W470–W474 *Nucleic Acids Research, 2011, Vol. 39, Web Server issue* *Published online 6 June 2011*
doi:10.1093/nar/gkr408

Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing

Rubén Sánchez[1], François Serra[1], Joaquín Tárraga[2], Ignacio Medina[2], José Carbonell[2], Luis Pulido[2], Alejandro de María[2], Salvador Capella-Gutíerrez[3], Jaime Huerta-Cepas[3], Toni Gabaldón[3], Joaquín Dopazo[2] and Hernán Dopazo[1,*]

[1]Evolutionary Genomics Lab, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), [2]Functional Genomics Lab Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Autopista del Saler, 16-3, 46013 Valencia and [3]Comparative Genomics Group, Bioinformatics & Genomics Programme Centre for Genomic Regulation (CRG), UPF, Doctor Aiguader, 88, 08003 Barcelona, Spain

Received February 7, 2011; Revised May 3, 2011; Accepted May 7, 2011

ABSTRACT

Phylemon 2.0 is a new release of the suite of web tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. It has been designed as a response to the increasing demand of molecular sequence analyses for experts and non-expert users. Phylemon 2.0 has several unique features that differentiates it from other similar web resources: (i) it offers an integrated environment that enables evolutionary analyses, format conversion, file storage and edition of results; (ii) it suggests further analyses, thereby guiding the users through the web server; and (iii) it allows users to design and save phylogenetic pipelines to be used over multiple genes (phylogenomics). Altogether, Phylemon 2.0 integrates a suite of 30 tools covering sequence alignment reconstruction and trimming; tree reconstruction, visualization and manipulation; and evolutionary hypotheses testing.

INTRODUCTION

Phylogenetic analysis and model-based hypothesis testing are essential elements in current molecular evolution studies (1,2). Web servers for phylogenetic and evolutionary analyses range from those running single programs to those integrating multiple tools. Among the first are servers that execute multiple sequence alignment (MSA) tools such as ClustalW (3) (http://www.ebi.ac.uk/clustalw/), or sophisticated programs to test molecular adaptation such as the HyPhy environment (4) (http://www.datamonkey.org/). In the second category, resources such as the 'Pasteur server' (e.g. see http://bioweb.pasteur.fr/seqanal/phylogeny/intro-uk.html), Phylogeny.fr (http://www.phylogeny.fr/) (5), CIPRES (http://www.phylo.org/) and Phylemon (6) developed the concept of integrated platforms, which implement different phylogenetic analysis programs in a single server.

Phylemon was originally developed in 2007 as a web server providing a common framework to run the most frequent analyses on DNA and protein sequences from a phylogenetic and evolutionary perspective. Phylemon 2.0 covers a wide, yet selected, range of programs, integrating over 30 different tools for phylogenetic and evolutionary analyses. Phylemon 2.0 has several unique features that differentiates it from other resources: (i) it offers an integrated environment that enables the concatenation of evolutionary analyses, the storage of results and that handles format conversions transparently; (ii) once an output file is produced, Phylemon suggests other possible analyses that could logically follow, thus guiding the user through the web server; and finally (iii) users can build and save complete pipelines to be automatically used on many genes in subsequent sessions (phylogenomics).

The main objective of Phylemon is to provide to expert and non-expert users all necessary applications in a single integrated web framework that guides them through the whole sequence evolutionary analysis. Here, we outline the main characteristics of the server and the new developments added to this version. Phylemon 2.0 is accessible at http://phylemon.bioinfo.cipf.es

*To whom correspondence should be addressed. Tel: +(34) 96 328 96 80 ext: 1008; Fax: +(34) 96 328 97 01; Email: hdopazo@cipf.es

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

OUTLINE OF THE PROGRAM

Phylemon 2.0 resources are organized in five major sections: Alignment, Phylogeny, Evolutionary Tests, Pipeliner and Utilities. Phylemon verifies the format of the uploaded input file (non-aligned sequences, aligned sequences, distance matrix, tree or pipeline format) and stores files for the exclusive use of tools reading that format. Users can rename files to help their recognition throughout the project.

A basic phylogenetic analysis consists of: (i) the proposal of a hypothesis about positional homology in a multiple alignment of sequences; and (ii) the search for a tree topology and branch lengths, the main components of a phylogeny. Once the phylogenetic hypothesis is solved, users may test for additional and specific hypotheses related to their sequences, including molecular clock behavior, estimation of synonymous and non-synonymous distances, maximum likelihood (ML) based parameter estimation, statistical support among competing topologies, clades or adaptive events acting on the sequences. For these purposes, Phylemon 2.0 groups different tools under the web tabs: Alignment, Phylogeny and Evolutionary Tests (see sections below).

Table 1 lists all the programs implemented in Phylemon 2.0 and the main connections among them.

Alignment

Phylemon 2.0 integrates four different programs for the alignment of molecular sequences: ClustalW v2.0.10 (3), Muscle v3.7(7), Lagan v2.0 and M-Lagan v2.0 (8). The first two are among the most frequently used programs for MSA. In this version of the server, we added Lagan (Limited Area Global Alignment of Nucleotides) and Multi-Lagan, which run efficient algorithms specifically developed to produce pairwise and multiple alignments of long genomic sequences, respectively.

Lagan and M-Lagan use a single input file containing two or more sequences in Fasta format, respectively. Both programs use the Translated Anchoring option, translating coding regions to anchor sequences. This is

Table 1. Programs available in Phylemon 2.0 web server

Program[a]	Version	Function	Output to program	Pplin[b]
Alignment				
1 T ClustalW	2.0.10	Multiple alignments. DNA and protein sequences	5, 8, 9, 11–16, 21–23, 26, 28	Y
2 T Muscle	3.7	Multiple alignments. DNA and protein sequences	5, 8, 9, 11–16, 21–23, 26, 28	Y
3 T Lagan	2.0	Pairwise alignment. Long and distant genomic sequences	5, 8, 9, 11	–
4 T M-Lagan	2.0	Multiple alignments. Long and distant genomic sequences	5, 8, 9, 11–16, 21–23, 26, 28	–
5 U TrimAl	1.3	Automated trimming of MSAs	8, 9, 11–16, 21–23, 26, 28	Y
6 U CDS-ProtAl	1.0	Alignment of DNA coding sequence using protein template	8, 11–16, 22, 28–30	Y
7 U ConcatenAl	1.0	Concatenation of MSAs	8, 9, 11–16, 21–23, 25, 26	–
8 U ReadAl	1.3	File format conversion	1–5, 8, 9, 11–16, 21–23, 25–30	–
Phylogeny reconstruction				
9 T Seqboot	Phylip 3.68	Bootstrap, jackknife or permutation resampling methods	11–16	Y
10 T Consense	Phylip 3.68	Consensus tree reconstruction	20	Y
11 T Dnadist	Phylip 3.68	DNA pairwise distances computation	17, 18	Y
12 T Protdist	Phylip 3.68	Protein pairwise distances computation	17, 18	Y
13 T DnaML	Phylip 3.68	ML tree reconstruction from DNA data	10, 20	–
14 T ProML	Phylip 3.68	ML tree reconstruction from protein data	10, 20	–
15 T DnaPars	Phylip 3.68	Maximum parsimony tree reconstruction from DNA data	10, 20	–
16 T ProtPars	Phylip 3.68	Maximum parsimony tree reconstruction from protein data	10, 20	–
17 T Neighbor	Phylip 3.68	Tree reconstruction using UPGMA and NJ methods	10, 20	Y
18 T Fitch	Phylip 3.68	Tree reconstruction using LS and ME methods	10, 20	Y
19 U TreeDist	Phylip 3.68	Distance computation among tree topologies	–	–
20 U ETE	2.1 beta	Tree visualization	–	Y
21 T PhyML-Best-AIC-Tree	1.0	ML tree with the best model fitting data under AIC estimation	20	Y
22 T PhyML	3.00	Maximum likelihood analysis (MLA) of DNA & protein data	20	Y
23 T Tree-Puzzle	5.2	MLA of DNA & protein sequences using quartets	20	–
24 T MrBayes	3.1.2	Bayesian phylogenetic analysis of DNA and protein sequences	20	–
Evolutionary tests				
25 T ProtTest	1.4	ML fitting of protein sequences to evolutionary models	–	–
26 T jModelTest	0.1.0	Model testing and phylogeny averaging	–	–
27 T RRTree	1.1.11	Relative rate test	–	–
28 T SLR	1.3	Site-wise analysis of positive and negative selection	–	–
29 T YN00	PAML 4.4c	Pairwise analysis of positive selection (PS) with counting methods	–	–
30 T CodeML	PAML 4.4c	MLA of PS using sites, branch and branch-site models	–	–

Programs are assembled in three main blocks: (i) alignment and files format conversion; (ii) phylogenetic reconstruction; and (iii) evolutionary tests. New resources in this version are shown in cursive.
[a]T-U: tools/utilities.
[b]Pplin: programs able to run in the Pipeliner.

useful when distantly related sequences are compared (i.e. primates and fishes). The Reverse Complement option in Lagan is useful to search for positional homology on the opposite DNA strand of the second sequence. Both programs produce a single output file of aligned sequences in Fasta format. Multiple alignments in Phylemon 2.0 can be sent to distance, parsimony and statistical tree reconstruction (ML and Bayesian) tools. Format conversion or alignment edition can be performed using ReadAl and TrimAl (see 'Utilities' section).

Phylogeny

Phylemon 2.0 incorporates distances, parsimony, ML and Bayesian methods for tree reconstruction. Distance and parsimony methods for DNA or protein sequence data are provided by algorithms of the Phylip package (9) v3.68: DnaDist, ProtDist, DnaPars and ProtPars, respectively. ML analysis can be performed using Phylip (DnaML, ProML), Tree-Puzzle v5.2 (10) and PhyML v3.0 (11,12) programs. Bayesian phylogenetic analysis runs in MrBayes (13) v3.1.2. Users have the option to interact with the program, thus monitoring the progress of the analysis. Program allows the user to specify sump and sumt parameters. Users interested to build the MrBayes commands block can fill the form that summarizes the main parameters. A useful list of command line parameters is available on the fly.

PhyML-Best-AIC-tree v1.02 b is a new tool in Phylemon 2.0. It is a Python script allowing the reconstruction of ML trees using the best AIC-DNA or protein model (14).

Evolutionary tests

For users interested in evolutionary hypotheses testing, Phylemon 2.0 collects tools of: Model Selection, Molecular Adaptation and Relative Rate Test.

In this version, we added jModelTest (15) v0.1, and a new version of ProtTest (16) v1.4 to improve the search for the best model of evolution for DNA and protein explaining the data. One of the interesting results of jModelTest is the average topology obtained by models within 95% confidence interval. This topology can be used as the intree file required for all programs testing for molecular adaptation in Phylemon 2.0.

Adaptation tests on protein-coding DNA sequences run in Phylemon 2.0 by means of Site-wise Likelihood Ratio (SLR) test program vs1.3 (17) and CodeML & YN00 (18) from PAML vs 4.4c (19). Finally, deviations from the molecular clock hypothesis can be tested using the RRTree (20) program vs1.1.11. RRTree computes relative rates tests among user-defined lineages with a weighted or unweighted scheme of species based on the tree topology provided by the user. The program accepts different parameters: the number of synonymous substitutions and synonymous transitions per synonymous site (Ks and As, respectively), the number of non-synonymous substitutions and non-synonymous transversions per non-synonymous site (Ka and Ba, respectively) and, finally, the number of synonymous transversions per 4-fold degenerate site (B4). Kimura two parameters (K2P) (21) and Jukes and Cantor (JC) models are available for non-coding DNA sequences. For protein sequences, RRTree computes a modification of JC model.

Users interested in ML comparison of topologies (paired-sites test) can select evaluation of user-defined trees option in Tree-Puzzle program.

PIPELINE AND PHYLOGENOMICS

Phylogenomic analyses sometimes involve repeating a certain set of analysis over several groups of genes. In such cases, it is necessary to apply the same set of phylogenetic algorithms to different sequence data using a single pipeline of tools. To satisfy this requirement, we developed the Pipeliner tool. Users interested in such kind of studies can upload a zip file containing sequences to run in a pipeline. Previous version of Pipeliner provides basic programs derived from the Phylip package and pipelines like ClustalW, Seqboot, DnaDist/ProtDist, Neighbor and Consense may be used in that order, for a phylogenetic reconstruction with bootstrap values.

Pipeliner in Phylemon 2.0 added PhyML and PhyML-Best-AIC-tree to select the best tree after comparing all AIC (Akaike Informaion Criteria) estimations of DNA or protein models.

Users can add tools from the list of tools and connect them using the 'create link' option. Once all the options of the tools are completed, the user can run and save the pipeline for future jobs.

UTILITIES

Phylemon 2.0 implements three new utilities. First, TrimAl vs1.3 (22) for automatically removing poorly aligned regions from MSAs. The user can select a set of columns to be removed or set specific thresholds based on the fraction of gaps or the similarity of residues in a column. Additionally, TrimAl implements a series of automated algorithms that apply different optimized thresholds, based on the characteristics of each alignment. Second, CDS-ProtAl, a new tool for multiple coding sequence alignment based on protein sequences. This program uses Muscle to compute protein alignments using default parameters but capped at 5 h running time or 9999 iterations on a translation (universal genetic code) of the coding DNA sequences provided as input. Finally, ReadAl v 1.3 a new tool for file conversion among the most popular format files used in phylogeny has been included.

ETE vs2.1 (23) allows users to visualize and interact with trees. The new version allows rooting, collapsing, expanding or swapping nodes and incorporates the possibility to search for distances, support values or names (including the use of Perl-based regular expressions). These options and many others are available by clicking on the nodes, and in the close framework of the tree by using left mouse buttons.

REGISTRATION, ACCOUNTS, PROJECTS AND SPACE

Phylemon 2.0 can be accessed by anonymous login or by registered users. The only difference between these choices is that registered users, from whom only an e-mail is required, can have many different projects and use the server to store up to 1.0 GB of data for future use. Files from anonymous users are deleted after 24 h. Projects and jobs in Phylemon 2.0 can be created, renamed and deleted by users. The number of jobs finished and waiting to be visited, visited, running and those waiting to be run in the queue are colored green, blue, red and yellow, respectively. Users have two icons to access to files, projects and data management.

Technical details

Phylemon 2.0 has been completely reengineered. The server-side is implemented in Java, the client-side is implemented in AJAX (Asynchronous JavaScript And XML). JSON (JavaScript Object Notation) exchanges client and server data. Consequently, the new interface allows asynchronous use of tools (a program can be left running to later come back to see the results), including new facilities for the management of projects and jobs. Moreover, a queue system has been implemented in the server. This release makes an intensive use of new web technologies and standards, so the supported browsers for this version are as follows: Chrome 7+, Firefox 3.5+, Safari 4+, Opera 10+ and Internet Explorer 8. Internet Explorer 6 and 7 are no longer supported. Pipeliner was developed in HTML5 JavaScript and makes use of Scalable Vector Graphics (SVG), therefore it runs in Chrome7+, Firefox4+ or Internet Explorer 9+. More details are available at the Wiki-Help: http://docs.bioinfo.cipf.es/phylemonwiki/doku.php.

DISCUSSION

Molecular evolution, phylogenetics, phylogenomics and evolutionary hypothesis testing embrace a wide range of scientific enquiries in biology. Following the last developments in the field, Phylemon 2.0 combines tools and programs ranging from the simplest distance phylogenetic reconstruction, or the basic relative rate test, to the newest ML model-averaged estimation of the tree topology or the analysis of molecular adaptation. The incorporation of new tools in Phylemon 2.0 extends its usefulness to advanced users trying to find answers to more complex questions of phylogeny and evolution in a web server. Phylemon 2.0 addresses an important requirement of users and students of evolution and phylogeny; namely, the need for a public web server providing a core set of format-compatible, classical and advanced tools truly integrated in an independent web platform.

ACKNOWLEDGEMENTS

We sincerely thank all authors providing the agreement to include their programs in this web server.

FUNDING

Grants (BFU2009-13409-C02-01, BIO2008-04212, BFU2009-09168); 'Plan de Estímulo a la Economía y el Empleo' (Plan E) from MICINN; PROMETEO/2010/001 from the GVA-FEDER. The CIBER de Enfermedades Raras and the INB are initiatives of the ISCIII. Funding for open access charge: Funding of Open Access publication charges was provided by MICINN project to HD.

Conflict of interest statement. None declared.

REFERENCES

1. Felsenstein,J. (2004) *Inferring Phylogenies*, 1st edn. Sinauer Associates Inc., Massachusetts.
2. Huelsenbeck,J.P. and Rannala,B. (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, **276**, 227–232.
3. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
4. Pond,S.L., Frost,S.D. and Muse,S.V. (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, **21**, 676–679.
5. Dereeper,A., Guignon,V., Blanc,G., Audic,S., Buffet,S., Chevenet,F., Dufayard,J.F., Guindon,S., Lefort,V., Lescot,M. *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.*, **36**, W465–W469.
6. Tarraga,J., Medina,I., Arbiza,L., Huerta-Cepas,J., Gabaldon,T., Dopazo,J. and Dopazo,H. (2007) Phylemon: a suite of web tools for molecular evolution, phylogenetics and phylogenomics. *Nucleic Acids Res.*, **35**, W38–W42.
7. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
8. Brudno,M., Do,C.B., Cooper,G.M., Kim,M.F., Davydov,E., Green,E.D., Sidow,A. and Batzoglou,S. (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.*, **13**, 721–731.
9. Felsenstein,J. (2005) *PHYLIP (Phylogeny Inference Package) version 3.6*. Department of Genome Sciences, University of Washington, Seattle.
10. Schmidt,H.A., Strimmer,K., Vingron,M. and von Haeseler,A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**, 502–504.
11. Anisimova,M. and Gascuel,O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.*, **55**, 539–552.
12. Guindon,S. and Gascuel,O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704.
13. Ronquist,F. and Huelsenbeck,J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
14. Akaike,H. (1974) A new look at the statistical model identification. *IEEE Trans. Automatic Control*, **9**, 716–723.
15. Posada,D. (2008) jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.*, **25**, 1253–1256.
16. Abascal,F., Zardoya,R. and Posada,D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, **21**, 2104–2105.
17. Massingham,T. and Goldman,N. (2005) Detecting amino acid sites under positive selection and purifying selection. *Genetics*, **169**, 1753–1762.
18. Yang,Z. and Nielsen,R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.*, **17**, 32–43.

19. Yang,Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.
20. Robinson-Rechavi,M. and Huchon,D. (2000) RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics*, **16**, 296–297.
21. Kimura,M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**, 111–120.
22. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
23. Huerta-Cepas,J., Dopazo,J. and Gabaldon,T. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics*, **11**, 24.