



BioMaster: An Integrated Database and Analytic Platform to Provide Comprehensive Information About BioBrick Parts

Beibei Wang^{1,2*}, Huayi Yang^{1†}, Jianan Sun^{1†}, Chuhaou Dou^{1†}, Jian Huang^{1,2} and Feng-Biao Guo^{1,2}

¹ School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China,

² Centre for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China

OPEN ACCESS

Edited by:

Trevor Carlos Charles,
University of Waterloo, Canada

Reviewed by:

Yi Xiong,
Shanghai Jiao Tong University, China
Rahmi Lale,
Norwegian University of Science
and Technology, Norway

*Correspondence:

Beibei Wang
bbwang@uestc.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 31 August 2020

Accepted: 04 January 2021

Published: 21 January 2021

Citation:

Wang B, Yang H, Sun J, Dou C,
Huang J and Guo F-B (2021)
BioMaster: An Integrated Database
and Analytic Platform to Provide
Comprehensive Information About
BioBrick Parts.
Front. Microbiol. 12:593979.
doi: 10.3389/fmicb.2021.593979

Synthetic biology seeks to create new biological parts, devices, and systems, and to reconfigure existing natural biological systems for custom-designed purposes. The standardized BioBrick parts are the foundation of synthetic biology. The incomplete and flawed metadata of BioBrick parts, however, are a major obstacle for designing genetic circuit easily, quickly, and accurately. Here, a database termed BioMaster <http://www.biomaster-uestc.cn> was developed to extensively complement information about BioBrick parts, which includes 47,934 items of BioBrick parts from the international Genetically Engineered Machine (iGEM) Registry with more comprehensive information integrated from 10 databases, providing corresponding information about functions, activities, interactions, and related literature. Moreover, BioMaster is also a user-friendly platform for retrieval and analyses of relevant information on BioBrick parts.

Keywords: synthetic biology, genetic circuit design, database integration, BioBrick parts, iGEM Registry

INTRODUCTION

As a branch of the emerging biological sciences in the 21st century, synthetic biology is dedicated to design and construct novel biological parts, devices, and systems for specific purposes, and to make the process of engineering biology easier (Heinemann and Panke, 2006; Bartley et al., 2017). One of the important directions for synthetic biology is to build standard systems using standard devices which are comprised of standard parts. Therefore, to standardize biological parts and devices, Tom Knight at MIT introduced the concept of BioBrick standard in 2003 (Smolke, 2009).

BioBrick parts are biological parts that conform to a restriction-enzyme-based assembly standard, containing defined prefix and suffix sequences which can be recognized by specific restriction endonucleases (Canton et al., 2008; Shetty et al., 2008). Different assembly methods are constantly being reported, such as 3A assembly (Shetty et al., 2008), Gibson assembly (Gibson et al., 2009), Golden Gate assembly (Engler et al., 2008), MIDAS (van Dolleweerd et al., 2018), and CasHRA (Zhou et al., 2016). The standardization of BioBrick parts makes the compatibility and assembly of different biological parts more efficient, and the development of tools for biological computer-aided design (CAD) possible. Since the first successfully synthetic toggle switch and repressilator (Gardner et al., 2000), several related CAD tools, aiming at making the design easier and high throughput, have been developed, such as GenoCAD (Cai et al., 2010), TinkerCell

(Chandran et al., 2009), Gene Designer (Villalobos et al., 2006), and Cello (Nielsen et al., 2016; Chen et al., 2020). GenoCAD is a computer tool that helps users assemble genetic circuits from a rich library of parts, compliantly with any of six implemented BioBrick standards. TinkerCell is a visual modeling tool that supports a hierarchy of biological parts. Gene Designer is a software for fast and easy design of synthetic DNA segments and uses advanced algorithms for codon optimization. And Cello provides a design environment for genetic circuit design automation. Thus, engineering biology, both manually and with CAD platforms, requires large amounts of BioBrick data of high quality (Nora et al., 2019). However, these powerful design processes are not supported by adequate information of standardized biological parts (Decoene et al., 2018).

So far, several data registries and repositories have been established to collect biological parts. For example, the Virtual Parts Repository consists of about 3000 biological parts (Cooling et al., 2010; Misirli et al., 2014), PAMDB contains 118 circuits and 165 parts (Huynh and Tagkopoulos, 2016), JBEI-ICE includes about 2302 items (Ham et al., 2012), SEVA-DB has a number of 185 vectors (Martinez-Garcia et al., 2015), while the international Genetically Engineered Machine (iGEM) Registry of standard biological parts (Smolke, 2009; Vilanova and Porcar, 2014) is, currently, the largest BioBrick database, containing over 40,000 items, cataloged as promoters, ribosomal binding sites (RBS), coding sequences, terminators, and so on. However, the iGEM Registry lacks joint analysis and retrieval with other databases, which are numerous and contain various biochemical, genetic, and molecular biological information.

Therefore, in this study, to complement and improve the metadata of iGEM BioBrick parts, we integrated 11 traditional biological databases (**Supplementary Table S1**), including iGEM Registry, UniProt (Bateman et al., 2017), QuickGO (Binns et al., 2009), KEGG (Kanehisa et al., 2017), BioGRID (Oughtred et al., 2019), BRENDA (Jeske et al., 2019), ExplorEnz (McDonald et al., 2009), STRING (Szklarczyk et al., 2017), PubMed (Sayers et al., 2020), EPD (Perier et al., 2000), and PromEC (Hershberg et al., 2001). These databases supplement the information of functions, sites, interactions, feature keys, and references. The integrated database, termed BioMaster, also provides a more user-friendly platform for searching, browsing, and analyzing BioBrick information.

MATERIALS AND METHODS

In this study, the UniProt database was taken as a hub for connecting the iGEM Registry with other databases. We first downloaded the entire BioBrick database from the iGEM Registry¹. The database contains the parts table, the sequence, and features of BioBrick parts (O'Donovan et al., 2002). Then BLAST was used to find matched sequence in the UniProt database. The other nine databases were then integrated through UniProt unique identifier. Finally, we developed a user-friendly database,

termed BioMaster, for easy searching, browsing, and analyzing all these BioBrick information (**Figure 1**).

BioMaster² was developed with PHP, HTML, CSS, JavaScript, and Python on a Linux platform. Laravel (version 7.0.0³), a PHP web framework, was used in the back end. Bootstrap⁴ and Vue⁵ were mainly used to build interactive pages of the front end of the website. The data were stored in a MySQL database and can be automatically updated. We also packed all data, programs, and the operating environment, and uploaded to the docker hub⁶. Elastic search⁷, a safe, efficient, and well-structured search engine, was used to implement full-text search, fuzzy search, and keyword search.

Data Sorting of the Sequence Alignment Results

There was no information associated with other databases in the iGEM Registry. To get the corresponding information, UniProt was used as a hub for connecting the iGEM Registry with other nine databases. The connection between the iGEM Registry and UniProt was established by sequence alignment on featured (functional) sequences in iGEM Registry or full BioBrick sequences in case of no available feature information. The sequence alignment was performed by BLAST (Altschul et al., 1997). Each alignment result of BLAST includes three main parameters: Score, expect value (E-value), and Identity. The Score is the sum of similarity of each pair between two sequence, so it is related to the length of the sequence. The E-value is the expected number of high score pairs and decreases exponentially as the score increases. The Identity is the proportion of which two sequences have the same residues at the same position after alignment. The results of BLAST are always displayed

²<http://www.biomaster-uestc.cn>

³<https://laravel.com/>

⁴<https://v2.bootcss.com/>

⁵<https://cn.vuejs.org/>

⁶<https://hub.docker.com/u/igem2019uestcsoftware>

⁷<https://www.elastic.co/cn/>

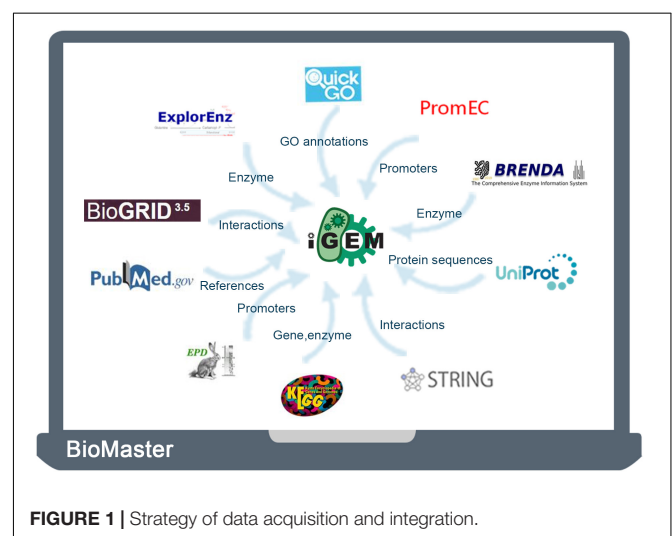


FIGURE 1 | Strategy of data acquisition and integration.

¹http://parts.igem.org/Registry_API

with a descending Score. In most cases, however, the order is different if a different standard is used for sorting. Therefore, to comprehensively consider all three parameters, we used their weighted average (p), calculated as the following:

$$p = Identity \times w_1 + \frac{Score}{S_{max} \times length} \times w_2 + \frac{\log_6(-\lg(Evalue))}{3} \times w_3,$$

where w_1 , w_2 , and w_3 are their weights, respectively. As the Identity is ranged from 0 to 1, the Score and E-value were adjusted to this range. The Score was divided by the maximum in the substitution matrix (S_{max}) (Altschul, 1991) and the length of the query sequence. After taking the negative logarithm of 10 of E-value, the value falls into a range of 1–200. To adjust it to 0–1, we calculated the logarithm of 6 again and divide it by 3 as shown in the above equation.

To determine the value of w_1 , w_2 , and w_3 , we tested from 0.1 to 0.8 with an interval of 0.1, respectively. About 200 items of BLAST results, manually selected, formed a training set, to obtain optimal combination. We found that the screening accuracy was highest when $w_1 = 0.1$, $w_2 = 0.5$, and $w_3 = 0.4$.

To validate the data sorting model, we labeled about 8500 BLAST results, which were classified by our model and support vector machine (SVM). The threefold validation accuracy of SVM is around 0.93, lower than the accuracy of our model (about 0.98). The area under ROC curve (AUC) value of our model is 0.99, which is also better than the AUC value of SVM (0.94) (Supplementary Figure S1A). Meanwhile, the average of Score, E-value, and Identity of the positive items also demonstrates that the performance of our model is better than SVM (Supplementary Figure S1B).

RESULTS AND WEB INTERFACE

To complete the information of iGEM BioBrick parts, we developed a database termed BioMaster, which provides a user-friendly query interface and a set of tools for a comprehensive overview of BioBrick parts.

The Amount of Data

BioMaster has all the data from the iGEM Registry. In addition, it also integrates corresponding information from 10 databases (UniProt, STRING, BioGRID, KEGG, QuickGO, PubMed, BRENDA, ExplorEnz, EPD, and PromEC) to complement the iGEM BioBrick datasheet. The iGEM Registry contains 47,934 entries of BioBrick information, among which 44,565 entries have feature information. Through sequence alignment with UniProt database and data screening, we found corresponding 3747 items in UniProt. Via their UniProt IDs, information from other nine databases was integrated: 100,000 items of interaction information from STRING and BioGRID, 94,260 pieces of annotations from QuickGO, 6203 pieces of gene information from KEGG, 2694 pieces of homology information from KO, information and annotations of 1336 enzymes from

KEGG ENZEMY, BRENDA, and ExplorEnz, some promoters information from EPD and PromEC, and more than 1 million associated references from PubMed. According to the iGEM Registry, BioBrick parts are divided into 25 categories (Supplementary Table S2), of which the eight main categories are promoters, RBS, coding sequences, translational units, terminators, DNA, plasmids, and composite parts.

Search Module

BioMaster supports a total of four search methods: ID search, Keyword search, Team Wiki search, and sequence alignment by BLAST (Figure 2B). The ID search automatically recognizes IGEM_ID, EPD_ID, UniProt_ID, and Gene_Name. Keywords search returns parts in the iGEM Registry related to the keywords. Team Wiki can search for the team names and keywords related to the iGEM teams. Searching for an amino acid sequence or nucleotide sequence is also practicable via BLAST against BioMaster database.

The search results include a total of no more than 1000 related items sorted into different groups (Figure 2C), such as terminators, DNA, RNA, and reporters. The detailed information can be accessed by selecting any item of results. It shows the information obtained from the iGEM Registry (Figure 2D) as well as other databases related to the part in detail (Figure 2E). In order to display the interaction between different parts or parts and proteins clearly, cytoscape.js was implemented to show interaction diagrams. The information of the corresponding BioBrick parts will be shown by clicking on the nodes and links. In addition, synthetic biology open language (SBOL) is used to visualize the feature information and display it as a graph. The search results can be freely downloaded and is open for user's comments. Detailed usage tutorials are available on our website in the form of videos⁸.

Application

Taking the part BBa_K1745002 for example, Figure 2A shows all the information for this part in the iGEM Registry, from which we found that this part is an enzyme, aralkylamine N-acetyltransferase (AANAT), but there is almost no information other than the sequence and features. In our database, BBa_K1745002 was matched to the UniProt_ID Q16613 correctly (Figure 2E), which provides the information about its function, catalytic activity, structure, related reference, and so on. Via the UniProt_ID, massive information, such as related pathways, optimal experimental conditions, KM values, IC50 values, inhibitors, and interaction networks, was integrated from KEGG, Brenda, String, and BioGrid. Therefore, this part is better characterized in our database.

Tools for Analyses and Visualization of BioBrick Parts

Several tools for analyses and visualization of BioBrick parts were integrated in BioMaster. We integrated a JAVA-based visual aiding design software, SBOL designer, into BioMaster to

⁸<http://www.biomaster-uestc.cn/public/index.php/main/tutorial>

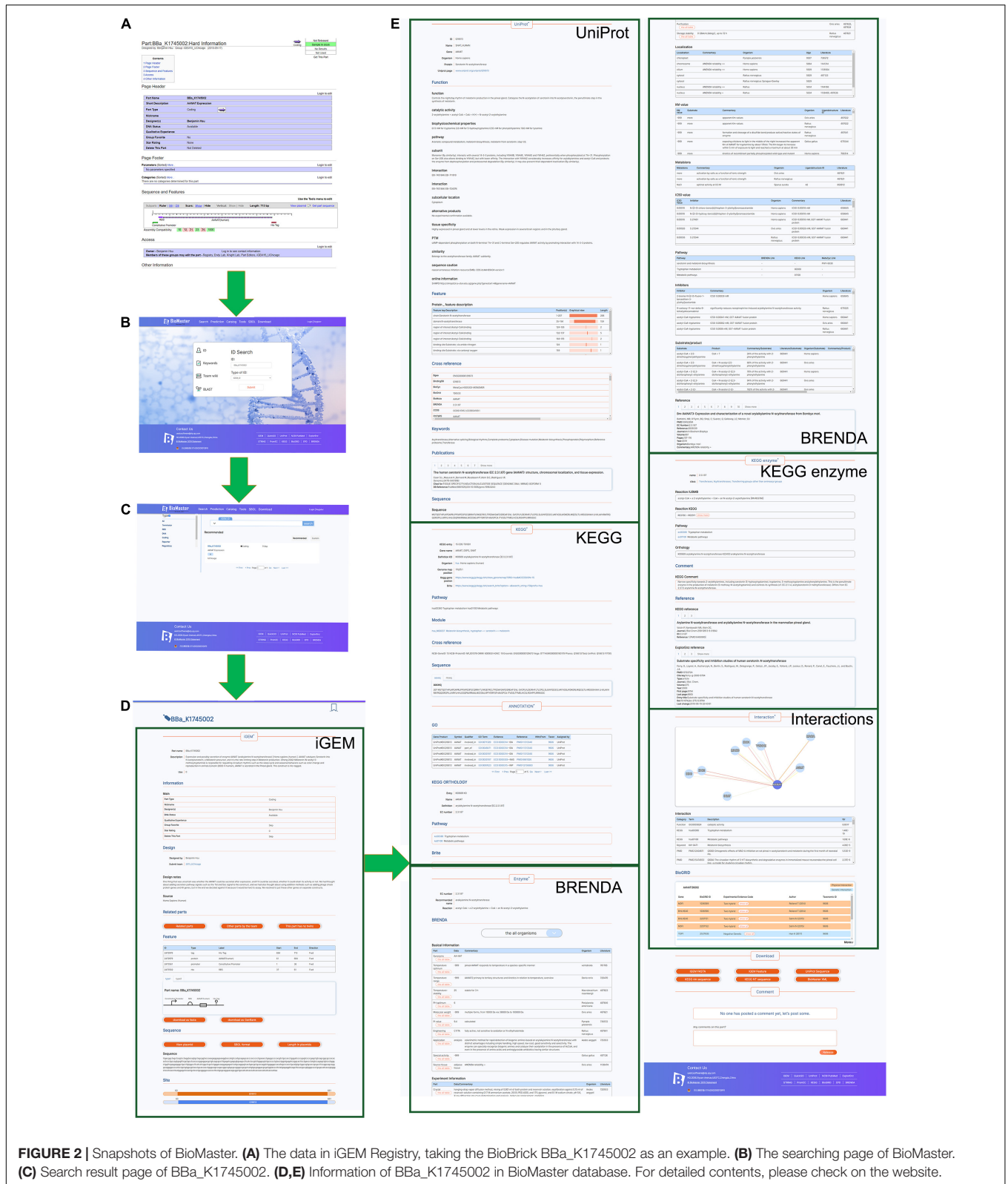
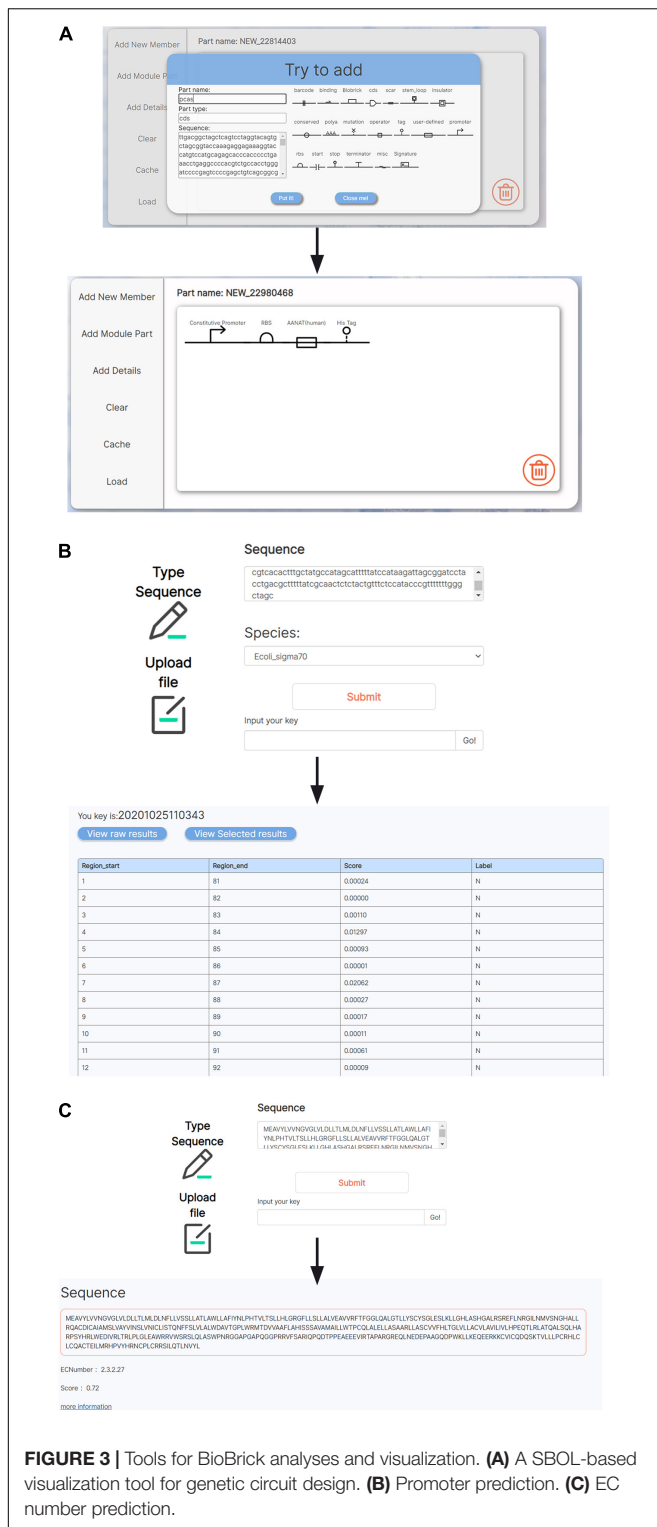


FIGURE 2 | Snapshots of BioMaster. **(A)** The data in iGEM Registry, taking the BioBrick BbA_K1745002 as an example. **(B)** The searching page of BioMaster. **(C)** Search result page of BbA_K1745002. **(D,E)** Information of BbA_K1745002 in BioMaster database. For detailed contents, please check on the website.

provide visualization and designing of biological components (Figure 3A) (Roehner et al., 2016; Madsen et al., 2019). On the other hand, concerning the limit amount of BioBrick

entries, BioMaster provides two prediction tools to conduct bioinformatic exploration on unknown sequences: the promoter prediction (Umarov and Solovyev, 2017) and the enzyme



commission (EC) number prediction (Dalkiran et al., 2018). The promoter prediction implemented uses convolutional neural network to recognize both prokaryotic and eukaryotic promoters (Figure 3B). The EC prediction was integrated into BioMaster to predict the enzymatic function for uncharacterized protein

sequence. An EC number is given as the prediction result, and more information can be obtained in BRENDA via the EC number (Figure 3C).

SUMMARY

The design of genetic circuit requires reliable, comprehensive, and complete information of BioBrick parts. To complete the BioBrick metadata of the iGEM Registry, we developed a database called BioMaster. Various information from 10 popular databases was integrated to the iGEM parts with diverse and detailed information. Meanwhile, BioMaster provides convenient retrieval, user-friendly interface, and potential predictions. In the future, we will integrate more data in breadth and depth from other synthetic biological databases and publications to connect the massive amounts of data pieces (Nora et al., 2019), and further improve the analysis tools on BioMaster. It is hoped that BioMaster could provide a platform to connect and standardize scattered biological part data, and to make designing of new biological systems easier both automatically and manually.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: <http://www.biomaster-uestc.cn>.

AUTHOR CONTRIBUTIONS

HY, JS, and CD collected the data, analyzed the data, developed the web, and wrote the manuscript. JH and F-BG designed the database and wrote the manuscript. BW designed the database, analyzed the data, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 31971176 and 31800616) and the Fundamental Research Funds for the Central Universities (No. A03018023601045).

ACKNOWLEDGMENTS

Thanks to the contribution of iGEM team members of UESTC-software-2018 and UESTC-software-2019.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.593979/full#supplementary-material>

REFERENCES

- Altschul, S. F. (1991). Amino-acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* 219, 555–565. doi: 10.1016/0022-2836(91)90193-a
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Bartley, B. A., Kim, K., Medley, J. K., and Sauro, H. M. (2017). Synthetic biology: engineering living systems from biophysical principles. *Biophys. J.* 112, 1050–1058. doi: 10.1016/j.bpj.2017.02.013
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., et al. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169.
- Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., and Apweiler, R. (2009). QuickGO: a web-based tool for gene ontology searching. *Bioinformatics* 25, 3045–3046. doi: 10.1093/bioinformatics/btp536
- Cai, Y., Wilson, M. L., and Peccoud, J. (2010). GenoCAD for iGEM: a grammatical approach to the design of standard-compliant constructs. *Nucleic Acids Res.* 38, 2637–2644. doi: 10.1093/nar/gkq086
- Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26, 787–793. doi: 10.1038/nbt1413
- Chandran, D., Bergmann, F. T., and Sauro, H. M. (2009). TinkerCell: modular CAD tool for synthetic biology. *J. Biol. Eng.* 3:19. doi: 10.1186/1754-1611-3-19
- Chen, Y., Zhang, S., Young, E. M., Jones, T. S., Densmore, D., and Voigt, C. A. (2020). Genetic circuit design automation for yeast. *Nat. Microbiol.* 5, 1349–1360. doi: 10.1038/s41564-020-0757-2
- Cooling, M. T., Rouilly, V., Misirli, G., Lawson, J., Yu, T., Hallinan, J., et al. (2010). Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics* 26, 925–931. doi: 10.1093/bioinformatics/btq063
- Dalkiran, A., Rifaioğlu, A. S., Martin, M. J., Cetin-Atalay, R., Atalay, V., and Dogan, T. (2018). ECPred: a tool for the prediction of the enzymatic functions of protein sequences based on the EC nomenclature. *BMC Bioinform.* 19:334. doi: 10.1186/s12859-018-2368-y
- Decoene, T., De Paepe, B., Maertens, J., Coussement, P., Peters, G., De Maeseneire, S. L., et al. (2018). Standardization in synthetic biology: an engineering discipline coming of age. *Crit. Rev. Biotechnol.* 38, 647–656. doi: 10.1080/07388551.2017.1380600
- Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3:e3647. doi: 10.1371/journal.pone.0003647 doi: 10.1371/journal.pone.0003647
- Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. doi: 10.1038/35002131
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J., and Keasling, J. D. (2012). Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* 40:e141. doi: 10.1093/nar/gks531
- Heinemann, M., and Panke, S. (2006). Synthetic biology—putting engineering into biology. *Bioinformatics* 22, 2790–2799. doi: 10.1093/bioinformatics/btl469
- Hershberg, R., Bejerano, G., Santos-Zavaleta, A., and Margalit, H. (2001). PromEC: an updated database of *Escherichia coli* mRNA promoters with experimentally identified transcriptional start sites. *Nucleic Acids Res.* 29, 277–277. doi: 10.1093/nar/29.1.277
- Huynh, L., and Tagkopoulou, I. (2016). A parts database with consensus parameter estimation for synthetic circuit design. *ACS Synth. Biol.* 5, 1412–1420. doi: 10.1021/acssynbio.5b00205
- Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019). BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 47, D542–D549.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361.
- Madsen, C., Goni Moreno, A., Umesh, P., Palchick, Z., Roehner, N., Atallah, C., et al. (2019). Synthetic biology open language (SBOL) version 2.3. *J. Integr. Bioinform.* 16:20190025.
- Martinez-Garcia, E., Aparicio, T., Goni-Moreno, A., Fraile, S., and de Lorenzo, V. (2015). SEVA 2.0: an update of the standard european vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* 43, D1183–D1189.
- McDonald, A. G., Boyce, S., and Tipton, K. F. (2009). ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.* 37, D593–D597.
- Misirli, G., Hallinan, J., and Wipat, A. (2014). Composable modular models for synthetic biology. *ACM J. Emerg. Tech. Com.* 11, 1–19. doi: 10.1145/2631921
- Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., et al. (2016). Genetic circuit design automation. *Science* 352:aac7341.
- Nora, L. C., Westmann, C. A., Martins-Santana, L., Alves, L. D. F., Oliveira Monteiro, L. M., Guazzaroni, M.-E., et al. (2019). The art of vector engineering: towards the construction of next-generation genetic tools. *Microb. Biotechnol.* 12, 125–147. doi: 10.1111/1751-7915.13318
- O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., and Apweiler, R. (2002). High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief. Bioinform.* 3, 275–284. doi: 10.1093/bib/3.3.275
- Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., et al. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47, D529–D541.
- Perier, R. C., Praz, V., Junier, T., Bonnard, C., and Bucher, P. (2000). The eukaryotic promoter database (EPD). *Nucleic Acids Res.* 28, 302–303. doi: 10.1093/nar/28.1.302
- Roehner, N., Beal, J., Clancy, K., Bartley, B., Misirli, G., Grunberg, R., et al. (2016). Sharing structure and function in biological design with SBOL 2.0. *ACS Synth. Biol.* 5, 498–506. doi: 10.1021/acssynbio.5b00215
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., et al. (2020). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 48, D9–D16.
- Shetty, R. P., Endy, D., and Knight, T. F. Jr. (2008). Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2:5.
- Smolke, C. D. (2009). Building outside of the box: iGEM and the BioBrick parts Foundation. *Nat. Biotechnol.* 27, 1099–1102. doi: 10.1038/nbt1209-1099
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368.
- Umarov, R. K., and Solovyev, V. V. (2017). Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. *PLoS One* 12:e0171410. doi: 10.1371/journal.pone.0171410 doi: 10.1371/journal.pone.0171410
- van Dolleweerd, C. J., Kessans, S. A., Van de Bittner, K. C., Bustamante, L. Y., Bundela, R., Scott, B., et al. (2018). MIDAS: a modular DNA assembly system for synthetic biology. *ACS Synth. Biol.* 7, 1018–1029. doi: 10.1021/acssynbio.7b00363
- Vilanova, C., and Porcar, M. (2014). iGEM 2.0-refoundations for engineering biology. *Nat. Biotechnol.* 32, 420–424. doi: 10.1038/nbt.2899
- Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinform.* 7:285. doi: 10.1186/1471-2105-7-285
- Zhou, J., Wu, R., Xue, X., and Qin, Z. (2016). CasHRA (Cas9-facilitated homologous recombination assembly) method of constructing megabase-sized DNA. *Nucleic Acids Res.* 44:e124. doi: 10.1093/nar/gkw475

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Yang, Sun, Dou, Huang and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.