W580–W584 *Nucleic Acids Research, 2015, Vol. 43, Web Server issue* *Published online 06 April 2015*
doi: 10.1093/nar/gkv279

The EMBL-EBI bioinformatics web and programmatic tools framework

Weizhong Li†, Andrew Cowley†, Mahmut Uludag, Tamer Gur, Hamish McWilliam, Silvano Squizzato, Young Mi Park, Nicola Buso and Rodrigo Lopez*

European Bioinformatics Institute, EMBL Outstation, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, Cambridge, UK

Received January 30, 2015; Revised March 18, 2015; Accepted March 21, 2015

ABSTRACT

Since 2009 the EMBL-EBI Job Dispatcher framework has provided free access to a range of mainstream sequence analysis applications. These include sequence similarity search services (https://www.ebi.ac.uk/Tools/sss/) such as BLAST, FASTA and PSI-Search, multiple sequence alignment tools (https://www.ebi.ac.uk/Tools/msa/) such as Clustal Omega, MAFFT and T-Coffee, and other sequence analysis tools (https://www.ebi.ac.uk/Tools/pfa/) such as InterProScan. Through these services users can search mainstream sequence databases such as ENA, UniProt and Ensembl Genomes, utilising a uniform web interface or systematically through Web Services interfaces (https://www.ebi.ac.uk/Tools/webservices/) using common programming languages, and obtain enriched results with novel visualisations. Integration with EBI Search (https://www.ebi.ac.uk/ebisearch/) and the dbfetch retrieval service (https://www.ebi.ac.uk/Tools/dbfetch/) further expands the usefulness of the framework. New tools and updates such as NCBI BLAST+, InterProScan 5 and PfamScan, new categories such as RNA analysis tools (https://www.ebi.ac.uk/Tools/rna/), new databases such as ENA non-coding, WormBase ParaSite, Pfam and Rfam, and new workflow methods, together with the retirement of depreciated services, ensure that the framework remains relevant to today's biological community.

INTRODUCTION

The European Bioinformatics Institute (EMBL-EBI https://www.ebi.ac.uk) has provided free and open access to a range of bioinformatics applications for sequence analysis since 1998 (1). In 2009 the Job Dispatcher framework (2,3) was released to provide consistent, robust and updatable access to modern bioinformatics tools such as NCBI BLAST+ (4) and PSI-Search (5) for sequence similarity searching; InterProScan (6) and PfamScan (7) for protein functional analysis; and multiple sequence alignment tools such as Clustal Omega (8), Kalign2 (9) and MAFFT (10). Through these applications the latest mainstream bioinformatics databases can be searched, for example ENA (11), Ensembl Genomes (12), UniProt (13), InterPro (14) and Pfam (15).

The framework is used by academic and industry scientists, and in 2014 handled roughly 110 million analysis jobs, up from 65 million in 2013. Help pages, tutorials and user guides (available as protocols (16)) are provided, together with training courses and helpdesk support. Continued feedback from the biological community, collaboration with bioinformatics tools and data providers and comprehensive metrics analysis helps to drive improvements to the accessibility and quality of the services.

THE TOOLS FRAMEWORK

The EMBL-EBI Job Dispatcher is a modular and configuration-driven framework aimed at both novice and expert users. A uniform web browser interface enables users to upload their data or select existing data from our databases for analysis in a wide range of applications (Table 1). Browser inputs are checked to validate all the parameters required for successful job submission and guidance is provided to the user in the case of failure. Default parameter choices are set in collaboration with the tool authors for the intended uses of the tools, and can be adjusted by the user. Results are presented visually and enriched with data from other applications, for example cross-reference annotations via EBI Search (17) or functional domain predictions via InterPro (14). Biological data entries discovered as part of the analysis can be retrieved via the dbfetch service (3). SOAP and REST Web Services provide stable APIs for programmatic use. Input validation and parameter help is also built-in to Web Service use and results can be retrieved in a range of graphical and machine

*To whom correspondence should be addressed. Tel: +44 1223 494 423; Fax: +44 1223 494 468; Email: rls@ebi.ac.uk
†These authors contributed equally to the paper as first authors.

Table 1. Tool services available in the Job Dispatcher framework

Category	Tool
EMBOSS Programs (https://www.ebi.ac.uk/Tools/emboss/)	needle, stretcher, water, matcher, transeq, sixpack, backtranseq, backtranambig, pepinfo, pepstats, pepwindow, cpgplot, newcpgreport, isochore & seqret
Multiple Sequence Alignment (https://www.ebi.ac.uk/Tools/msa/)	clustal omega, clustalw2, dbclustal, kalign, mafft, mafft_addseq, muscle, mview, tcoffee & prank
Pairwise Sequence Alignment (https://www.ebi.ac.uk/Tools/psa/)	needle, stretcher, water, matcher, lalign, wise2dba, genewise & promoterwise
Phylogeny Analysis (https://www.ebi.ac.uk/Tools/phylogeny/)	clustalw2 phylogeny & raxml_epa
Protein Functional Analysis (https://www.ebi.ac.uk/Tools/pfa/)	censor, fingerprintscan, interproscan 5, pfamscan, phobius, pratt, prosite scan & radar
RNA Analysis (https://www.ebi.ac.uk/Tools/rna/)	infernal_cmscan & mapmi
Sequence Format Conversion (https://www.ebi.ac.uk/Tools/sfc/)	seqret, readseq & mview
Sequence Operation (https://www.ebi.ac.uk/Tools/so/)	censor & seqcksum
Sequence Similarity Search (https://www.ebi.ac.uk/Tools/sss/)	ncbiblast+, fasta, ggsearch, glsearch, psiblast, psisearch, ssearch & wublast
Sequence Statistics (https://www.ebi.ac.uk/Tools/seqstats/)	pepinfo, pepstats, pepwindow, saps, cpgplot, newcpgplot & isochore
Sequence Translation (https://www.ebi.ac.uk/Tools/st/)	transeq, sixpack, backtranseq & backtranambig

readable formats. Sample Web Services clients are available in a range of programming languages (e.g. C#, Java, Perl, Python and Ruby).

New analysis tools and databases

New tool developments include NCBI BLAST+ for sequence similarity searching, InterProScan 5 (6) and PfamScan (7) for protein functional analysis, Infernal_cmscan (18) and MapMi (19) for RNA analysis, RAxML_EPA (20) for phylogenetic analysis and MAFFT_addseq (10) for multiple sequence alignment. Please see the Supplementary Information for sample inputs for PfamScan, Infernal_cmscan and MapMi. New sequence databases include ENA Coding and Non-coding sequence databases, WormBase ParaSite (21), Pfam (15), Rfam (22) along with many new genomes and proteomes as existing databases are updated.

Tool and database retirements

Legacy applications SRS (23), InterProScan 4 (24) and NCBI BLAST (non-plus) have been retired. EMBLCDS (11), HGVbase (25), IPI (26) and proteomes databases (13) have been removed from sequence similarity searching services.

New functionalities

As a result of user-feedback we have incorporated additional workflow functionalities to the framework. Interactive workflows help the investigator move between different tool categories and can be utilised through both the web browser interface and Web Services. Figure 1 illustrates an example workflow constructing a phylogenetic tree from sequence similarity search results. The top BLAST hit sequences (Figure 1a) are selected and aligned using the Clustal Omega tool (8); the alignment (Figure 1b) is then used to generate a phylogenetic analysis (Neighbor-Joining clustering (27)) and the final phylogenetic tree is displayed (Figure 1c). The user can control which sequences are selected at each stage in the process. Sequences are retrieved behind the scenes and robust filtering and validation procedures and additional pre- and post-processing steps have been implemented to ensure successful job submissions across the workflow.

New result representations

The web interfaces have adopted the latest EMBL-EBI web style guidelines (https://www.ebi.ac.uk/web/guidelines) and are more user-friendly as a result of extensive usability testing. Feature annotations can be displayed that (Figure 2) highlight UniProt sequence features present within well-aligned regions and are available in the FASTA (28), PSI-Search and LALIGN services. The result summary tables (Figure 1a) for sequence similarity searching can now be downloaded in XML, CSV, TSV and JSON formats. The NCBI BLAST+ service now offers more BLAST alignment views, including ASN archive format. Phylogenetic analysis offers output in percentage identity matrix (PIM) and a new tree viewing (Figure 1c) component is now available that uses JavaScript technologies such as BioJS (29) and D3 (d3js.org).

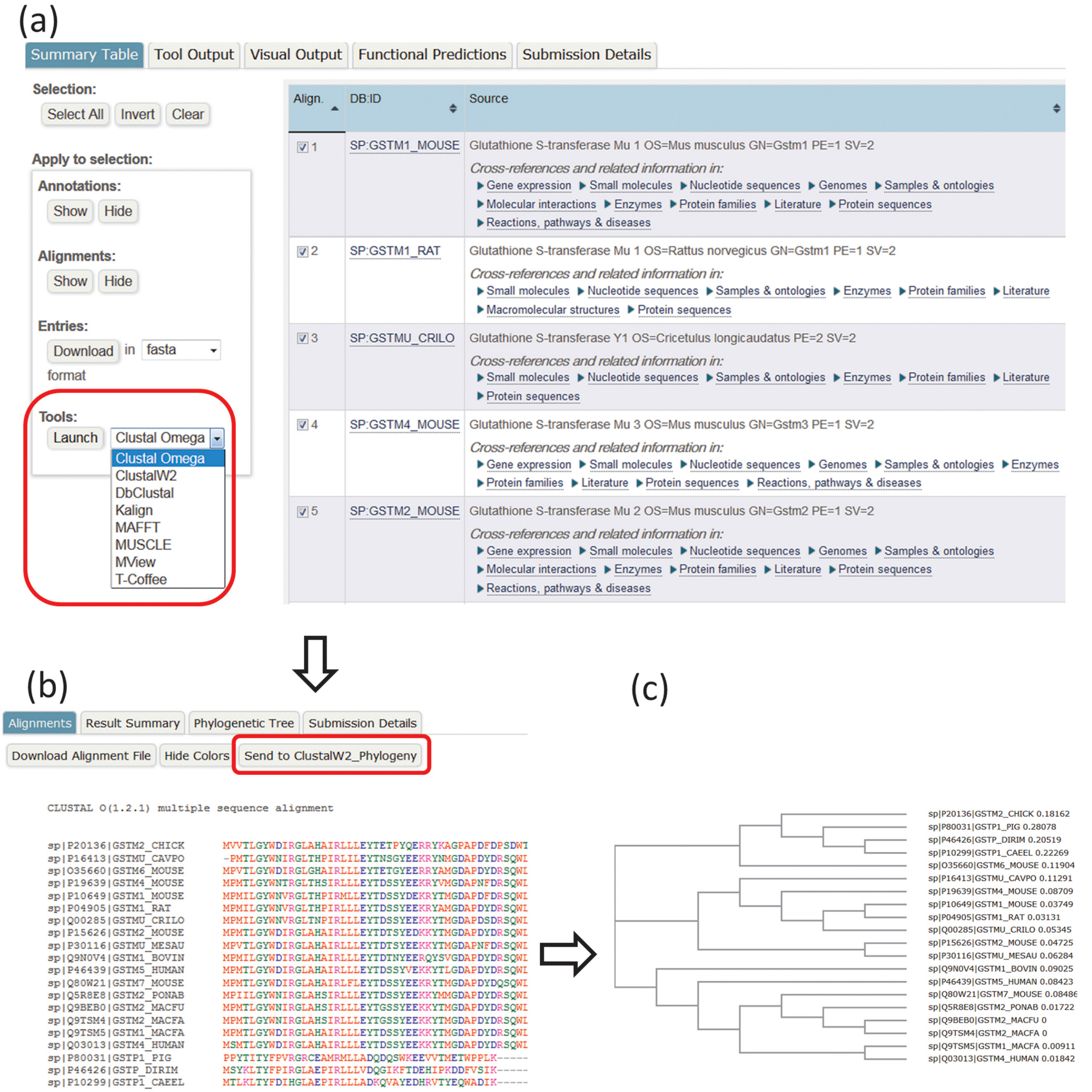
WSDLs for the SOAP Web Services API have been provided since the first availability of the framework in 2009. Users of the REST API are now supported through the provision of equivalent WADLs for all tools. The parameter settings of analysis jobs can be accessed through the REST API as well. Integrated tests have been implemented to make the Web Services more robust and stable.

Help and documentation

EMBL-EBI offers helpdesk support and training courses for the use of the tool services provided by the framework. General help, FAQ pages, tutorials and example protocols (16) are available for using the services via web browser interfaces and sample clients for Web Services. A brief guide to Web Services technologies is also provided for those wishing to learn more and develop their own client programs (https://www.ebi.ac.uk/Tools/webservices/tutorials/00_contents).

FUTURE DEVELOPMENTS

As well as continuing to maintain existing services, future planned developments include the integration of new tools


(a)
Summary Table
Tool Output
Visual Output
Functional Predictions
Submission Details
Selection:
Select All
Invert
Clear
Apply to selection:
Annotations:
Show
Hide
Alignments:
Show
Hide
Entries:
Download
in
fasta
format
Tools:
Launch
Clustal Omega
Clustal Omega
ClustalW2
DbClustal
Kalign
MAFFT
MUSCLE
MView
T-Coffee
(b)
Alignments
Result Summary
Phylogenetic Tree
Submission Details
Download Alignment File
Hide Colors
Send to ClustalW2_Phylogeny
(c)
sp|P20136|GSTM2_CHICK 0.18162
sp|P80031|GSTP1_PIG 0.28078
sp|P46426|GSTP_DIRIM 0.20519
sp|P10299|GSTP1_CAEEL 0.22269
sp|O35660|GSTM6_MOUSE 0.11904
sp|P16413|GSTMU_CAVPO 0.11291
sp|P19639|GSTM4_MOUSE 0.08709
sp|P10649|GSTM1_MOUSE 0.03749
sp|P04905|GSTM1_RAT 0.03131
sp|Q00285|GSTMU_CRILO 0.05345
sp|P15626|GSTM2_MOUSE 0.04725
sp|P30116|GSTMU_MESAU 0.06284
sp|Q9N0V4|GSTM1_BOVIN 0.09025
sp|P46439|GSTM5_HUMAN 0.08423
sp|Q80W21|GSTM7_MOUSE 0.08486
sp|Q5R8E8|GSTM2_PONAB 0.01722
sp|Q9BEB0|GSTM2_MACFU 0
sp|Q9TSM4|GSTM2_MACFA 0
sp|Q9TSM5|GSTM1_MACFA 0.00911
sp|Q03013|GSTM4_HUMAN 0.01842

Align.	DB:ID	Source
☑ 1	SP:GSTM1_MOUSE	Glutathione S-transferase Mu 1 OS=Mus musculus GN=Gstm1 PE=1 SV=2 *Cross-references and related information in:* ▶ Gene expression ▶ Small molecules ▶ Nucleotide sequences ▶ Genomes ▶ Samples & ontologies ▶ Molecular interactions ▶ Enzymes ▶ Protein families ▶ Literature ▶ Protein sequences ▶ Reactions, pathways & diseases
☑ 2	SP:GSTM1_RAT	Glutathione S-transferase Mu 1 OS=Rattus norvegicus GN=Gstm1 PE=1 SV=2 *Cross-references and related information in:* ▶ Small molecules ▶ Nucleotide sequences ▶ Samples & ontologies ▶ Enzymes ▶ Protein families ▶ Literature ▶ Macromolecular structures ▶ Protein sequences
☑ 3	SP:GSTMU_CRILO	Glutathione S-transferase Y1 OS=Cricetulus longicaudatus PE=2 SV=2 *Cross-references and related information in:* ▶ Small molecules ▶ Nucleotide sequences ▶ Samples & ontologies ▶ Enzymes ▶ Protein families ▶ Literature ▶ Protein sequences
☑ 4	SP:GSTM4_MOUSE	Glutathione S-transferase Mu 3 OS=Mus musculus GN=Gstm3 PE=1 SV=2 *Cross-references and related information in:* ▶ Gene expression ▶ Small molecules ▶ Nucleotide sequences ▶ Genomes ▶ Samples & ontologies ▶ Enzymes ▶ Protein families ▶ Literature ▶ Protein sequences ▶ Reactions, pathways & diseases
☑ 5	SP:GSTM2_MOUSE	Glutathione S-transferase Mu 2 OS=Mus musculus GN=Gstm2 PE=1 SV=2 *Cross-references and related information in:* ▶ Gene expression ▶ Small molecules ▶ Nucleotide sequences ▶ Genomes ▶ Samples & ontologies ▶ Molecular interactions ▶ Enzymes ▶ Protein families ▶ Literature ▶ Protein sequences ▶ Reactions, pathways & diseases

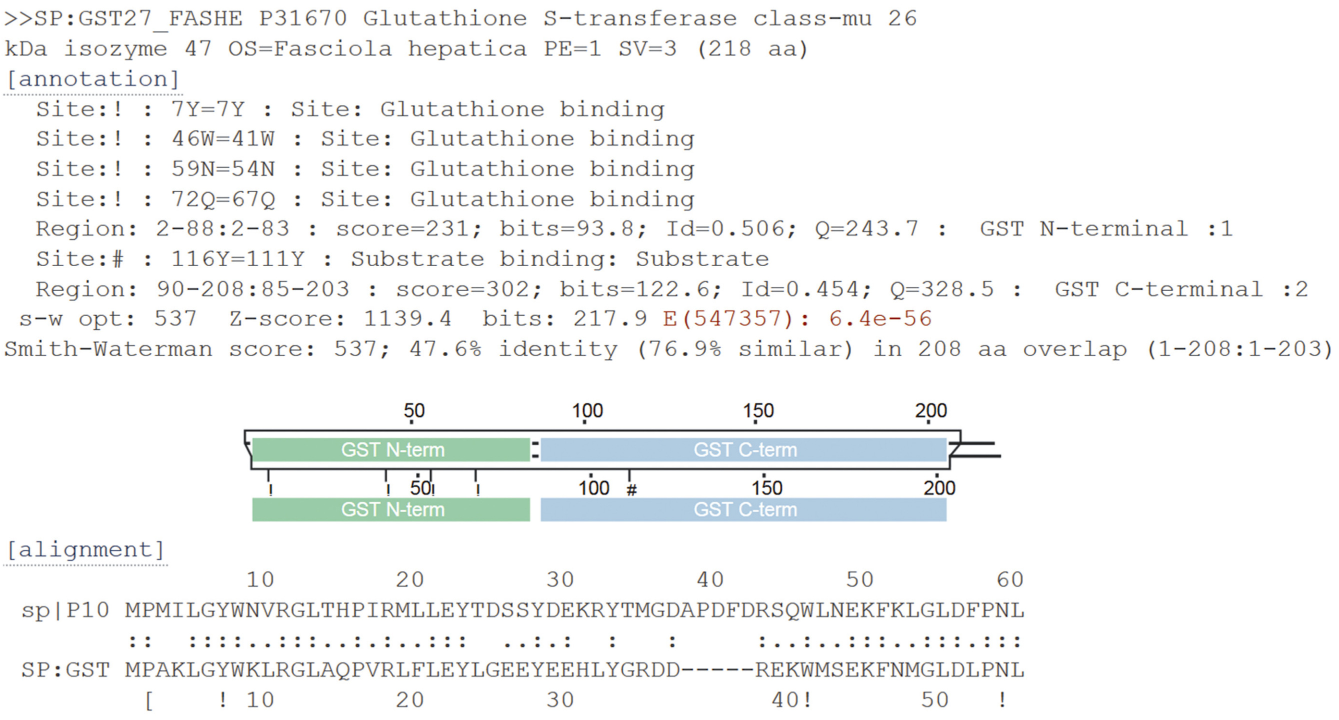
```
CLUSTAL O(1.2.1) multiple sequence alignment


sp|P20136|GSTM2_CHICK      MVVTLGYWDIRGLAHAIRLLLEYTETPYQERRYKAGPAPDFDPSDWI
sp|P16413|GSTMU_CAVPO      -PMTLGYWNIRGLTHPIRLILEYTNSGYEEKRYNMGDAPDYDRSQWI
sp|O35660|GSTM6_MOUSE      MPVTLGYWDIRGLGHAIRLLLEYTETGYEERRYAMGDAPDYDRSQWI
sp|P19639|GSTM4_MOUSE      MPMTLGYWNTRGLTHSIRLLLEYTDSSYEEKRYVMGDAPNFDRSQWI
sp|P10649|GSTM1_MOUSE      MPMILGYWNVRGLTHPIRMLLEYTDSSYDEKRYTMGDAPDFDRSQWI
sp|P04905|GSTM1_RAT        MPMILGYWNVRGLTHPIRLLLEYTDSSYEEKRYAMGDAPDYDRSQWI
sp|Q00285|GSTMU_CRILO      MPMILGYWNVRGLTNPIRLLLEYTDSSYEEKKYTMGDAPDSDRSQWI
sp|P15626|GSTM2_MOUSE      MPMTLGYWDIRGLAHAIRLLLEYTDTSYEDKKYTMGDAPDYDRSQWI
sp|P30116|GSTMU_MESAU      MPVTLGYWDIRGLAHAIRLLLEYTDTSYEEKKYTMGDAPNFDRSQWI
sp|Q9N0V4|GSTM1_BOVIN      MPMILGYWDIRGLAHAIRLLLEYTDTNYEERQYSVGDAPDYDRSQWI
sp|P46439|GSTM5_HUMAN      MPMTLGYWDIRGLAHAIRLLLEYTDSSYVEKKYTLGDAPDYDRSQWI
sp|Q80W21|GSTM7_MOUSE      MPMTLGYWDIRGLAHAIRLFLEYTDSSYEEKRYTMGDAPDYDQSQWI
sp|Q5R8E8|GSTM2_PONAB      MPIILGYWNIRGLAHSIRLLLEYTDSSYEEKKYMMGDAPDYDRSQWI
sp|Q9BEB0|GSTM2_MACFU      MPMTLGYWNIRGLAHSIRLLLEYTGSSYEEKKYTMGDAPDYDRSQWI
sp|Q9TSM4|GSTM2_MACFA      MPMTLGYWNIRGLAHSIRLLLEYTGSSYEEKKYTMGDAPDYDRSQWI
sp|Q9TSM5|GSTM1_MACFA      MPMTLGYWDIRGLAHAIRLLLEYTDSSYEEKKYTMGDAPDYDRSQWI
sp|Q03013|GSTM4_HUMAN      MSMTLGYWDIRGLAHAIRLLLEYTDSSYEEKKYTMGDAPDYDRSQWI
sp|P80031|GSTP1_PIG        PPYTITYFPVRGRCEAMRMLLADQDQSWKEEVVTMETWPPLK-----
sp|P46426|GSTP_DIRIM       MSYKLTYFPIRGLAEPIRLLLVDQGIKFTDEHIPKDDFVSIK-----
sp|P10299|GSTP1_CAEEL      MTLKLTYFDIHGLAEPIRLLLADKQVAYEDHRVTYEQWADIK-----
```

Figure 1. An example workflow from NCBI BLAST+ to Clustal Omega and construction of a phylogenetic tree. (**a**) Perform a NCBI BLAST+ similarity search and select sequence hits from the summary table to align with Clustal Omega; (**b**) Perform a simple phylogenetic analysis on the Clustal Omega alignment; (**c**) Visualise the phylogenetic tree.

such as HMMER 3 (30), R-COFFEE (31) and new data resources such as ENA Barcode and Geospatial databases (11). Further cross-resource integration will be available, such as additional annotations to sequence similarity results using the EBI Search Web Services (17) and visualisations using novel client-side technologies that render complex data faster and in more efficient ways than traditional server-side methods. Ensembl data (32) will be available via the NCBI BLAST+ service.

Some applications have been flagged for retirement from EMBL-EBI in 2015. These include ClustalW2 (27), DaliLite (33), DbClustal (34), MaxSprout (35), ReadSeq (36) and WU-BLAST (37). Further details will be announced on the web site.

Additional support for users in the future will include webinars and the production of video-based tutorials and other integrated online learning capabilities.



```
>>SP:GST27_FASHE P31670 Glutathione S-transferase class-mu 26
kDa isozyme 47 OS=Fasciola hepatica PE=1 SV=3 (218 aa)
[annotation]
  Site:! : 7Y=7Y : Site: Glutathione binding
  Site:! : 46W=41W : Site: Glutathione binding
  Site:! : 59N=54N : Site: Glutathione binding
  Site:! : 72Q=67Q : Site: Glutathione binding
  Region: 2-88:2-83 : score=231; bits=93.8; Id=0.506; Q=243.7 :  GST N-terminal :1
  Site:# : 116Y=111Y : Substrate binding: Substrate
  Region: 90-208:85-203 : score=302; bits=122.6; Id=0.454; Q=328.5 :  GST C-terminal :2
 s-w opt: 537  Z-score: 1139.4  bits: 217.9 E(547357): 6.4e-56
Smith-Waterman score: 537; 47.6% identity (76.9% similar) in 208 aa overlap (1-208:1-203)
```

```
[alignment]
                10        20        30        40        50        60
 sp|P10 MPMILGYWNVRGLTHPIRMLLEYTDSSYDEKRYTMGDAPDFDRSQWLNEKFKLGLDFPNL
        ::  ::::..:::..:.:..:::  ..:.:  :   :     :..:..:::..:::.:::
 SP:GST MPAKLGYWKLRGLAQPVRLFLEYLGEEYEEHLYGRDD-----REKWMSEKFNMGLDLPNL
         [    ! 10        20        30             40!       50   !
```

Figure 2. An example domain display from PSI-Search output, showing UniProt sequence features that are present in significantly aligned regions.

DISCUSSION

Having a tools framework for EMBL-EBI applications allows users access to a range of services through uniform interfaces and helps the maintenance of a robust, relevant service by enabling individual applications to be added, updated, or retired as required. Improvements to the web browser interface help usability and allow more complex analyses to be carried out through the provision of workflow mechanisms between tools. Integration of other resources such as EBI Search and dbfetch expands the resources the framework can draw on and facilitates user acquisition of biological data. New and updated tools and databases ensure that scientists have access to the most recent analyses and data available, while retirement of depreciated services helps to ensure that the application set is well maintained and resources are dedicated to the most relevant services.

Since becoming available in 2009, the framework has been used by academic and industry users for almost 260 million analysis jobs and the volume of usage has been increasing significantly with roughly 110 million analyses in 2014 alone. Web Services in particular lend themselves to integration in third party pipelines, and the applications have been of use to commercial and academic organisations as well as to other EMBL-EBI teams such as Ensembl Genomes, Pfam and UniProt. Where such integrations are present it is especially important not to break dependencies. So, a careful process of communication and change management is in place, including updates through a range of channels that include mailing lists, news feeds, web site announcements and Twitter.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Prof. William Pearson from the University of Virginia for his support and feedback and the web administrators and Systems team at EMBL-EBI for their assistance in the provision of the tools framework service.

FUNDING

Funding for open access charge: European Molecular Biology Laboratory.
Conflict of interest statement. None declared.

REFERENCES

1. Brooksbank,C., Bergman,M.T., Apweiler,R., Birney,E. and Thornton,J. (2014) The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Res.*, **42**, D18–D25.
2. Goujon,M., McWilliam,H., Li,W., Valentin,F., Squizzato,S., Paern,J. and Lopez,R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, **38**, W695–W699.
3. McWilliam,H., Li,W., Uludag,M., Squizzato,S., Park,Y.M., Buso,N., Cowley,A.P. and Lopez,R. (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.*, **41**, W597–W600.
4. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications.*BMC Bioinformatics*, **10**, 421–429.
5. Li,W., McWilliam,H., Goujon,M., Cowley,A., Lopez,R. and Pearson,W.R. (2012) PSI-Search: iterative HOE-reduced profile SSEARCH searching. *Bioinformatics*, **28**, 1650–1651.
6. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G. *et al.* (2014)

InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.

7. Mistry,J., Bateman,A. and Finn,R.D. (2007) Predicting active site residue annotations in the Pfam database.*BMC Bioinformatics*, **8**, 298–311.
8. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Söding,J.J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539–544.
9. Lassmann,T., Frings,O. and Sonnhammer,E.L. (2009) Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.*, **37**, 858–865.
10. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
11. Silvester,N., Alako,B., Amid,C., Cerdeño-Tárraga,A., Cleland,I., Gibson,R., Goodgame,N., Ten Hoopen,P., Kay,S., Leinonen,R. *et al.* (2014) Content discovery and retrieval services at the European Nucleotide Archive. *Nucleic Acids Res.*, **43**, D23–D29.
12. Kersey,P.J., Allen,J.E., Christensen,M., Davis,P., Falin,L.J., Grabmueller,C., Hughes,D.S., Humphrey,J., Kerhornou,A., Khobova,J. *et al.* (2014) Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res.*, **42**, D546–D552.
13. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
14. Mitchell,A., Chang,H.Y., Daugherty,L., Fraser,M., Hunter,S., Lopez,R., McAnulla,C., McMenamin,C., Nuka,G., Pesseat,S. *et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.*, **43**, D213–D221.
15. Finn,R.D., Bateman,A., Clements,J., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Heger,A., Hetherington,K., Holm,L., Mistry,J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
16. Lopez,R., Cowley,A., Li,W. and McWilliam,H. (2014) Using EMBL-EBI Services via Web Interface and Programmatically via Web Services. *Curr Protoc Bioinformatics.*, **48**, 3.12.1–3.12.50.
17. Valentin,F., Squizzato,S., Goujon,M., McWilliam,H., Paern,J. and Lopez,R. (2010) Fast and efficient searching of biological data resources–using EB-eye. *Brief Bioinform.*, **11**, 375–384.
18. Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.
19. Guerra-Assunção,J.A. and Enright,A.J. (2010) MapMi: automated mapping of microRNA loci. *BMC Bioinformatics*, **11**, 133–139.
20. Rokas,A. (2011) Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum Likelihood (RAXML) Program. *Curr. Protoc. Mol. Biol.*, doi:10.1002/0471142727.mb1911s96.
21. Howe,K., Davis,P., Paulini,M., Tuli,M.A., Williams,G., Yook,K., Durbin,R., Kersey,P. and Sternberg,P.W. (2012) WormBase: Annotating many nematode genomes. *Worm*, **1**, 15–21.
22. Burge,S.W., Daub,J., Eberhardt,R., Tate,J., Barquist,L., Nawrocki,E.P., Eddy,S.R., Gardner,P.P. and Bateman,A. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.*, **41**, D226–D232.
23. Etzold,T., Ulyanov,A. and Argos,P. (1996) SRS: information retrieval system for molecular biology data banks. *Meth. Enzymol.*, **266**, 114–128.
24. Quevillon,E., Silventoinen,V., Pillai,S., Harte,N., Mulder,N., Apweiler,R. and Lopez,R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Res.*, **33**, W116–W120.
25. Fredman,D., Munns,G., Rios,D., Sjöholm,F., Siegfried,M., Lenhard,B., Lehväslaiho,H. and Brookes,A.J. (2004) HGVbase: a curated resource describing human DNA variation and phenotype relationships. *Nucleic Acids Res.*, **32**, D516–D519.
26. Kersey,P.J., Duarte,J., Williams,A., Karavidopoulou,Y., Birney,E. and Apweiler,R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, **4**, 1985–1988.
27. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
28. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2444–2448.
29. Gómez,J., García,L.J., Salazar,G.A., Villaveces,J., Gore,S., García,A., Martín,M.J., Launay,G., Alcántara,R., Del-Toro,N. *et al.* (2013) BioJS: an open source JavaScript framework for biological data visualization. *Bioinformatics*, **29**, 1103–1104.
30. Mistry,J., Finn,R.D., Eddy,S.R., Bateman,A. and Punta,M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.*, **41**, e121.
31. Wilm,A., Higgins,D.G. and Notredame,C. (2008) R-Coffee: a method for multiple alignment of non-coding RNA. *Nucleic Acids Res.*, **36**, e52.
32. Cunningham,F., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
33. Holm,L. and Park,J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics*, **16**, 566–567.
34. Thompson,J.D., Plewniak,F., Thierry,J. and Poch,O. (2000) DbClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res.*, **28**, 2919–2926.
35. Holm,L. and Sander,C. (1991) Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors. *J. Mol. Biol.*, **218**, 183–194.
36. Gilbert,D. (2003) Sequence file format conversion with command-line readseq. *Curr. Protoc. Bioinformatics*, doi:10.1002/0471250953.bia01es00.
37. Lopez,R., Silventoinen,V., Robinson,S., Kibria,A. and Gish,W. (2003) WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.*, **31**, 3795–3798.