EchinoDB: An update to the web-based application for genomic and transcriptomic data on echinoderms

Varnika Mittal^{1*} (ORCID: 0000-0002-3033-4891); Robert W. Reid¹ (ORCID: 0000-0001-7125-7151); Denis Jacob Machado¹ (ORCID: 0000-0001-9858-4515); Vladimir Mashanov² (ORCID: 0000-0001-6875-7518); and Daniel A. Janies¹ (ORCID: 0000-0002-7890-9906).

¹Department of Bioinformatics and Genomics, College of Computing and Informatics, University of North Carolina at Charlotte, 9331 Robert D. Snyder Rd, Charlotte, NC 28223, USA.

²Wake Forest Institute for Regenerative Medicine, 91 Technology Way NE, Winston-Salem, NC 27101, USA.

Abstract

Here we release a new version of EchinoDB (<https://echinodb.uncc.edu>). EchinoDB is a database of genomic and transcriptomic data on echinoderms. The initial database consisted of groups of 749,397 orthologous and paralogous transcripts arranged in orthoclusters by sequence similarity. The new version of EchinoDB includes RNA-seq data of the brittle star *Ophioderma brevispinum* and high-quality genomic assembly data of the green sea urchin *Lytechinus variegatus*. In addition, we enabled keyword searches for annotated data and installed an updated version of Sequenceserver to allow BLAST searches. The data are downloadable in FASTA format. The first version of EchinoDB appeared in 2016 and was implemented in GO on a local server. The new version has been updated using R Shiny to include new features and improvements in the application. Furthermore, EchinoDB now runs entirely in the cloud for increased reliability and scaling. EchinoDB enjoys a user base drawn from the fields of phylogenetics, developmental biology, genomics, physiology, neurobiology, and regeneration. As use cases, we illustrate how EchinoDB is used in discovering pathways and gene regulatory networks involved in the tissue regeneration process.

Database | Echinoderms | Echinoids | Gene Family | Genome | Ophiuroids | Orthocluster | Ortholog | Paralog | Transcriptome
*Correspondence: vmittal@uncc.edu.

Introduction

EchinoDB (1) is an open-source web-based application accessed via <https://echinodb.uncc.edu> (19). An effective bioinformatics tool must keep up with new data, advances in software, server architecture, and programming languages. To scale well with increasing amounts of data, reliability and number of users, we have updated EchinoDB to take advantage of cloud services. We have also added new datasets along with new features such as sequence downloading functionality, Basic Local Alignment Search Tool (BLAST) (13) and keyword search to efficiently query the annotated data.

EchinoDB is a repository of genomic and amino acid sequence data from across the phylum Echinodermata. The primary dataset is organized into clusters of orthologous genes (orthoclusters). Additional datasets include RNA-seq data of the brittle star *Ophioderma brevispinum* (Say, 1825) (Echinodermata: Ophiuroidea: Ophiacanthida: Ophiodermatidae) (5) and genome assembly data of green sea urchin *Lytechinus variegatus* (Lamarck, 1816) (Echinodermata: Echinoidea: Camarodonta: Toxopneustidae) (6).

EchinoDB serves a large and diverse community of researchers focused on different aspects of biology including phylogenetics, developmental biology, genomics, physiology, neurobiology, and regeneration. As use cases, we illustrate how EchinoDB is used in discovering pathways and gene regulatory networks involved in the tissue regeneration process.

Echinoderms are a phylum of marine invertebrate deuterostomes and thus share a deep common ancestor with vertebrates (9, 10, 14). However, unlike vertebrates, many echinoderm species can regenerate all their tissue types after injury without inducing cancers (15). Therefore, with the correct resources, including genomic and transcriptomic data, echinoderms can be developed as a model organism to research regulatory mechanisms for regeneration.

The raw data in EchinoDB were generated by RNA-seq profiling adult tissues from 42 echinoderm specimens from 24 orders and 37 families (1). The RNA-seq data were assembled, and translated into protein sequences. The *de novo* transcriptome assembly consisted of 1,198,706 amino acid sequences across 42 species. These data were clustered using OrthoMCL, an algorithm for grouping orthologous protein sequences based on sequence similarity (11). The resulting orthoclusters database consisted of groups of 749,397 orthologous and paralogous transcripts. The orthocluster data have been annotated through sequence similarity with respect to a purple sea urchin *Strongylocentrotus purpuratus* genome which was the best annotated echinoderm genome at the origin of the project (12). These annotations provide the basis for keyword searches.

The EchinoDB web-based application was initially built in the GO programming language in 2016 (1). The updated EchinoDB has been rewritten in R Shiny (3), which is highly extensible, easy to write, and maintain compared to the previous implementation. R Shiny supports faster development of user interfaces by providing a framework that requires no knowledge of scripting languages like HTML, CSS or JavaScript. In addition, the new version facilitates the inclusion of new data from collaborations. We have taken advantage of this feature to extend the application's capabilities to expose new data such as the *Lytechinus* (6) and *Ophioderma* (5) sequences to interface for BLAST searches via Sequenceserver (4).

Materials and Methods

In the latest EchinoDB release, we added a text box that allows users to conduct searches using accession numbers and other keywords with or without the use of wildcard entries. Results include protein sequence(s), annotated description(s), known NCBI GenInfo Identifier (GI ids), and orthocluster(s). The annotations are borrowed from aligning our sequences to the well-characterized protein sequence dataset of *Strongylocentrotus purpuratus* (i.e., sequences attributed to taxon 7668 in NCBI's RefSeq, accessed in August 2012). These results can be further filtered by name or GenInfo Identifier (GI ids) in the search box in the top right corner. Additionally, users are able to expand or narrow their search based on taxonomic class, order, and family via toggle switches. Figure 1 depicts the design created in R Shiny for the EchinoDB application. Each row of the result table represents an orthocluster with the sequence similarity count or total hits. The number of hits is clickable, facilitating the viewing and downloading of related amino acid and nucleotide sequences in a FASTA format.

Figure 1: Screenshot of our EchinoDB landing page (19), available at <https://echinodb.uncc.edu>. Users could search against all extant echinoderm classes, orders, and families or un-toggle to retrieve information for a particular taxon.

New data resources for ophiuroid and echinoid

We have added RNA-seq data of *O. brevispinum* (5), a common brittle star found in shallow waters of the western Atlantic Ocean ranging from Canada to Venezuela. We have also added high-quality genome assembly data of *L. variegatus* (6), a sea urchin found in shallow waters throughout the western Atlantic Ocean ranging from the United States to Venezuela.

OphiuroidDB

We have provided a transcriptome dataset, translated and annotated using BLASTX functionality against the NCBI collection of predicted proteins of *S. purpuratus* (25) and protein models from UniProt's Swiss-Prot (16) and NCBI's RefSeq (17). The application can be accessed via “Link to *O. brevispinum* transcriptome” in EchinoDB. This page is referred to as “OphiuroidDB” and it serves the data exclusive to the brittle star, *O. brevispinum* (20).

These data were first used to characterize the downstream genes controlled by the Notch signaling pathway, which is believed to play a role in brittle star regeneration (5). Then, the raw sequencing reads were submitted to the NCBI as a GEO dataset under the accession number GSE142391 (5, 24), and transcriptome sequences can be downloaded from OphiuroidDB. A total of 30,149 genes were identified and included in the application.

EchinoidDB

EchinoidDB facilitates access to a recently published annotated high-quality chromosomal-scale genome assembly of *L. variegatus* (6, 21). The data (*Lvar_3.0*) includes 27,232 nucleotide and protein sequences which were annotated using BLASTP (18) against UniProt Swiss-Prot (16), *S. purpuratus* (23) and non-*S.*

purpuratus RefSeq invertebrate protein models (17). These annotations can be downloaded from EchinoidDB.

Use case example

To demonstrate the usage of EchinoDB and associated resources, we can search for genes associated with the Notch signaling pathway. This should serve as a relevant example because this pathway is required for regeneration in echinoderms (5). Knowledge of the Notch signaling pathway is important because it is highly conserved in the animal kingdom and regulates various cellular events, including proliferation, differentiation, fate specification, and cell death (26-29).

For example, the user can search for Notch-related genes and obtain the corresponding sequences and metadata from our web resources. To do this, the user can search for the keyword Notch in our web resources to locate Notch-related sequences in brittle stars and other echinoderms. The results include NCBI's accession numbers, other unique identifiers, descriptions, start and end positions, and other details depending on the application used. In addition, the interface allows the user to visualize any selected sequence or cluster of sequences from the related repository.

Results and Discussion

EchinoDB is designed to facilitate comparative transcriptomic studies of deeply-sampled clades of echinoderms (1). EchinoDB is re-factored in R Shiny. R Shiny is extensible and code developed with R Shiny can be integrated, readily with CSS themes, HTML widgets, and scripting languages like JavaScript. In addition, as R Shiny is widely adopted, the code can be modified and tuned at later stages in the development cycle by many developers.

EchinoDB is hosted using the Nginx web server (2) in Amazon Web Services (AWS) (7). AWS offers on-demand cloud computing services to build your own web-based applications independent of university information technology bureaus.

Application features

As many echinoderm "omic" data are not yet well annotated, blast search is an important complement to keyword or accession search.

Using Sequenceserver to run BLAST

EchinoDB contains an instance of Sequenceserver (4), a web-based BLAST server that supports sequence similarity searches against nucleotide and protein sequence databases. EchinoDB provides nucleotide and protein databases to facilitate sequence similarity searches using default or user-selected parameters.

Integration with BLAST allows users of EchinoDB to search data resources with strings of the query sequence. Figure 2 illustrates Sequenceserver for BLAST functionality and can be accessed via "Link to BLAST Sequenceserver" in the EchinoDB application.

SequenceServer
2.0.0.rc8
Help & Support

Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...

Nucleotide databases
[Select all]

☐ Lytechinus_variegatus_Nucleotide_Sequences
☐ Ophioderma_brevispinum_Nucleotide_Sequences
☐ OrthoCluster_Nucl_Seqs

Protein databases
[Select all]

☐ Lytechinus_variegatus_Protein_Sequences
☐ Ophioderma_brevispinum_Protein_Sequences
☐ OrthoCluster_Prot_Seqs

Advanced parameters:

eg: -evalue 1.0e-5 -num_alignments 100
?

☐ Open results in new tab

BLAST

Easy BLASTing with SequenceServer.
Tweet

Please cite relevant data sources and: Priyam et al. (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases.

Figure 2: Screenshot of our Sequenceserver integrated in EchinoDB (22). Users can perform BLAST searches against nucleotide and protein sequences of included datasets in the application via <https://echinodb.uncc.edu/sequenceserver/>.

Literature

We provide a repository that contains links to the research papers associated with EchinoDB by their title. The literature repository is updated regularly.

Additional data

A link is added in the Literature section to allow users to download data. For example, one dataset provides evidence that *Xyloplax* sp. is a velatid (order of class Asteroidea) asteroid rather than a new class (8). The data included in EchinoDB includes tables and phylogenomic data from large amounts of transcriptome data used in this paper. The additional data repository is updated regularly.

Usage and documentation

EchinoDB, EchinoidDB, and OphiuroidDB user manuals (Additional files 2–4: Files S1–3, respectively) are available in a tab named "Documentation" in the EchinoDB website. The user manuals are downloadable and provide instructions with screenshots to assist the user in navigating through the application.

Figure 3 demonstrates the usefulness of EchinoDB and some of its outputs. The figure depicts the usage in a step-by-step process by locating individual or clusters of Notch-related amino acid sequences in brittle stars and other echinoderms.

6

Conclusion

The updated EchinoDB provides, via a cloud-based server, additional tools and data from collaborations and our lab that can be of interest to a variety of scientific communities. One of our focal points in the future is to extend the genomic, transcriptomic, and orthocluster contents of EchinoDB.

Supplementary Materials

Assembled sequences and orthoclusters are available in EchinoDB (<https://echinodb.uncc.edu>; 19). Raw reads from the various echinoderm species are available in NCBI's SRA (see accession numbers in Additional file 1: Table S1). Additionally, the user manuals for EchinoDB, EchinoidDB, and OphiuroidDB are available as Additional file 2: File S1, Additional file 3: File S2, and Additional file 4: File S3, respectively. Additional files are available in Zenodo (<https://doi.org/10.5281/zenodo.5803547>).

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R15 GM128066-01. We acknowledge the support of several entities of the University of North Carolina at Charlotte including the College of Computing and Informatics, the Graduate School, the Department of Bioinformatics and Genomics, University Research Computing, and the Bioinformatics Research Center. Additionally, we thank Benjamin Stalcup for issuing a SSL certificate for EchinoDB application and Steven Blanchard for helping us with certificates deployments and instructions for web server setup. We also thank Shantoy Hansel and Jan Kofsky for their testing of the application. Finally, we thank Greg Wray for providing the genomic data of the green sea urchin, *Lytechinus variegatus* (6).

Bibliography

1. Janies DA, Witter Z, Linchangco GV, Foltz DW, Miller AK, Kerr AM, Jay J, Reid RW, Wray GA. EchinoDB, an application for comparative transcriptomics of deeply-sampled clades of echinoderms. BMC Bioinformatics. 2016; doi: 10.1186/s12859-016-0883-2.
2. Nginx web server. <https://www.nginx.com/resources/wiki/start/>. Accessed 21 January 2021.
3. Shiny from RStudio. <https://shiny.rstudio.com/>. Accessed 31 October 2020.
4. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, Chowdhary H, Pieniak I, Maynard LJ, Gibbins MA, Moon H. Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol. Biol. Evol. 2019; 36(12):2922-4; doi: 10.1093/molbev/msz185.
5. Mashanov V, Akiona J, Khoury M, Ferrier J, Reid R, Machado DJ, Zueva O, Janies D. Active Notch signaling is required for arm regeneration in a brittle star. PLoS One. 2020; 15(5):e0232981; doi: 10.1371/journal.pone.0232981.
6. Davidson PL, Guo H, Wang L, Berrio A, Zhang H, Chang Y, Soborowski AL, McClay DR, Fan G, Wray GA. Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. Genome Biol. Evol. 2020; 12(7):1080-6; doi: 10.1093/gbe/evaa101.
7. Amazon Web Services. <https://aws.amazon.com/>. Accessed 10 December 2020.
8. Linchangco Jr GV, Foltz DW, Reid R, Williams J, Nodzak C, Kerr AM, Miller AK, Hunter R, Wilson NG, Nielsen WJ, Mah CL. The phylogeny of extant starfish (Asteroidea: Echinodermata) including *Xyloplax*, based on comparative transcriptomics. Mol. Phyl. Evol. 2017; 115:161-70; doi: 10.1016/j.ympev.2017.07.022.
9. Swalla BJ, Smith AB. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. Philos. T. R. Soc. B: Biol. Sci. 2008; 363(1496):1557-68; doi: 10.1098/rstb.2007.2246.

10. Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, De Rosa R. The new animal phylogeny: reliability and implications. *P. Natl. A. Sci.* 2000; 97(9):4453-6; doi: 10.1073/pnas.97.9.4453.
11. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003; 13(9):2178-89. Version 2.0.9; doi: 10.1101/gr.1224503.
12. NCBI Protein RefSeqs for taxon 7668. RefSeq assembly accession GCA_000002235.4, Spur_3.1. <http://www.ncbi.nlm.nih.gov>. Accessed August 2012.
13. Madden T. The BLAST sequence analysis tool. The NCBI handbook. 2003 Aug 13.
14. Blair JE, Hedges SB. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* 2005; 22(11):2275-84; doi: 10.1093/molbev/msi225.
15. Dupont S, Thorndyke MC. Growth or differentiation? Adaptive regeneration in the brittle star *Amphiura filiformis*. *Jpn. J. Exp. Biol.* 2006; 209(19):3873-81; doi: 10.1242/jeb.02445.
16. Bateman A. Uniprot: a universal hub of protein knowledge. In: *Protein Science 2019* (Vol. 28, pp. 32-32). HOBOKEN: WILEY.
17. O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016; 44(D1):D733-45; doi: 10.1093/nar/gkv1189.
18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009; 10(1):1-9; doi: 10.1186/1471-2105-10-421.
19. EchinoDB Database. <https://echinodb.uncc.edu>. Accessed 23 December 2021.
20. Ophiuroid Database. <https://echinodb.uncc.edu/BStarApp/>. Accessed 23 December 2021.
21. Echinoid Database. <https://echinodb.uncc.edu/SUrchinApp/>. Accessed 23 December 2021.
22. EchinoDB Sequenceserver Integration. <https://echinodb.uncc.edu/sequenceserver/>. Accessed 23 December 2021.
23. *Strongylocentrotus purpuratus* Spur_5.0 Protein Models. www.ncbi.nlm.nih.gov/assembly/GCF_000002235.5. Accessed 26 February 2020.
24. Geo Dataset; GSE142391. Active Notch signaling is required for arm regeneration in a brittle star. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142391>. Accessed 15 November 2021.
25. The NCBI collection of predicted proteins of the sea urchin *Strongylocentrotus purpuratus*. ftp://ftp.ncbi.nih.gov/genomes/Strongylocentrotus_purpuratus/protein/. Accessed 26 February 2020
26. Eve Gazave, Pascal Lapébie, Gemma S. Richards, Frédéric Brunet, Alexander V. Ereskovsky, Bernard M. Degnan, Carole Borchiellini, Michel Vervoort, and Emmanuelle Renard. Origin and evolution of the Notch signalling pathway: an overview from eukaryotic genomes. *BMC Evol Biol*, 9:249, 2009. doi: 10.1186/1471-2148-9-249.
27. Heather Marlow, Eric Roettinger, Michiel Boekhout, and Mark Q. Martindale. Functional roles of Notch signaling in the cnidarian *Nematostella vectensis*. *Dev Biol*, 362(2):295–308, 2012. doi: 10.1016/j.ydbio.2011.11.012.
28. Michael J. Layden and Mark Q. Martindale. Non-canonical Notch signaling represents an ancestral mechanism to regulate neural differentiation. *Evodevo*, 5:30, 2014. doi: 10.1186/2041-9139-5-30.

29. María Belén Favarolo and Silvia L López. Notch signaling in the division of germ layers in bilaterian embryos. *Mech Dev*, 154:122–144, 2018. doi: 10.1016/j.mod.2018.06.005.