Published online 29 April 2011 *Nucleic Acids Research, 2011, Vol. 39, Web Server issue* ***W479–W485***
doi:10.1093/nar/gkr243

PhyleasProg: a user-oriented web server for wide evolutionary analyses

Joël Busset[1,2,3,4], Cédric Cabau[5], Camille Meslin[1,2,3,4] and Géraldine Pascal[1,2,3,4,*]

[1]INRA, UMR85, Physiologie de la Reproduction et des Comportements, [2]CNRS, UMR6175, F-37380 Nouzilly, [3]Université François Rabelais de Tours, F-37041 Tours, [4]IFCE, F-37380 Nouzilly and [5]INRA, SIGENAE, UR83 Recherches Avicoles, F-37380 Nouzilly, France

Received February 15, 2011; Revised March 22, 2011; Accepted April 5, 2011

ABSTRACT

Evolutionary analyses of biological data are becoming a prerequisite in many fields of biology. At a time of high-throughput data analysis, phylogenetics is often a necessary complementary tool for biologists to understand, compare and identify the functions of sequences. But available bioinformatics tools are frequently not easy for non-specialists to use. We developed PhyleasProg (http://phyleasprog.inra.fr), a user-friendly web server as a turnkey tool dedicated to evolutionary analyses. PhyleasProg can help biologists with little experience in evolutionary methodologies by analysing their data in a simple and robust way, using methods corresponding to robust standards. Via a very intuitive web interface, users only need to enter a list of Ensembl protein IDs and a list of species as inputs. After dynamic computations, users have access to phylogenetic trees, positive/purifying selection data (on site and branch-site models), with a display of these results on the protein sequence and on a 3D structure model, and the synteny environment of related genes. This connection between different domains of phylogenetics opens the way to new biological analyses for the discovery of the function and structure of proteins.

INTRODUCTION

Today, more and more eukaryotic genomes have been sequenced thanks to second-generation sequencing technologies thereby providing an extraordinary wealth of information for evolutionary analyses. Currently, the GOLD web site (1) lists more than 3000 eukaryotic genomes whose sequencing is complete or ongoing. Under these circumstances, bioinformatics tools could help to understand the evolutionary histories of proteins especially by connecting phylogenetics analysis and positive selection calculations. These approaches constitute the core of many biological research areas, and as stated by Theodosius Dobzhansky 'Nothing in biology makes sense except in the light of evolution'. Indeed, present protein sequences are the result of a long, complex and extensive evolutionary process. Proteins have different levels of conservation. Active sites or protein–protein interaction domains are often well conserved, while highly variable regions may carry sites under positive selection. Such positively selected sites may be interpreted as being a consequence of molecular adaptation, which may confer an evolutionary advantage to the organism (2–4).

Accordingly, the association of (i) the establishment of orthology and paralogy relationship; (ii) the functional inference by reconstruction of the phylogenetic tree; and (iii) the identification of sites/genes under positive selection is an important step, not only in studies of evolutionary biology, but also in functional studies. By projecting the results of positive selection onto the 3D structure of proteins, this becomes a powerful and very useful tool for biologists. The combined data could help biologists plan site-directed mutagenesis experiments. However, obtaining a phylogenetic tree requires successive computations including identification of homologous sequences, multiple alignment, phylogenetic reconstructions and graphic representation of the inferred tree. Obtaining positive selection data require the use of mathematical methods, such as PAML (5), which are designed for specialists.

Several web sites offer phylogenetic tree reconstruction. Some are turnkey systems such as PhyloBuilder (6) and POWER (7). Some offer a single tool, while others bring together many of the most popular programs for phylogenetic reconstruction such as Mobyle (8). The web server Phylogeny.fr (9) is designed for non-specialists and has up-to-date programs that are often designed for experts. In parallel, two phylogenetic tree databases, PhylomeDB (10) and TreeFam (11), offer a large number of pre-computed

*To whom correspondence should be addressed. Tel: 33 247427795; Fax: 33 247427743; Email: geraldine.pascal@tours.inra.fr

trees based on all genes of all genomes. A number of web sites are also available for analysing evolutionary forces. The web server Selecton (12) offers a user-friendly tool to compute positive selection and displays results on a 3D structure of proteins. However, it only allows calculation of one set of orthologues. The DataMonkey server (13) enables detection of signatures of positive and negative selection from coding sequence alignments using a wide range of statistical models. The Selectome (14) database provides the results of a branch-site-specific likelihood test for positive selection based on whole gene families from the TreeFam database. Phylemom (15) enables experts to build a complete pipeline dedicated to phylogenetics and evolution.

Many tools are already available to reply to phylogenetics and evolutionary questions. However, they are complex to use and do not allow all the necessary computations to be carried out on a single server. Phylogenetic tree reconstruction, positive selection detection and protein 3D structure modelling require (i) installation/use of numerous tools; (ii) knowledge of up-to-date tools; and (iii) substantial computational resources. In particular, when biologists analyse several proteins of interest, they want to repeat bioinformatics methods on their data in the same conditions and they want to obtain results in a reasonable amount of time. This is why we built PhyleasProg web server in such a way that it could be used by the largest possible number of biologists. Our aim was to combine usefulness and usability. Such a server is a helpful guide for biologists with little experience in evolutionary methodologies as it can analyse their data in a simple and robust way, using methods corresponding to well-accepted standards.

Via a very simple interface, users enter one or a list of Ensembl protein IDs (16) and choose a set of species about which they wish to obtain evolutionary information among the sequenced vertebrates in Ensembl. Once submitted, each ID is treated independently and the computations are performed on both orthologues and paralogues of the related genes. As output, PhyleasProg provides (i) phylogenetic trees; (ii) positive/purifying selection data (on site and branch-site models) with visualization of these outcomes on the protein sequence and whenever possible, on a 3D structure; and (iii) the genomic environment of related genes. To our knowledge, no other web server performs all these tasks on several input sequences simultaneously. In addition, PhyleasProg computes the degree of purifying selection and positive Darwinian selection for each site in the protein sequence and displays these data on the modelled molecular structure of the protein. To guide users through these different evolutionary methods, which are not always very easy for non-experts, the pipeline only returns results if they are statistically significant.

This unique connection between phylogenetic trees, synteny studies, positive/purifying selection data and 3D structures opens the way to new biological analyses to improve our understanding of function and structure of proteins.

OVERVIEW

The PhyleasProg pipeline is a combination of Perl modules and external software (Figure 1). As input data, it requires one or a list of Ensembl protein IDs and a list of species selected among completely or partially sequenced vertebrates in Ensembl (16). Once the process is complete, users can obtain evolutionary results on each ID submitted, treated independently but simultaneously, on orthologues and paralogues of the related genes.

We intentionally chose to not embed an exhaustive number of similar methodologies in our platform. We chose rapid, up-to-date, accurate and proven tools. Multiple sequence alignments are performed by MUSCLE (17) and are refined by GBLOCKS (18), itself improved by a home-made Perl program. TREEBEST (http://treesoft.sourceforge.net/treebest.shtml) reconstructs phylogenetic trees. CODEML, a PAML program (5), performs positive selection computation. MODELLER (19) builds homology models of the 3D structure of proteins.

Data visualization was an important goal for the development of this platform. JALVIEW (20) is used to display multiple sequence alignments, ARCHAEOPTERYX (21) for interactive manipulation of phylogenetic trees and JMOL (22) to display the 3D structure of proteins. We were careful to present processes and results very simply to enable biologists to navigate through a user-friendly environment. To guide users, the pipeline only returns significant results. Moreover, all input and output data can be downloaded as flat files.

A cluster computer manages the execution of the whole pipeline. This choice allows a very reasonable execution speed and authorizes PhyleasProg to work on several proteins simultaneously. The user interface was optimized for Firefox browser developed in Perl CGI.

PHYLEASPROG PIPELINE

Data acquisition

Input. For a very simple use of PhyleasProg, only Ensembl IDs of the proteins to be studied and a list of the species with which they should be compared are required as inputs. Protein IDs can be separated by a comma, a space or a new line character. Ensembl protein IDs are unique, they start with 'ENS' and their last letter must be a 'P' (e.g. ENSMUSP00000099398). To choose species for which they want evolutionary results, users simply tick the name of the species in the lists of completely and partially sequenced genomes. The *Job summary page* summarizes the list of IDs submitted, the selected species and displays the status of process for each ID.

Interrogation of Ensembl database. We chose to work with Ensembl protein IDs because Ensembl provides high-quality genome annotation across vertebrate species and allows computer scientists to retrieve a lot of data very quickly, thanks to a Perl application programming interface (API) (23).

Using this API, for each protein ID submitted, we retrieved protein and related transcript sequences,

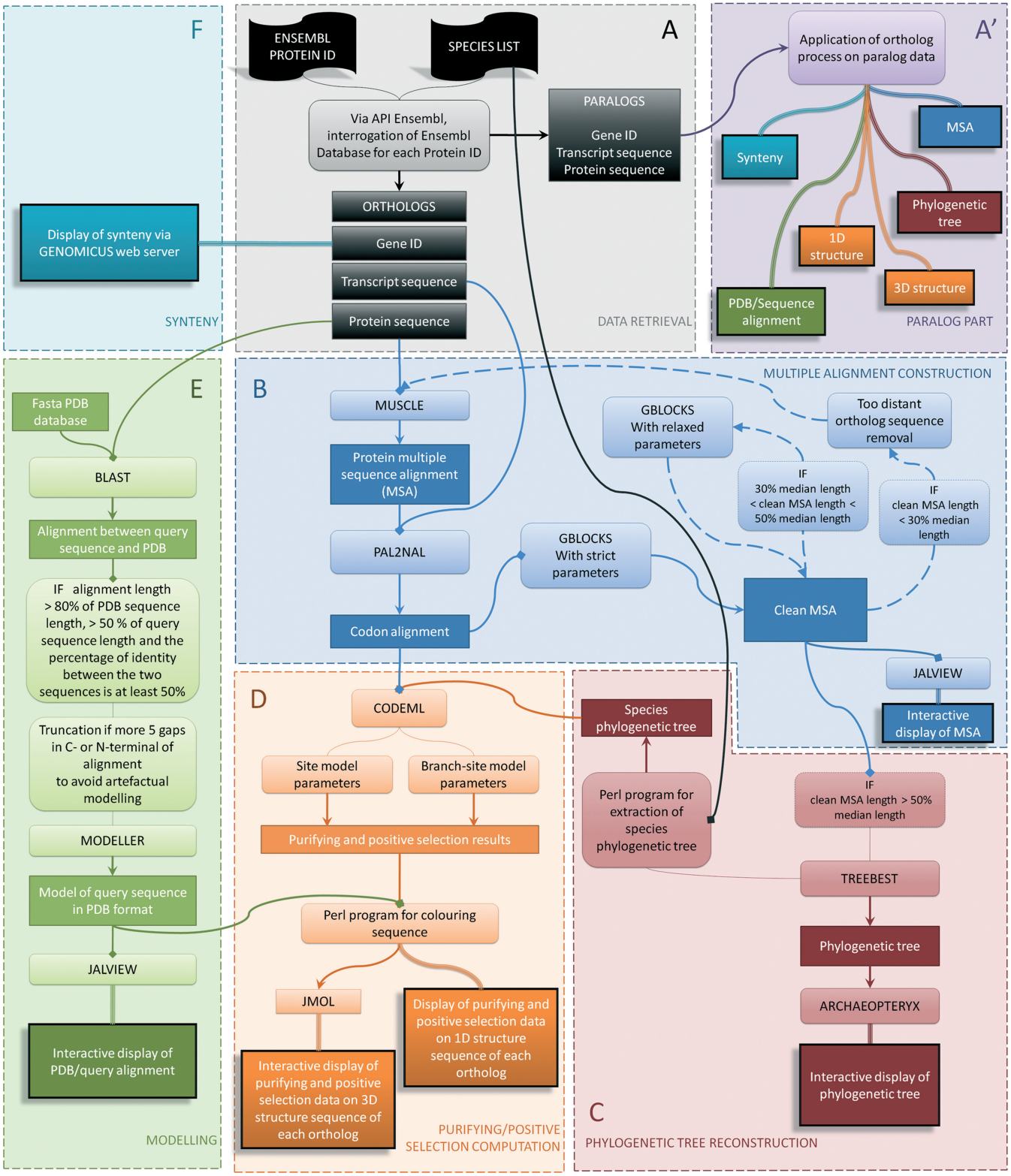

Figure 1. The workflow of PhyleasProg web server.

related gene ID, orthologous and paralogous protein IDs, orthologue and paralogue protein sequences and related transcript sequences (Figure 1A and A′). Among the numerous orthologues identified in Ensembl, we chose to keep either the one-to-one orthologues or the related gene with the shortest evolutionary distance among the one-to-many or the many-to-many orthologues (24).

Reconstruction of phylogenetic trees

Multiple sequence alignment and refinement. For each protein ID submitted, PhyleasProg reconstructs phylogenetic trees of both orthologues and paralogues. And for each orthologue related to one of the protein IDs submitted, a phylogenetic tree of paralogues is also reconstructed.

As shown in Figure 1B, multiple sequence alignment (MSA) of proteins is generated by MUSCLE. This alignment is then converted into multiple codon alignment by PAL2NAL (25). As our pipeline offers a turnkey process, we had to pay particular attention to the quality of MSA because this is essential for the quality of the related phylogenetic tree. Thus, GBLOCKS is used to edit MSA. This software removes all sites containing at least one gap and sites that are too divergent because these positions might not be homologous or might be saturated by multiple substitutions. First of all, GBLOCKS is performed with strict parameters (type = codons; maximum number of contiguous non-conserved positions = 8; minimum length of a block = 10; no gaps allowed). After this first step, the generated MSA can be very short, which would seriously damage the rest of the computations in the PhyleasProg pipeline. Consequently, refinement step are performed recursively: if after GBLOCKS, the MSA length is <30% of the median length of sequences in the raw MSA, the sequence that induces most of the gaps is removed from the dataset, and a new MSA is computed. If the length of the clean MSA is between 30 and 50%, a new editing with GBLOCKS is performed on the raw MSA with relaxed parameters (type = codons; maximum number of contiguous non-conserved positions = 10; minimum length of a block = 5; no gaps allowed). If after this last step, the length of the MSA is still too short, computation is aborted. Thus, it is important to estimate the quality of the MSA (downloadable through the flat files menu) before analysing the other results of the pipeline (Figure 2).

Phylogenetic reconstruction. The clean MSA from the previous step is used to reconstruct the phylogenetic tree by TreeBeST (Figure 1C). TreeBeST integrates multiple tree topologies, in particular both DNA- and protein-level models and combines them with a species-tree aware penalization of topologies, which is inconsistent with known species relationships. TreeBeST is run with the option *best*. This enables the combination of (i) a maximum likelihood (ML) tree built using PhyML (26) based on the protein alignment with the Whelan And Goldman model; (ii) a ML tree built using PhyML based on the codon alignment with the Hasegawa–Kishino–Yano (HKY) model; (iii) a neighbour-joining (NJ) tree using *p*-distance based on the codon alignment; (iv) a NJ tree using d*N* distance (rate of non-synonymous substitutions) based on the codon alignment; and (v) a NJ tree using d*S* distance (rate of synonymous substitutions) based on the codon alignment. As TreeBeST runs with a species tree, the final phylogenetic tree is rooted by minimizing gene duplications and then losses, the best rooting strategy for this type of input.

Visualization. Archaeopteryx, the successor of ATV (27), is a Java application used as applet for the display and manipulation of annotated phylogenetic trees.

Positive/purifying selection calculations

Overview. PhyleasProg gives positive and purifying selection data using ML calculations which underlie the stochastic process of evolution. CODEML, from the package PAML (Figure 1D) (5), evaluates the ratio of non-synonymous/synonymous substitution rates (d*N*/d*S*), denoted ω, which is a measure of selective pressure. Values of $\omega < 1$, $= 1$ and > 1 are indicators of purifying selection, neutral evolution and positive selection, respectively. Two distinct categories of codon substitution models are used: site models (M1a versus M2a, M7 versus M8 and M8a versus M8) and branch-site models. For the two types of analyses, two models are compared: one model which allows positive selection and one model which does not allow positive selection. For each model, the lnL (log likelihood) value is retrieved ($\ln L_1$ for the model allowing positive selection, $\ln L_0$ for the other) and a LRT (likelihood ratio test) is calculated [$LRT = 2 \times (\ln L_1 - \ln L_0)$] to assess the significance of the results. The LRT value follows a χ^2 law which allows the *P*-value of the LRT to be obtained. If the LRT is significant for the comparison, PhyleasProg lists sites under positive selection detected by Bayes empirical Bayes (BEB) with posterior probabilities >95% and sites under purifying selection.

As shown in Figure 2, selection pressure data appear in two separate menus. One is dedicated to results of site models and the other one to results of branch-site models. In the second case, these models allow the ω ratio to vary both among sites in the protein and across branches on the tree and aim to detect positive selection affecting a few sites along particular lineages (foreground branches). In the pipeline, all branches of the tree are tested as foreground branches for positive selection. Two models are used, one called alternative and one called null. In the alternative model, three classes of sites are admitted for the foreground branch, ω_0: $dN/dS < 1$, ω_1: $dN/dS = 1$ and ω_2: $dN/dS \geq 1$. In the null model, ω_2 is fixed to 1. Significant results with branch-site models are accessible on a clickable tree. Branches under positive selection are represented by a purple star and are highlighted in green. Raw result files (rst) of CODEML are also available.

Visualization. Results of selection pressure calculation with site and branch-site models share the same presentation (Figure 2). Data are visualized on 1D and 3D structures on the same results page. A dropdown menu embedded in the positive selection results web page enables users to visualize data on each protein in the orthologue or paralogue dataset. For the two types of representations, a discrete colour scale is used to distinguish the different values of ω for each site. The scale from green to yellow represents purifying selection, i.e. $\omega < 0.3$, while red and orange represent positive selection with posterior probabilities >99% or 95%, respectively. White means that no information is available for this site

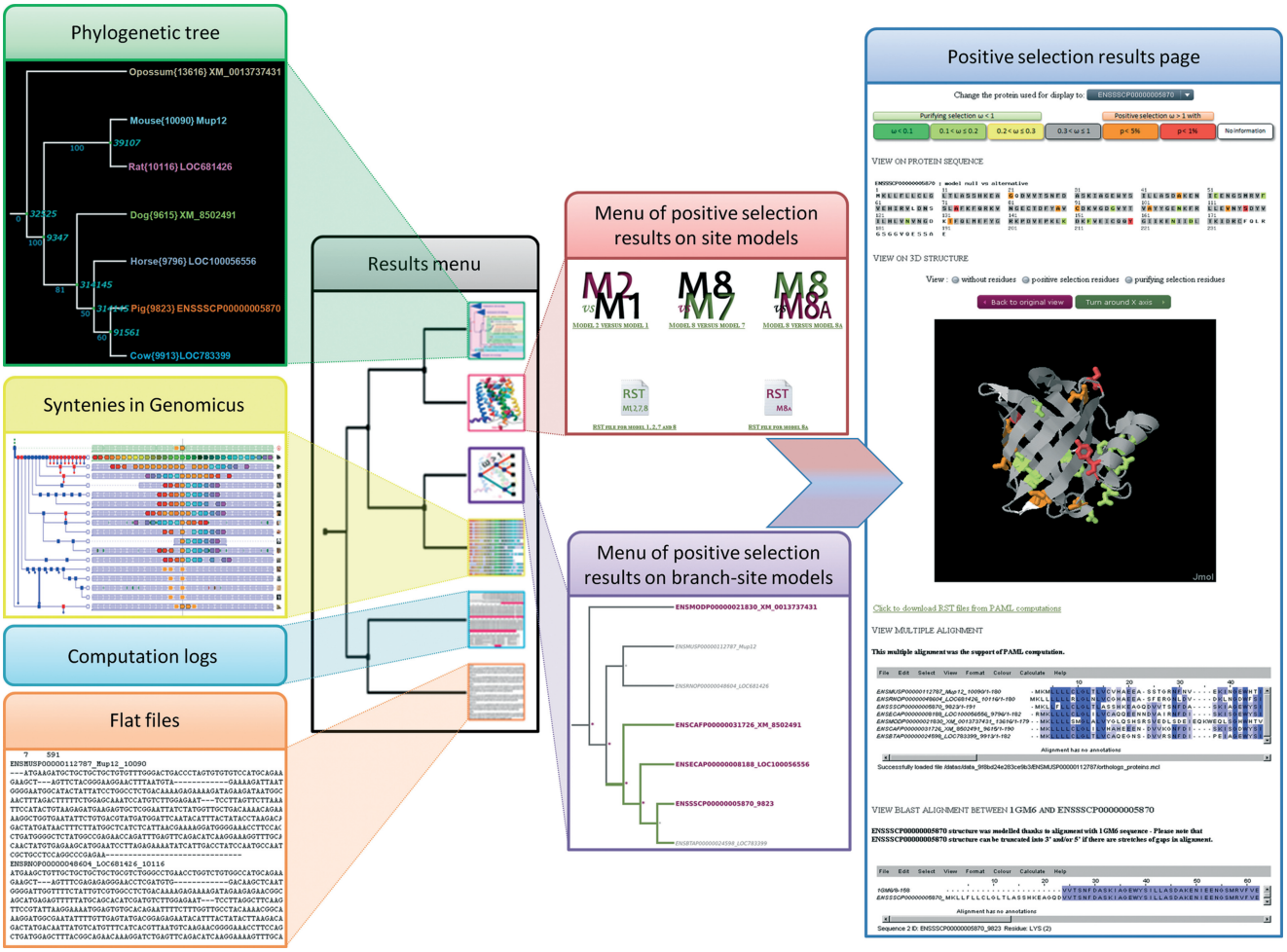

Figure 2. Overview of the results menu of PhyleasProg and its results pages.

because no calculation was performed by CODEML due to at least one gap in the MSA at this position. Grey means results are not significant enough to infer either purifying or positive selection. To locate different amino acids in different organisms, the MSA used for PAML computation is displayed using the JalView applet.

These data can greatly help biologists to plan site-directed mutagenesis experiments to target essential functional residues. This was the main reason to have PhyleasProg display results on a 3D structure, if one can be modelled (Figure 1E). To model the 3D structure, a BLAST (28) search is performed to find a similar structure in the PDB database (29) in order to use it as a template to calculate a model with Modeller. 3D structure is sometimes difficult to predict, mostly when the template is too distant from the sequence to be modelled. To avoid models of insufficient quality, a model is built only if: (i) the alignment between the sequence to be modelled and the length of the PDB template covers at least 80% of PDB sequence and at least 50% of the query sequence and (ii) the percentage of identity between the two sequences is at least 50%. If the query sequence is shorter than the template, amino acids in the C- or N-terminal are removed. In order to enable users to locate differences between a raw query sequence and the model, the alignment between the PDB sequence and raw query sequence is displayed using JalView. Hence, when a homology model can be built, evolutionary results are directly visualized on the modelled structure, while if homology modelling is not possible, results are only presented on the 1D sequence.

Synteny exploration

In order to achieve complete evolutionary analysis of the protein submitted, PhyleasProg offers the possibility to explore the genetic environment of related genes. Indeed, in the results menu (Figure 2) the user has a link to Genomicus (30). This database is a synteny browser that can represent and compare numerous genomes in a broad phylogenetic view. In addition, Genomicus includes the reconstructed organization of ancestral gene, thus greatly facilitating interpretation of the data. We chose not to develop our own genome browser because this web tool is really accurate, complete, up-to-date, user-oriented and also based on Ensembl data.

CONCLUSION AND FUTURE DEVELOPMENTS

With PhyleasProg, we offer biologists a tool specially developed for non-specialists of phylogenetics, which is user-oriented, fast, complete, up-to-date, ready-to-use and accessible via a web interface, and allows the user to submit several jobs at the same time. All computations are dynamically produced and displayed as soon as the results are available, so the user can begin to analyse results without waiting for the whole process to end.

Thanks to the modular architecture of our pipeline, it is relatively easy to update and to incorporate new tools. In the short term, our main plan is to extend the range of possible inputs. With the present system, only proteins from organisms available in Ensembl can be treated in PhyleasProg. A FASTA sequence as input, for example, could be useful. We also want to let users upload their own PDB files. In the very near future, we will offer a 3D structure model based on a multiple alignment including several proteins from the PDB database, which would improve the quality of the models. Finally, to provide more accurate pressure selection data, we are already thinking about a way to minimize the guanine-cytosine bias in positive selection results.

ACKNOWLEDGEMENTS

The project was hosted by the Toulouse Midi-Pyrénées bioinformatics platform. Sincere thanks to Didier Laborie and Sylvain Thomas for their technical support. We thank Anne Poupon for sharing expertise on Modeller, for critical reading and suggestions on the manuscript. Particular thanks goes to Delphine Capela for critical reading and suggestions on the manuscript. We would like to thank Philippe Monget for his supportive discussions on project. We are grateful to Alexis Dereeper, Ziheng Yang and Li Heng for helpful discussions on clickable tree, Codeml and Treebest, respectively. We are grateful to Daphne Goodfellow for attention to the English-language version.

FUNDING

MENRT PhD fellowship (to C.M.). Funding for open access charge: INRA.

Conflict of interest statement. None declared.

REFERENCES

1. Liolios,K., Chen,I.M., Mavromatis,K., Tavernarakis,N., Hugenholtz,P., Markowitz,V.M. and Kyrpides,N.C. (2010) The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.*, **38**, D346–D354.
2. Eyre-Walker,A. (2006) The genomic rate of adaptive evolution. *Trends Ecol. Evol.*, **21**, 569–575.
3. Graur,D. and Li,W.H. (2000) *Fundamentals of Molecular Evolution*, 2nd edn. Sinauer Associates, Sunderland, MA, USA.
4. Studer,R.A., Penel,S., Duret,L. and Robinson-Rechavi,M. (2008) Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. *Genome Res.*, **18**, 1393–1402.
5. Yang,Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.
6. Glanville,J.G., Kirshner,D., Krishnamurthy,N. and Sjolander,K. (2007) Berkeley Phylogenomics Group web servers: resources for structural phylogenomic analysis. *Nucleic Acids Res.*, **35**, W27–W32.
7. Lin,C.Y., Lin,F.K., Lin,C.H., Lai,L.W., Hsu,H.J., Chen,S.H. and Hsiung,C.A. (2005) POWER: PhylOgenetic WEb Repeater–an integrated and user-optimized framework for biomolecular phylogenetic analysis. *Nucleic Acids Res.*, **33**, W553–W556.
8. Neron,B., Menager,H., Maufrais,C., Joly,N., Maupetit,J., Letort,S., Carrere,S., Tuffery,P. and Letondal,C. (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics*, **25**, 3005–3011.
9. 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.
10. Huerta-Cepas,J., Capella-Gutierrez,S., Pryszcz,L.P., Denisov,I., Kormes,D., Marcet-Houben,M. and Gabaldon,T. (2010) PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res.*, **39**, D556–D560.
11. Ruan,J., Li,H., Chen,Z., Coghlan,A., Coin,L.J., Guo,Y., Heriche,J.K., Hu,Y., Kristiansen,K., Li,R. *et al.* (2008) TreeFam: 2008 Update. *Nucleic Acids Res.*, **36**, D735–D740.
12. Stern,A., Doron-Faigenboim,A., Erez,E., Martz,E., Bacharach,E. and Pupko,T. (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.*, **35**, W506–W511.
13. Delport,W., Poon,A.F., Frost,S.D. and Kosakovsky Pond,S.L. (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, **26**, 2455–2457.
14. Proux,E., Studer,R.A., Moretti,S. and Robinson-Rechavi,M. (2009) Selectome: a database of positive selection. *Nucleic Acids Res.*, **37**, D404–D407.
15. 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.
16. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2010) Ensembl 2011. *Nucleic Acids Res.*, **39**, D800–D806.
17. Edgar,R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
18. Talavera,G. and Castresana,J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *System. Biol.*, **56**, 564–577.
19. Eswar,N., Eramian,D., Webb,B., Shen,M.Y. and Sali,A. (2008) Protein structure modeling with MODELLER. *Methods Mol. Biol.*, **426**, 145–159.
20. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
21. Han,M.V. and Zmasek,C.M. (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, **10**, 356.
22. Herráez,A. (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ.*, **34**, 255–261.
23. Stabenau,A., McVicker,G., Melsopp,C., Proctor,G., Clamp,M. and Birney,E. (2004) The Ensembl core software libraries. *Genome Res.*, **14**, 929–933.
24. 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 Res.*, **19**, 327–335.
25. Suyama,M., Torrents,D. and Bork,P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.*, **34**, W609–W612.

26. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *System. Biol.*, **59**, 307–321.
27. Zmasek,C.M. and Eddy,S.R. (2001) ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics*, **17**, 383–384.
28. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
29. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D. *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
30. Muffato,M., Louis,A., Poisnel,C.E. and Roest Crollius,H. (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics*, **26**, 1119–1121.