the "evolutionary age" of a gene [30, 31]. To calculate the PAI, the Orthoscape uses KEGG Organisms database for taxonomic trees. It performs a search of orthologous genes (using sequence similarity thresholds mentioned above), populates the tree of species these genes belong to and then analyses the resulting tree. Ranged between 0 and N (where N is a number of phylums between root and a species in the taxonomic tree; for instance, $N = 14$ for human), this index shows the level of the last common ancestor node in the whole taxonomic tree, containing at least one species from the list of species possessing orthologs of a gene under analysis (Fig. 2). Consequently, the "Cellular Organisms" node (root) has PAI = 0, "Eukaryota" node has PAI = 1, etc.; "Homo" node has PAI =14).

The second characteristic is the Divergence Index (Ka/Ks index, DI) of a gene [32, 33]. It indicates the influence of natural selection on gene evolution. To calculate the DI, Orthoscape extracts nucleotide sequences and amino acid sequences of a gene/protein under analysis and its nearest ortholog (from the closest species). Amino acid sequences are retrieved via KEGG REST API and then aligned using the Needleman–Wunsch algorithm [34] using NW-align code (http://zhanglab.ccmb.med.umich.edu/NW-align/). The codon alignment is obtained using the protein alignment; Ka/Ks ratio is calculated using Java code from the program Ka/Ks calculator [35] (http://kakscalculator.fumba.me/index.jsp). In general, the

Fig. 2 PAI calculation. Part of the taxonomic tree illustrating the PAI calculation. For gene 1, the only ortholog gene was found in Bonobo (gene 2). It means the evolutionary age of gene1 and gene 2 is 13 ("Hominidae" – young genes). Contrariwise, for gene 3 we found ortholog even in opossum (gene 4). It means that genes diverged on Mammalia stage, so the PAI is equal to 7 ("Mammalia" – "moderate age" genes)

DI may be calculated through the analysis of any orthologs found, however it is recommended to use the closest taxa for DI calculation [30]. In Orthoscape, the user may specify the maximum distance between the reference species and other nodes on the species tree. Besides, for human gene networks we provide an additional option that allows explicit specifying of the closest species – Pan troglodytes (chimpanzee), Pongo abelii (sumatran orangutan), Pan Paniscus (bonobo) and working with orthologs belonging to these only organisms.

Additionally, Orthoscape reports phylogenetic profile [36] for each gene in the network. This is a table with columns corresponding to genes represented in the metabolic network, and rows corresponding to organisms which genomes contain at least one ortholog of the network genes. The cell in column i and row j contain '+' if in the organism j exists at least one ortholog of gene i, or '-'otherwise.

Report generator

Orthoscape generates reports with all calculated characteristics, indices, statistics etc. obtained for one or several gene networks. Report is generated in HTML format and also includes directory with additional image files. Report contains PAI histograms and tables of genes sorted by DI. Analyzing PAI histograms, the Orthoscape reports "average evolutionary age" of a gene network and/or gene set. First of all, it concerns to the mean and median PAI values for all genes in network/set. Once the user tried different SW-Score and identity parameters in his/her analysis, all of these attempts are included into the report.

Orthoscape provides both Gene Set and Network PAI statistics. The first one is calculated with the simple formula:

$$\text{Gene Set PAI} = \frac{\sum_{\text{genes}} PAI}{N}$$

Where N is the number of genes. Thus, if all genes are ancient (PAI = "Cellular organisms"), then the statistics value would be 0. If all genes are "young" (PAI = "Homo"), then the statistics value would be 14. Therefore, the more is the mean PAI value, the "younger" genes are in the gene set.

Network PAI statistics additionally takes into account the network topology, more specifically connectivity of nodes (i.e., their degrees). For example, if one node is connected with two edges, its degree would be 2, if it has no connections, the degree would be 0 etc. The modified formula is used:

$$\text{Network PAI} = \frac{\sum_{\text{nodes}} (PAI * d)}{2N}$$

Where d is the node degree. It makes highly connected nodes, which we believe are more important in gene networks functioning than lowly connected ones to

have higher contribution to the "evolutionary age" of a gene network.

There are also median, oldest and youngest statistics for PAI values in a network. Finally, the total number of orthologs analyzed is reported too. An example of statistical analysis from report is shown below in Results section.

Visualization of results

Analyzed network may be colored using two schemes: (1) Heatmap scheme colors "young" genes in red and "old" in blue. Intermediate colors are yellow, green, cyan; (2) Blue-white-red gradient scheme. Each style may be used for PAI, DI and homologous groups.

Results and discussion

As an example of the Orthoscape application, we have analyzed two pathways from KEGG database related to steroid metabolism: steroid biosynthesis pathway and steroid hormone biosynthesis pathway.

Steroids, such as cholesterol, are synthesized in almost all eukaryotic cells, which use these triterpenoid lipids to control the fluidity and flexibility of their cell membranes. Sterols also play a key role in such eukaryotic features as phagocytosis. In KEGG, steroid biosynthesis pathway (ko00100) follows the terpenoid backbone biosynthesis pathway and uses farnesyl diphosphate as input metabolite. A number of sterols produced as a result of this pathway: zymosterol, fecosterol, episterol, ergosterol and others. Interestingly, few bacteria can synthesize sterols [37], however, phylogenetic analysis demonstrated that they likely acquired homologs of enzymes of the sterol pathway via ancient horizontal gene transfer from eukaryotes [38].

An important metabolite produced within the sterol biosynthesis pathway is cholesterol, an essential structural component of all animal cell membranes and precursor for the biosynthesis of steroid hormones [39, 40]. Five major classes of steroid hormones include testosterone, progesterone and estradiol, which are known as sex-steroids, and cortisol/corticosterone and aldosterone, which are referred to as corticosteroids [39]. Steroid hormones are synthesized from cholesterol through a common precursor steroid, pregnenolone, which is formed by the enzymatic cleavage of a 6-carbon side-chain of the 27- carbon cholesterol molecule by the cytochrome P450 side-chain cleavage enzyme [39]. Their biosynthesis is described in KEGG by steroid hormone biosynthesis pathway (ko00140).

Visualization of these two pathways is shown in Figs. 3 and 4. The visualization demonstrates the PAI for each gene in the metabolic network graph, from blue (small PAI, older genes) to green (large PAI, younger genes). For example, for delta24-sterol reductase (gene ID
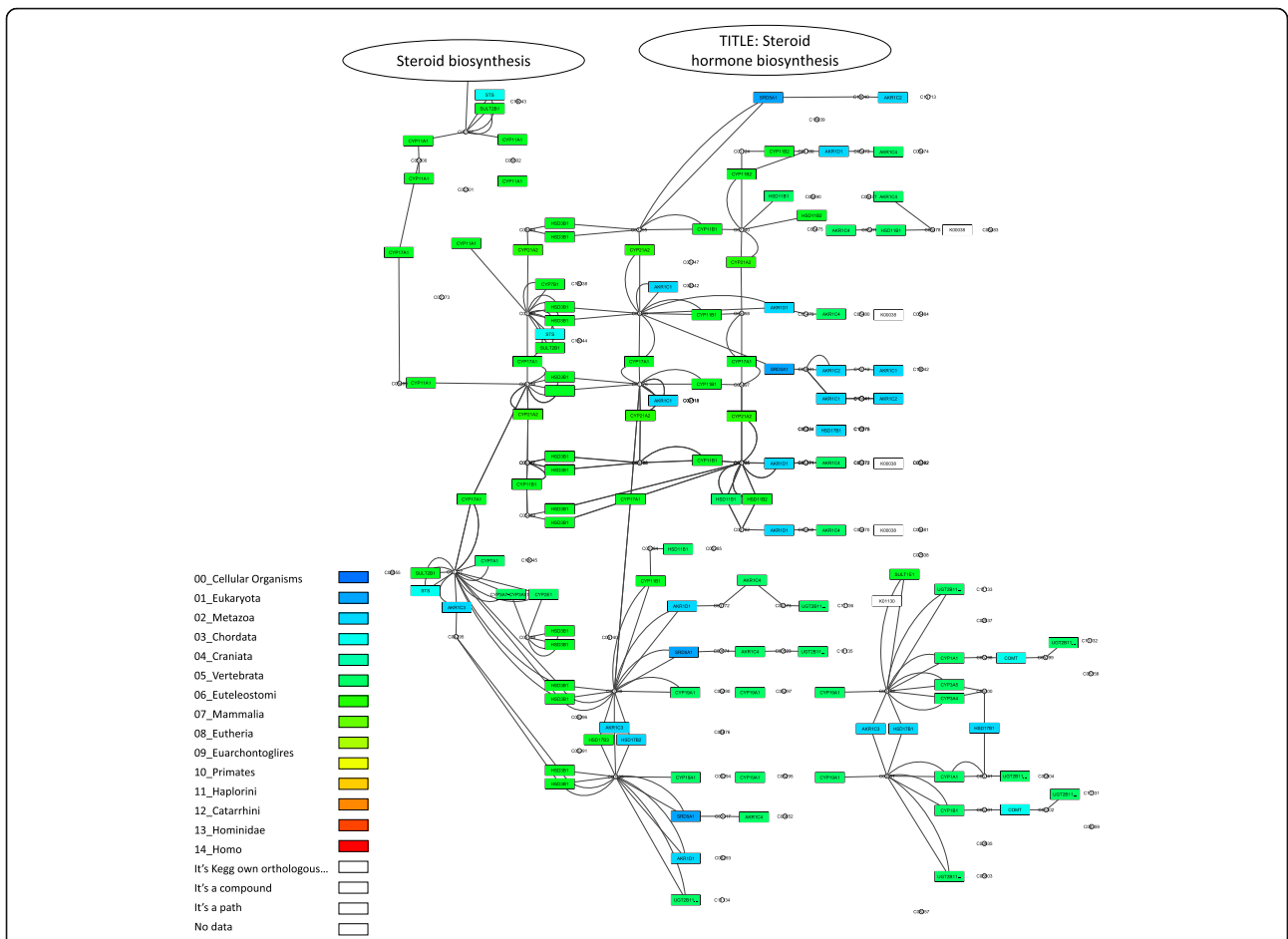
Fig. 3 KEGG steroid biosynthesis pathway. Visualization of the KEGG steroid biosynthesis pathway (ko00100) by the Orthoscape application using PAI heatmap color scheme. Gene node colors correspond to PAI values from smaller (PAI = 0, older genes, dark blue color) to larger (PAI = 5, young genes, green color)

DHCR24, EC:1.3.1.72, 1.3.1.-), which participates in sterol biosynthesis pathway (Fig. 3), the PAI is equal to 0 ('Cellular Organisms' taxonomic group, dark blue color). For vitamin D 25-hydroxylase (gene ID CYP24A1, EC:1.14.14.24) in the same pathway (Fig. 3), the PAI value is 5 ('Vertebrata' taxonomic group, green color). It could be seen from the comparison of these two pathway diagrams (Figs. 3 and 4) that the steroid hormone biosynthesis pathway contains larger fraction of younger genes as compared to the steroid biosynthesis pathway. It is demonstrated by comparison of the gene set PAI values for these two pathways for different orthology detection thresholds (Fig. 5). The PAI values for steroid hormone biosynthesis pathway are higher at all thresholds except in the case of identity is equal to 1. These results demonstrate that most genes for this pathway diverged from their ancestors and acquired their new function later as compared to genes from steroid biosynthesis pathway.

The obtained results of comparison of the gene 'ages' for the above pathways are consistent with the current knowledge of the eukaryotic evolution. Steroids participate in

the formation of membranes, which are basal cellular structures in eukaryotes. Thus, sterol biosynthesis is a fundamental feature of eukaryotic cells and it is generally accepted that the pathway of sterol biosynthesis appeared after the emergence of oxygenic photosynthesis and the oxygenation of the atmosphere and oceans (between 2.7 and 2.4 Ga) [41].

Steroid hormones regulate diverse physiological functions such as reproduction, blood salt balance, maintenance of secondary sexual characteristics, response to stress, neuronal function, various metabolic processes [40] and response to environmental factors [42]. In mammals, they are synthesized in a specific set of organs: ovary (granulosa cells, luteal cells), testis (Leydig cells), adrenal gland (zona glomerulosa and zona reticularis cells), placenta and brain (neurons, glial and Purkinje cells) [40]. The analysis of the evolution of the steroid hormone receptors demonstrates, that they are common to vertebrates [43], but might have originated before the divergence of vertebrates, suggesting the ancient origin of some steroid hormone systems (as was shown for the estrogen signaling [44]). It should be noted, however, that hormones are the important source of

Fig. 4 KEGG steroid hormone biosynthesis pathway. Visualization of the KEGG steroid hormone biosynthesis pathway (ko00140) by the Orthoscape application using PAI heatmap color scheme. Gene node colors correspond to PAI values from smaller (PAI = 1, older genes, blue color) to larger (PAI = 7, young genes, green color)
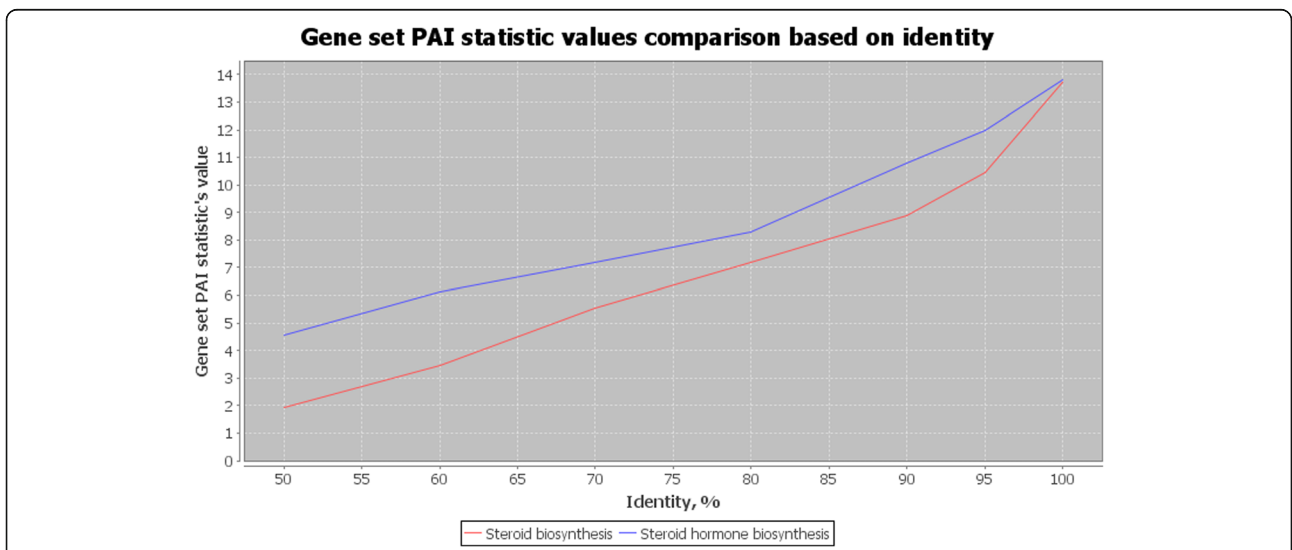

Fig. 5 Overall PAI comparison for steroid pathways. Comparison of the dependence of the PAI indices (Y axis) for steroid biosynthesis pathway (red line) and steroid hormone biosynthesis pathway (blue line) with respect to identity threshold for gene orthology detection (X axis)

the organism's phenotypic plasticity [45]. For example, placenta, the evolutionary innovation of mammals, demonstrates amazing diversity [46]. These phenotypic innovations should require evolutionary changes in the hormone biosynthesis pathway, for example, increasing its complexity to produce different types of hormones. This is the likely reason for the relatively modest age of the genes involved in the steroid hormone biosynthesis.

Figure 6 shows the histograms for the distribution of PAI values for genes from both networks (SW-Score = 500, identity = 0.5, 0 domains). On histograms, one could see that steroid hormone biosynthesis pathway (Fig. 6a) includes larger fraction of genes with PAI from 5 ('Vertebrata') to 7 ('Mammalia') as compared to steroid biosynthesis pathway (Fig. 6b). Orthoscape reports the following statistics for steroid biosynthesis pathway: Gene set PAI = 4.548; Network PAI = 4.828; Median taxon = Vertebrata; Oldest taxon = Eukaryota; Youngest taxon = Mammalia. Values for steroid biosynthesis pathway were as follows: Gene set PAI = 1.941; Network PAI = 1.4; Median taxon = Metazoa; Oldest taxon = Cellular Organisms; Youngest taxon = Vertebrata.

Note, that data from Figs. 5 and 6 and PAI statistics could be found in HTML reports output by Orthoscape. Global report (PAI values dependence on the sequence similarity thresholds for a set of networks, see Fig. 5) and networks specific reports (PAI distribution histograms, see Fig. 6) could be navigated using hyperlinks. The original data for the plots in text format and links to networks information at the KEGG web-site are provided.

Figure 7 demonstrates the additional layout scheme for the networks based on PAI identity. This type of network layout shown for the steroid hormone biosynthesis pathway: genes are located on the circles corresponding to different PAI values (genes with identical PAI are located on the same circle); non-gene nodes are located on the separate circles. It helps to identify genes with the same "age" easily.

The gene network layout based on the grouping of genes with respect to their membership in different homology clusters is demonstrated in (Fig. 8) for the steroid biosynthesis network. There were 14 homologous clusters found for this network. By expanding the node on this type layout user can obtain the list
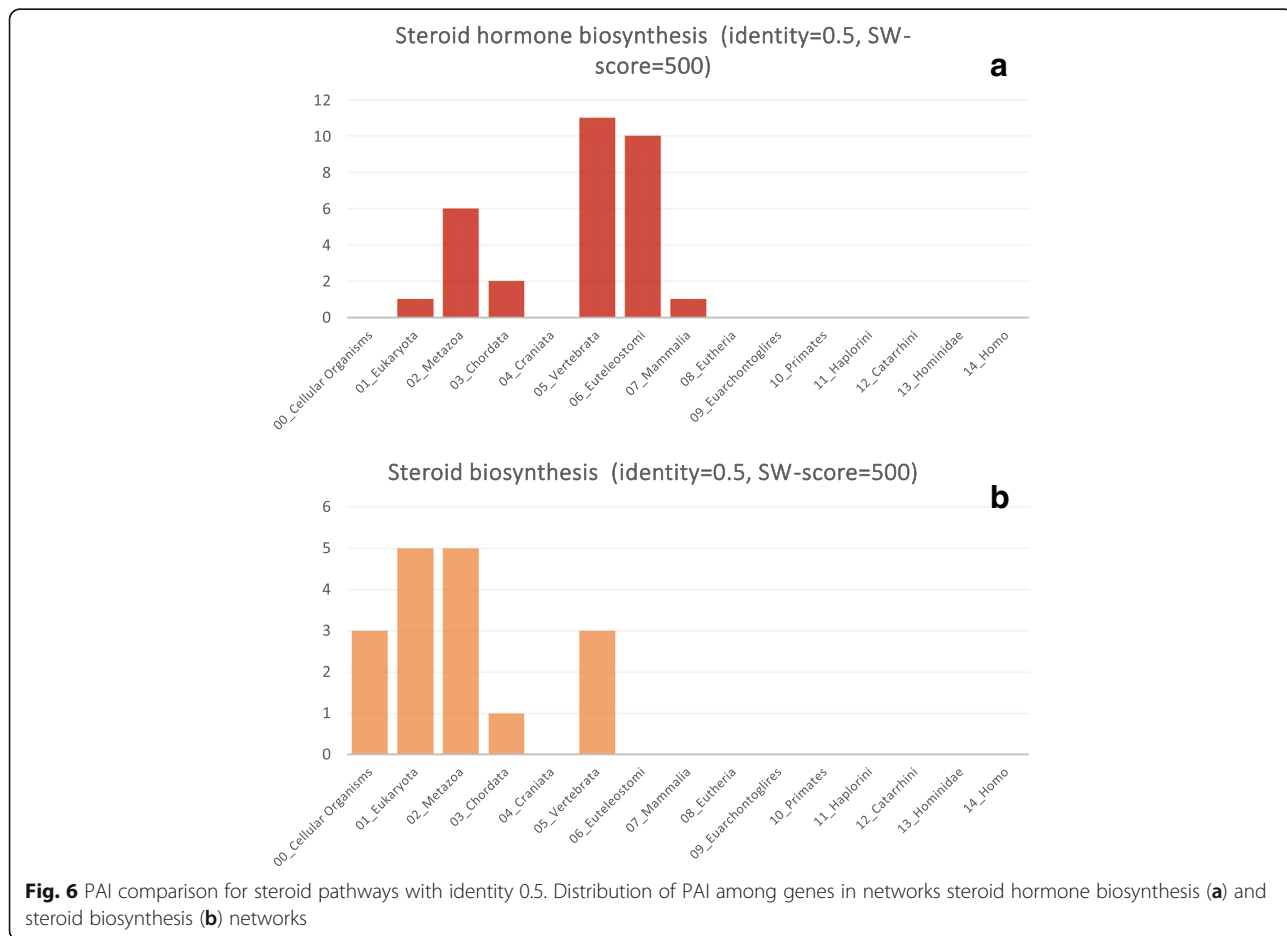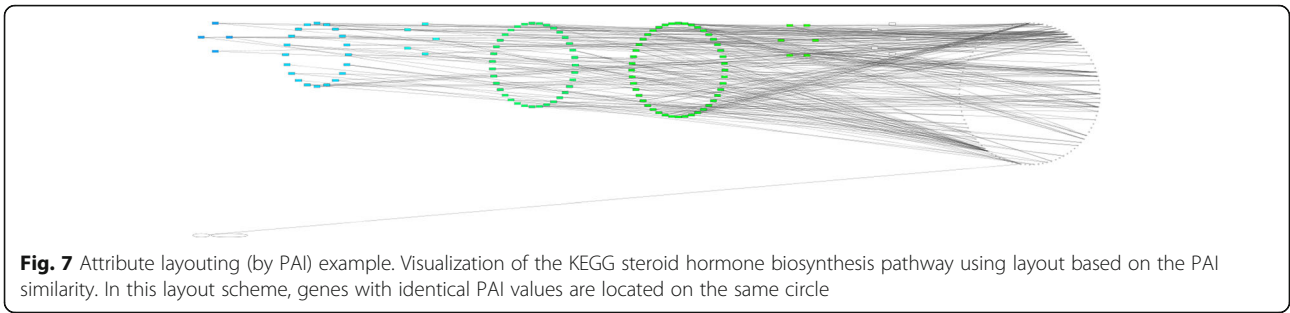

Fig. 6 PAI comparison for steroid pathways with identity 0.5. Distribution of PAI among genes in networks steroid hormone biosynthesis (**a**) and steroid biosynthesis (**b**) networks

Fig. 7 Attribute layouting (by PAI) example. Visualization of the KEGG steroid hormone biosynthesis pathway using layout based on the PAI similarity. In this layout scheme, genes with identical PAI values are located on the same circle

of genes in the cluster. Detailed report generated by the Orthoscape on this comparison is presented in the Additional file 3.

Conclusion

In summary, Orthoscape application is a tool for analysis of gene networks/sets. It allows to search for homologs (orthologs and/or paralogs), to perform phylostratigraphic analysis of genes and to investigate the divergence. User is aided in better understand of evolution of genes and their (sub)networks under selective pressure.

Fig. 8 Grouping to homology clusters example. Visualization of the KEGG steroid biosynthesis pathway using layout based on the homology cluster membership (all homologous genes are joined in the same nodes)

At each step of data analysis, Orthoscape also provides for convenient visualization and data manipulation.

Additional files

Additional file 1: Orthoscape-1.0. The plugin itself. (JAR 105 kb)

Additional file 2: Orthoscape manual. The manual with basics of Orthoscape. (DOCX 133 kb)

Additional file 3: Reports. 7z archive containing the Orthoscape reports example. (7Z 381 kb)

Abbreviations
API: Application program interface; DI: Divergence index;
PAI: Phylostratigraphic age index; REST: Representational state transfer

Availability of data and materials
Project name: Orthoscape
Operating system(s): Platform independent
Programming language: Java version 1.8
Other requirements: Cytoscape 3.0+
License: GPL
Source Code: https://github.com/ZakharM/Orthoscape
Any restrictions to use by non-academics: none

Authors' contributions
MZS and LSA developed and implemented Orthoscape. GKV suggested the initial idea of the project. MYUG, GKV, ADA and LSA formulated requirements and use cases for Orthoscape. MZS performed analysis of the networks. LSA and ADA coordinated the writing of the paper. All authors have read and approved the final manuscript.
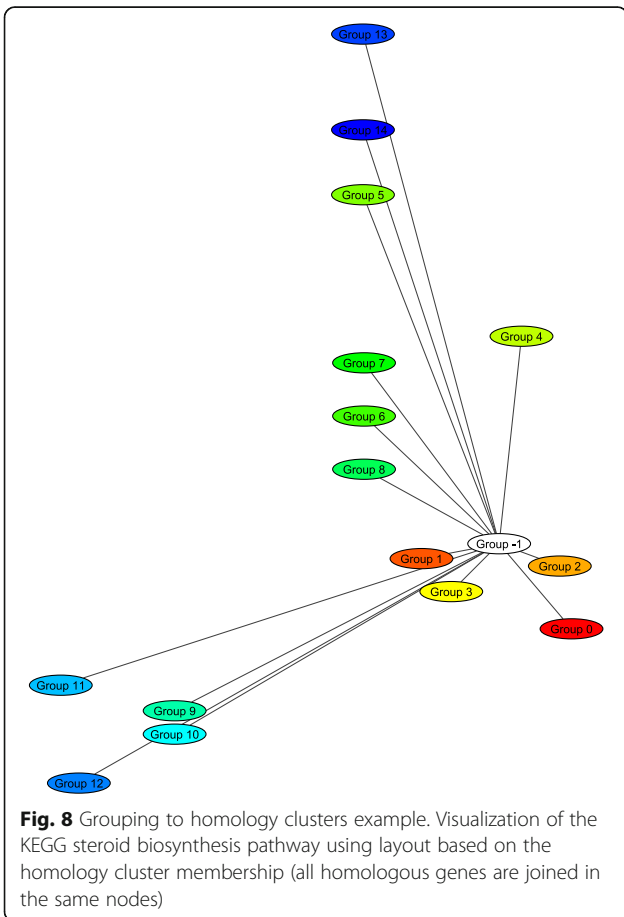
Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

References

1. Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes Dev. 2007;1010–1024.
2. Shade A, Teal TK. Computing workflows for biologists: a roadmap. PLoS Biol. 2015;13:e1002303.
3. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12:56–68.
4. Mitra K, Carvunis A-R, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. Nat Rev Genet. 2013;14:719–32.
5. Goh WWB, Lee YH, Chung M, Wong L. How advancement in biological network analysis methods empowers proteomics. Proteomics. 2012;12:550–63.
6. Pujol A, Mosca R, Farrés J, Aloy P. Unveiling the role of network and systems biology in drug discovery. Trends Pharmacol Sci. 2010;31:115–23.
7. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
8. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server):W214–20.
9. Thomas S, Bonchev D. A survey of current software for network analysis in molecular biology. Hum Genomics. 2010;4:353.
10. Mayerich D, Bjornsson C, Taylor J, Roysam B. NetMets: software for quantifying and visualizing errors in biological network segmentation. BMC Bioinformatics. 2012;13 Suppl 8:S7.
11. Demenkov PS, Ivanisenko TV, Kolchanov NA, Ivanisenko VA. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. In Silico Biol. 2012;11:149–61.
12. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang P-L, Lotia S, Pico AR, Bader GD, Ideker T. A travel guide to cytoscape plugins. Nat Methods. 2012;9:1069–76.
13. Winterhalter C, Nicolle R, Louis A, To C, Radvanyi F, Elati M. PEPPER: cytoscape app for protein complex expansion using protein-protein interaction networks. Bioinformatics. 2014;30:3419–20.
14. Micale G, Continella A, Ferro A, Giugno R, Pulvirenti A. GASOLINE: a cytoscape app for multiple local alignment of PPI networks. F1000Res. 2014;3:140.
15. Politano G, Benso A, Savino A, Di Carlo S. ReNE: a cytoscape plugin for regulatory network enhancement. PLoS One. 2014;9:e115585.
16. Montojo J, Khosravi P, Gazestani VH, Bader GD. SIREN Cytoscape plugin: Interaction Type Discrimination in Gene Regulatory Networks. 2015.
17. Tzfadia O, Diels T, De Meyer S, Vandepoele K, Aharoni A, Van de Peer Y. CoExpNetViz: comparative co-expression networks construction and visualization tool. Front Plant Sci. 2016;6.
18. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, Sartor MA, Stringer KA, Jagadish HV, Burant C, Athey B, Omenn GS. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. Bioinformatics. 2012;28:373–80.
19. Duren W, Weymouth T, Hull T, Omenn GS, Athey B, Burant C, Karnovsky A. MetDisease–connecting metabolites to diseases via literature. Bioinformatics. 2014;30:2239–41.
20. Jourdan F, Breitling R, Barrett MP, Gilbert D. MetaNetter: inference and visualization of high-resolution metabolomic networks. Bioinformatics. 2008;24:143–5.
21. Ritz A, Poirel CL, Tegge AN, Sharp N, Simmons K, Powell A, Kale SD, Murali T. Pathways on demand: automated reconstruction of human signaling networks npj. Syst Biol Appl. 2016;2:16002.
22. Poirel CL, Rodrigues RR, Chen KC, Tyson JJ, Murali TM. Top-down network analysis to drive bottom-up modeling of physiological processes. J Comput Biol. 2013;20:409–18.
23. Schivo S, Scholma J, Karperien HBJ, Langerak R, van de Pol JC, Post JN. ANIMO: a tool for modeling biological pathway dynamics. J Tissue Eng Regen Med. 2014;8(Suppl):54–5.
24. Wozniak M, Tiuryn J, Dutkowski J. MODEVO: exploring modularity and evolution of protein interaction networks. Bioinformatics. 2010;26:1790–1.
25. Zorzan S, Lorenzetto E, Ettorre M, Pontelli V, Laudanna C, Buffelli M. HOMECAT: consensus homologs mapping for interspecific knowledge transfer and functional genomic data integration. Bioinformatics. 2013;29:1574–6.
26. Hao Y, Merkoulovitch A, Vlasblom J, Pu S, Turinsky AL, Roudeva D, Turner B, Greenblatt J, Wodak SJ. OrthoNets: simultaneous visual analysis of orthologs and their interaction neighborhoods across different organisms. Bioinformatics. 2011;27:883–4.
27. Nersisyan L, Samsonyan R, Arakelyan A. CyKEGGParser: tailoring KEGG pathways to fit into systems biology analysis workflows. F1000Res. 2014;3:145.
28. Schivo S, Scholma J, Karperien M, Post JN, van de Pol J, Langerak R. Setting parameters for biological models with ANIMO. Electron Proc Theor Comput Sci. 2014;145:35–47.
29. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol. 1981;147:195–7.
30. Quint M, Drost H-G, Gabel A, Ullrich KK, Bönn M, Grosse I. A transcriptomic hourglass in plant embryogenesis. Nature. 2012;490:98–101.
31. Domazet-Lošo T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature. 2010;468:815–8.
32. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 2000;15:496–503.
33. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18:486–7.
34. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 1970;48:443–53.
35. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. Genomics Proteomics Bioinformatics. 2006;4:259–63.
36. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc Natl Acad Sci. 1999;96:4285–8.
37. Bode HB, Zeggel B, Silakowski B, Wenzel SC, Reichenbach H, Müller R. Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium Stigmatella aurantiaca. Mol Microbiol. 2003;47:471–81.
38. Desmond E, Gribaldo S. Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. Genome Biol Evol. 2010;1:364–81.
39. Payne AH, Hales DB. Overview of Steroidogenic Enzymes in the Pathway from Cholesterol to Active Steroid Hormones. Endocr Rev. 2004;25:947–70.
40. Hu J, Zhang Z, Shen W-J, Azhar S. Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones. Nutr Metab (Lond). 2010;7:47.
41. Summons RE, Bradley AS, Jahnke LL, Waldbauer JR. Steroids, triterpenoids and molecular oxygen. Philos Trans R Soc B Biol Sci. 2006;361:951–68.
42. Hau M. Regulation of male traits by testosterone: implications for the evolution of vertebrate life histories. BioEssays. 2007;29:133–44.
43. Baker ME. Evolution of adrenal and sex steroid action in vertebrates: a ligand-based mechanism for complexity. BioEssays. 2003;25:396–400.
44. Thornton JW. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science (80-. 2003;301:1714–7.
45. Dufty A. Hormones, developmental plasticity and adaptation. Trends Ecol Evol. 2002;17:190–6.
46. Strauss JF. Placental steroid hormone synthesis: unique features and unanswered questions. Biol Reprod. 1996;54:303–11.