

JBEI Performance Metric for FY16: Develop an improved metabolic engineering method for modifying microorganisms for biofuel production from cellulosic sugars

Q3 Metric: New tools available for facilitating the design of engineered pathways in microorganisms for biofuel production from cellulosic sugars.

1. BACKGROUND

It is mission-critical for JBEI to develop new tools that facilitate and improve the design of engineered pathways and microorganisms (as well as plants) for the production of biofuels from cellulosic sugars. While tools that address JBEI-needs are prioritized for development, most (if not all) of the tools deployed to overcome JBEI research challenges have wide applicability across academia, government, and industry. Through this broader use of JBEI tools (and their corresponding commercialization), the DOE further leverages JBEI in support of the greater U.S. bioeconomy. As such, it remains important that putative applications beyond the immediate JBEI mission are kept in mind during the tool design and development process.

In this report, we present a selection of tools developed at JBEI to extend research capabilities and boost productivity through facilitated and improved biological systems design: from a (digital as well as physical) repository of microorganisms, sequences, parts, and seeds; to the visual design of DNA constructs and the automatic generation of optimized protocols to build them; to the systematic storage and visualization of complex functional-genomics data; to the prediction of phenotypic outcomes that would result from genetic modifications. This report highlights emerging development efforts as well as published tools that have already positively and significantly impacted the broader research community.

2. ICE/VectorEditor: Repository (and tools) for strains, sequences, parts, and seeds

ICE (Inventory of Composable Elements) is a physical as well as a freely open-source (<https://github.com/JBEI/ice>) web-based digital repository platform for microbial strains, sequences, parts, and (*Arabidopsis*) seeds¹ (see Figure 1). ICE entries (whether strains, sequences, parts, or seeds) can be linked to the Experiment Data Depot (EDD, see below), for quick access to associated experiments. ICE supports community standards including the Synthetic Biology Open Language (SBOL) data-exchange format^{2, 3} as well as the SBOL Visual⁴ representation of annotated DNA sequences (via Pigeon⁵). ICE provides tools for the storage and visualization of sequence validation results (e.g., Sanger, PacBio/MiSeq) and the auto-annotation of DNA sequences. ICE integrates VectorEditor for the viewing, editing, and manual-annotation of DNA sequences. VectorEditor, also freely open-source (<https://github.com/JBEI/openVectorEditor>), is additionally used to specify parts and visualize designed DNA constructs for DeviceEditor and DIVA (see below).

ICE entries may have physical sample locations associated with them. A desktop application interfaces ICE with Matrix tube and plate 2D-barcode scanners for sample location curation and tracking management. ICE plasmid entries can be linked to their corresponding AddGene (<http://www.addgene.org/>) entries for an additional mode of public sample access.

Figure 1. ICE user interface. A plasmid entry⁶ in the Public JBEI Registry (<https://public-registry.jbei.org/entry/4931>) is shown. At left are featured collections of entries associated with JBEI publications. In the middle, plasmid-entry information is displayed, followed by links to open the sequence in VectorEditor, download the sequence (in SBOL 2, Genbank, FASTA formats), or auto-annotate the sequence; followed by the Pigeon (SBOL Visual) and VectorEditor renderings of the plasmid. This particular plasmid entry is publicly shared across the Web of Registries. At right are links to sequence validation, physical sample, and experiment data information; followed by access permission settings, and the collections that contain the plasmid.

Multiple repositories across the world can exchange information in a granular access-controlled manner using ICE’s “Web of Registries” functionality. From within a given repository (e.g., the JBEI Private Registry) it is possible to browse or search for entries in other repositories (e.g., the JBEI Public Registry). Data is managed so that access to an entry in one repository may be provided (with edit or read-only permissions) to a specific user on another repository. This Web of Registries “private remote” collaboration functionality is accessible to third-party tools and other repository platforms (i.e., those not running ICE) via ICE’s Application Program Interface.

The JBEI Private Registry stores over 46,000 entries. The JBEI Public Registry (<https://public-registry.jbei.org> - anyone can register for access) stores the strains, sequences, parts, and seeds associated with JBEI’s publications and features over 5,000 entries (1,500 of these have been created by over 1,000 non-JBEI registered users). In addition to the JBEI Private and Public Registries, there are other ICE repositories at the DOE Joint Genome Institute (JGI), Boston University, Manchester University, and SynBERC. The journal *ACS Synthetic Biology* has announced an initiative in which it had deployed its own ICE repository for authors to store and share all of the strains, sequences, parts, and seeds associated with their published manuscripts⁷.

3. DeviceEditor/j5: Biological computer-aided design and DNA assembly automation

DeviceEditor is a web-based biological computer-aided-design canvas that enables users to visually design combinatorial DNA constructs⁸ (see Figure 2). DeviceEditor depicts part types (e.g., promoter, terminator) using SBOL Visual⁴ standard glyphs. DeviceEditor integrates a subset of the Eugene⁹ design-rule specification language to enable the user to eliminate specific combinations that prior biological knowledge or Design of Experiments approaches suggest should not be pursued. DeviceEditor integrates ICE and EDD (see above and below) for characterization data-driven part suggestions, VectorEditor (see above) for part specification and design preview visualization, GeneDesign¹⁰ for back-translation and codon juggling, j5 for DNA assembly design automation, and DIVA (see below) for design storage.

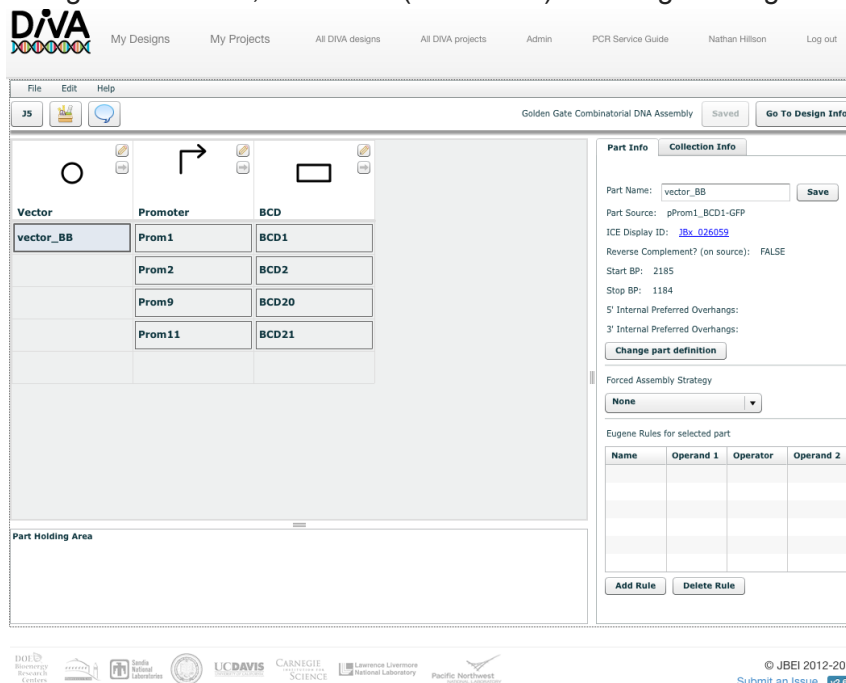


Figure 2. DeviceEditor user interface. Shown at left is a 16-variant combinatorial design (4 promoters combined with 4 bicistronic-devices)¹¹, with columns left to right corresponding to DNA 5' to 3', and rows top to bottom corresponding to different part variants. SBOL Visual glyph column headers indicate the type of parts contained within a column. At right is information for the selected “vector_BB” part, indicating its location within its source sequence (with a link to its entry within ICE, along with its forced assembly strategy and the Eugene rules specified for it.

j5 automates the design and cost-optimization of modern scar-less combinatorial DNA assembly methods (including Type IIS and flanking homology methods such as Golden Gate, Gibson, and Yeast Assembly)¹², and has been cited over 100 times. j5 receives design information from DeviceEditor, and interfaces with PR-PR^{11, 13} to enable the downstream execution of the designed DNA assembly protocols on laboratory automation devices including liquid handling robotics and microfluidic platforms¹⁴⁻¹⁶. A U.S. Patent for j5 has been issued¹⁷, and j5 has been exclusively-licensed for commercial-use and distribution to TeselaGen (a JBEI startup company founded on j5 technology, whose customers include biofuels, bio-based products, agriculture,

and therapeutics companies). Through the JBEI public j5 web-server (<https://j5.jbei.org>) there are more than 1,800 academic, non-profit, and government registered users of DeviceEditor and j5 at more than 460 institutions worldwide.

4. DIVA: DNA design, implementation, and validation automation

DIVA (Design, Implementation, and Validation Automation) is a web-based collaborative biological design and fabrication platform (see Figure 3) that increases research efficiency and productivity through enabling a division of labor (separation of design from fabrication tasks) and design aggregation to scales benefitting from laboratory automation devices. Users design with DeviceEditor (see above), and then (through DIVA) submit their designs for technical feasibility review and price estimates, submit the price-estimated designs for PI budgetary approval, and then track the fabrication process through sequence validation and construct delivery. A small team of technical staff, leveraging laboratory automation devices, build and verify all designs submitted for construction. DIVA integrates with ICE to store all of the resulting constructs. Over 3,000 constructs have been designed within DIVA at JBEI across all 4 of its Scientific and Technology Divisions, and over 1,100 of these have been submitted for construction. While the DIVA platform is primarily an internal research tool for JBEI at present, it has been deployed at the JGI, and has been licensed for non-commercial use to the University of Edinburgh and the National University of Singapore.

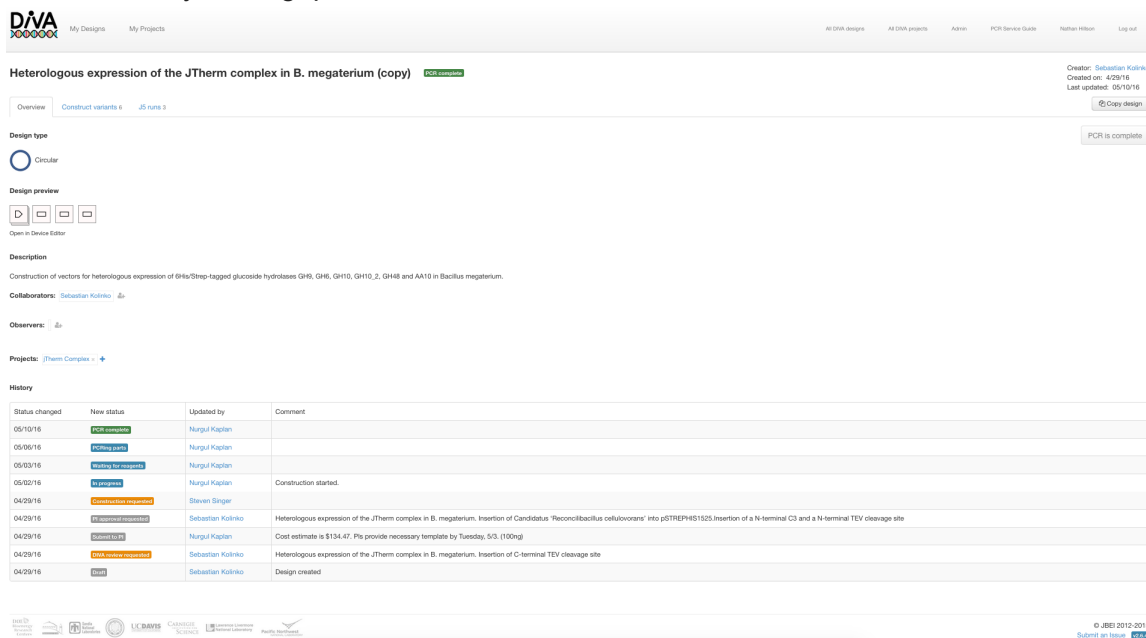


Figure 3. DIVA user interface. Shown is the overview tab for a 6-variant plasmid design. Tabs for construct variants and j5 runs (not shown), respectively contain statuses and links to ICE for each variant, and the statuses and results from all j5 runs for the design. Shown at top left is the design type (circular or linear) along with a SBOL Visual preview of the design, which may be opened in DeviceEditor. At middle left, the collaborators (edit privileges), observers (read privileges), and the projects associated with the design are shown. Shown at bottom left, the temporal history of the design, from creation, to technical review and cost estimate, to PI budgetary-approval, to the various stages of construction through completion.

5. jQMM: Algorithms for actionable metabolic engineering insights

Modeling of microbial metabolism is a topic of growing importance in biotechnology. Mathematical modeling helps provide a mechanistic understanding for the studied biotechnology process, separating the main drivers from the circumstantial ones, bounding the outcomes of experiments and guiding engineering approaches. Among different modeling schemes, the quantification of intracellular metabolic fluxes (*i.e.*, the rate of each reaction in cellular metabolism) is of particular interest for metabolic engineering because it describes how carbon and energy flow throughout the cell. Additionally, the effective use of the ever more readily available and abundant 'omics data (*e.g.*, transcriptomics, proteomics and metabolomics), is an imperative for the field.

The JBEI Quantitative Metabolic Modeling (jQMM, Birkel *et al* in preparation) library provides an open-source, python-based framework for modeling internal metabolic fluxes and leveraging other 'omics data for the scientific study of cellular metabolism and bioengineering purposes (see Figure 4). Firstly, it presents a complete toolbox for performing two different types of flux analysis that are typically disjoint: Flux Balance Analysis and ^{13}C Metabolic Flux Analysis. Moreover, it introduces the capability to use ^{13}C labeling experimental data to effectively constrain comprehensive genome-scale models through a technique called two-scale ^{13}C Metabolic Flux Analysis (2S- ^{13}C MFA)¹⁸. In addition, the library includes a demonstration of a method that uses proteomics data to produce actionable engineering strategies to increase biofuel production⁶. Finally, the use of the jQMM library is illustrated through several Jupyter iPython notebook demonstration files that provide complete reproducibility.

The jQMM library has been used to suggest gene knock-out targets in *S. cerevisiae* that increase fatty acid yield by ~40% (Ghosh *et al*, in preparation), to investigate the metabolism of glucose-repression (Shymansky *et al*, in preparation), and to use proteomics data to increase limonene and bisabolene production by ~40% and ~200% respectively⁶. As an open source software project, we hope it will attract additions from the community and grow with the rapidly changing field of metabolic flux analysis. Some jQMM functionality has been run on the KBase platform (<https://kbase.us>). Future efforts aim to make the full jQMM library available via KBase.

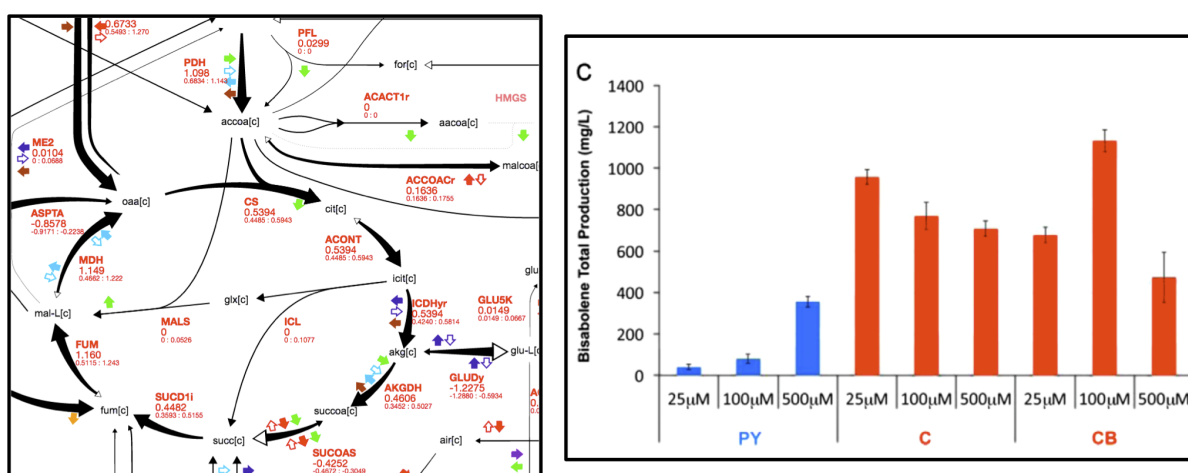


Figure 4. jQMM metabolic flux profiles and increases in bisabolene production. Shown at left are metabolic flux profiles obtained through 2S ¹³C MFA by constraining genome-scale models with ¹³C labeling data¹⁸. Right panel shows the increase (in red) of bisabolene production using plasmid designs obtained through PCAP⁶.

6. Experiment Data Depot (EDD): Storing the data generated by JBEI's experiments

Biology has undergone a radical transformation in the 20th century: whereas it had previously focused on classifying and explaining biological behavior, the advent of synthetic biology provides the possibility of changing the instruction set of biological entities and modify their behavior¹⁹. However, while our capability to create new biological designs improves quickly²⁰, our ability to predict the outcomes of engineered biological systems remains infantile²¹.

Greater predictability and reproducibility is enabled by efficient data and metadata collection and sharing²², as provided by the EDD (Morell *et al* in preparation). Computational biologists, modelers and data scientists worldwide are limited by experimental data with which to validate a growing number of computational approaches designed to predict biological behavior²³⁻²⁵. At the same time, the post-genomic revolution has provided experimentalists with large-scale data sets of 'omics data that are orders of magnitude larger than they are accustomed to analyzing. Hence, the robust exchange of data between experimentalist and computational specialists, as enabled by the EDD, is crucial to improving predictive models of biological systems. These data exchanges are particularly important for the further development of synthetic biology and biological computer aided design: as has been shown by Davidsohn *et al.*²⁶, the rigorous characterization of pathway parts enables the accurate prediction of the behavior of full pathways. These improved predictive capabilities can significantly improve strain-engineering efforts at JBEI and other research centers.

The EDD is an online tool designed as a repository of standardized data describing experiments performed at JBEI. EDD can easily uptake experimental data, provide quick quality checks and visualization of these data, and provide these data through several file standards and a RESTful API. The input of data to EDD is conveniently done through automated data streams: each of these input streams automatically parse the standard outputs of the instruments most commonly used at JBEI. So far, EDD can accept proteomics, HPLC, Biolector and metabolomics (LC-MS) data. EDD provides convenient data visualization (see Figure 5), which can be used as a quick sanity check for imported data. Furthermore, EDD can check the integrity of HPLC data through a built-in carbon balance calculation. Since data is stored internally in a standardized manner in a relational database, all data output is standardized. Outputs can be provided in terms of different standardized files (SBML or CSV) or through a RESTful API.

The EDD is not a laboratory information management system (LIMS): it does not store raw data or MS spectra, it only stores actionable data that can be immediately interpreted by the experimentalist (e.g., metabolite concentrations as opposed to MS spectra). At this moment, EDD stores data for fermentation experiments and enzyme characterization. These data can be used as input for the metabolic modeling techniques contained in jQMM (see above): ¹³C MFA, FBA, and PCAP, among others. The EDD provides a way to further leverage DOE's investment in JBEI by providing data that can be further analyzed outside of JBEI.

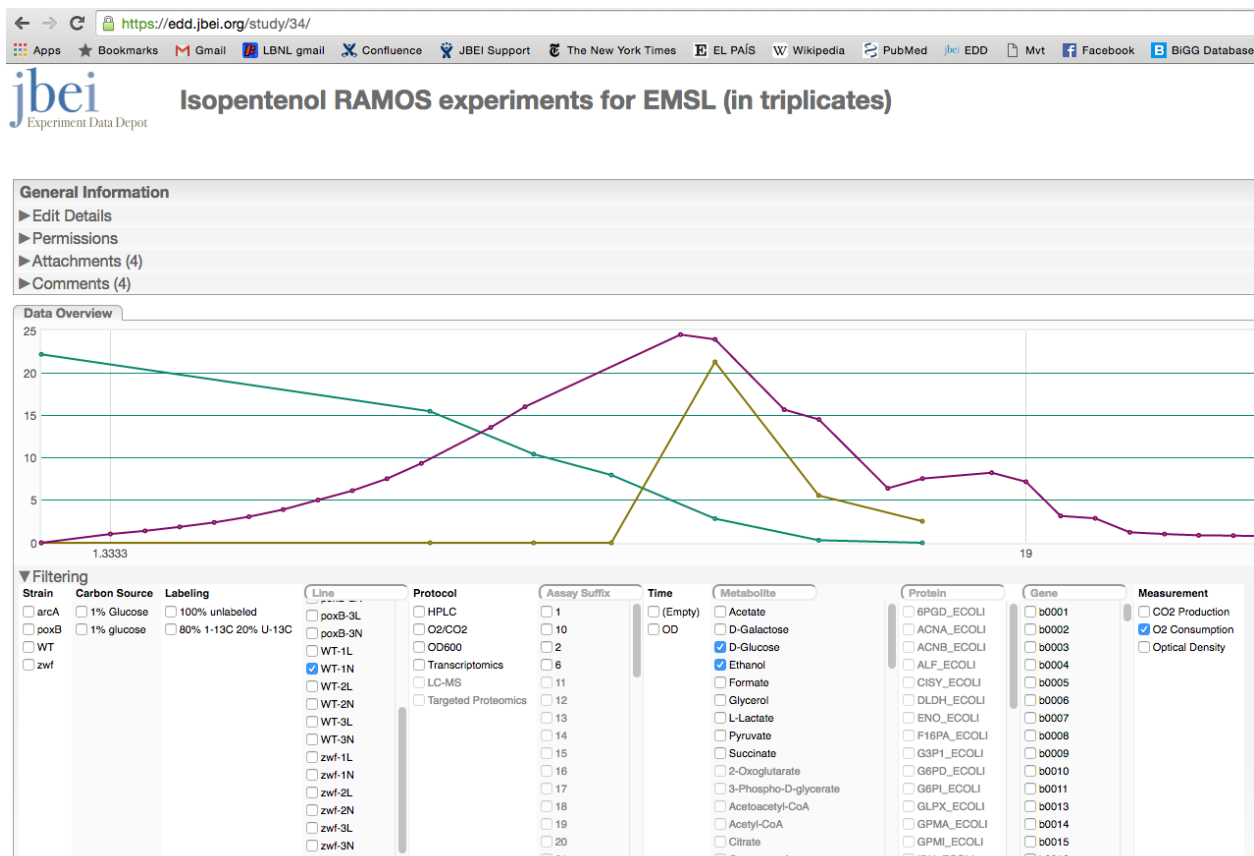


Figure 5. Experiment Data Depot user interface. Shown is the experimental data for an isopentenol fermentation run. At center is a visualization of relevant data, in this case: glucose concentration, oxygen input, and ethanol concentration. Other data such as proteomics and transcriptomics can be visualized by clicking in the menus below the graph.

7. Arrowland: Visualizing multi 'omics data

Much of this century's biology will focus around productively using genomics and functional genomics data. Biological researchers find themselves buried in a deluge of data orders or magnitude more abundant than they are accustomed to. They need methods able to produce *actionable items* from that data: methods able to condense all the information into the next *action* they need to perform in order to obtain their goals, along with a mechanistic explanation. However, we lack good tools to even visualize these functional genomics data effectively.

We developed Arrowland as a general tool to view functional genomics data (e.g., transcriptomics, proteomics, metabolomics and fluxomics), using the latest web technology. Arrowland provides a multi-scale map with different levels of resolution that can be interactively explored, in the same way as Google Maps does for geographical maps. Arrowland allows for the overlay of different types of 'omics data (see Figure 6), and the comparison of different time points or conditions. Furthermore, the latest feature in Arrowland allows users to click on a reaction and predict the flux profile if that reaction were to be knocked out, therefore providing some actionable results of use for metabolic engineers.

Arrowland has very recently finished its prototype stage and is now available at <https://arrowland.jbei.org/> for general use within JBEI. Current efforts are directed towards providing a web tool available for users outside of JBEI.

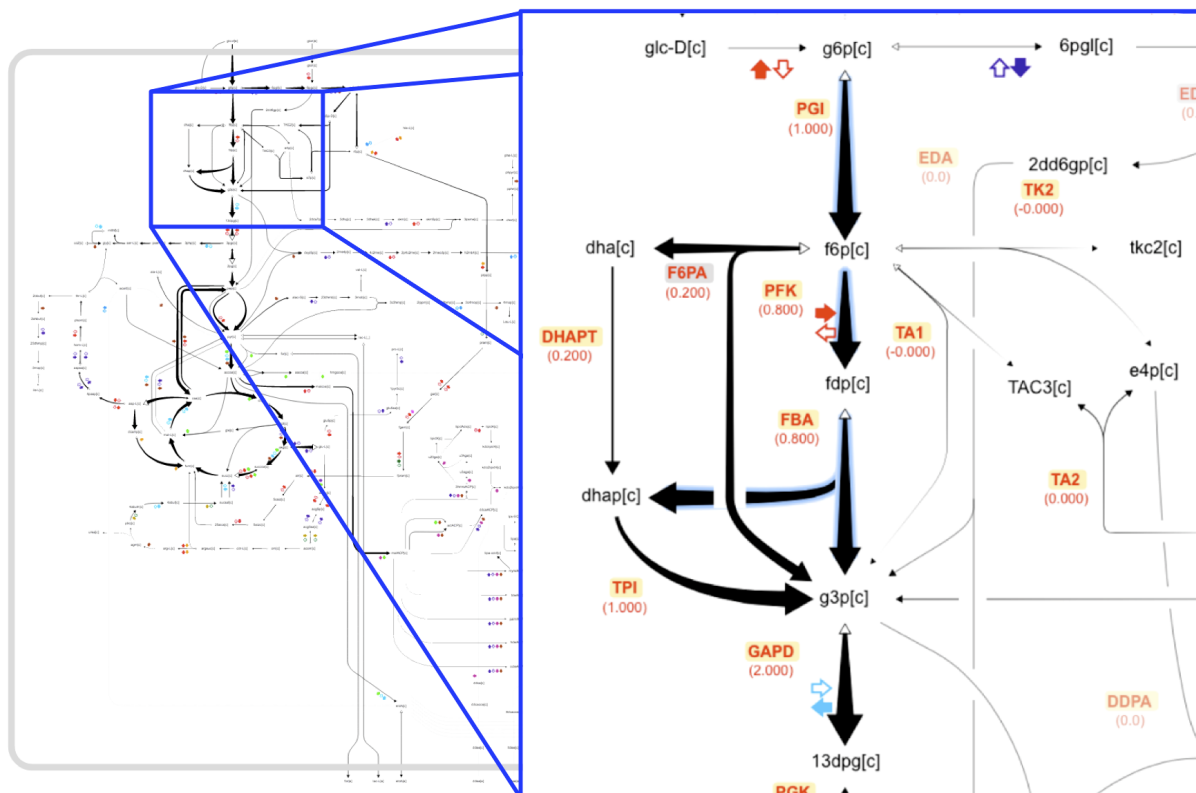


Figure 6. Arrowland is an online tool that presents ‘omics data in an intuitive and interactive fashion. The user can zoom in or out and obtain more information or a global picture. ‘Omics data is presented with flux data (proportional to flux thickness) through overlays (blue shadow for transcript data).

8. CONCLUDING REMARKS

JBEI has produced a set of tools aimed at easing the process of engineering biological systems for biofuel production. However, many of these tools can be generally used for a variety of biological engineering goals and are expected to have an impact in bioengineering beyond JBEI. In this report, we have described ICE as a digital and physical repository of microorganisms, sequences, parts, and seeds. We have also shown how DeviceEditor, j5, and DIVA constitute a powerful tool set for visually designing DNA constructs and automatically generate optimized protocols to build them. We have seen how the EDD constitutes a repository of standardized experimental data, of use to improve predictability and repeatability. In fact, these data can be leveraged through the jQMM library to predict biological behavior and produce actionable items for yield increase, one of the key enhancers of commercialization potential. Finally, we have shown how Arrowland can be used to provide an interactive, multi-scale visualization of ‘omics data. We hope that the tools showcased here and developed at JBEI will form the seed of a more robust, effective and predictive bioengineering process.

Q3 Report Authors: Hector Garcia Martin and Nathan J. Hillson.

JBEI Fuel Synthesis Leadership: Harry R. Beller, Taek Soon Lee, Hector Garcia Martin, Nathan J. Hillson, Jay D. Keasling, and Aindrila Mukhopadhyay.

REFERENCES

- [1] 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.
- [2] Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Roehner, N., Wilson, M. L., Adam, L., Anderson, J. C., Bartley, B. A., Beal, J., Chandran, D., Chen, J., Densmore, D., Endy, D., Grunberg, R., Hallinan, J., Hillson, N. J., Johnson, J. D., Kuchinsky, A., Lux, M., Misirli, G., Peccoud, J., Plahar, H. A., Sirin, E., Stan, G. B., Villalobos, A., Wipat, A., Gennari, J. H., Myers, C. J., and Sauro, H. M. (2014) The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology, *Nat Biotechnol* 32, 545-550.
- [3] Roehner, N., Beal, J., Clancy, K., Bartley, B., Misirli, G., Grunberg, R., Oberortner, E., Pocock, M., Bissell, M., Madsen, C., Nguyen, T., Zhang, M., Zhang, Z., Zundel, Z., Densmore, D., Gennari, J. H., Wipat, A., Sauro, H. M., and Myers, C. J. (2016) Sharing Structure and Function in Biological Design with SBOL 2.0, *ACS Synth Biol* 5, 498-506.
- [4] Quinn, J. Y., Cox, R. S., 3rd, Adler, A., Beal, J., Bhatia, S., Cai, Y., Chen, J., Clancy, K., Galdzicki, M., Hillson, N. J., Le Novere, N., Maheshwari, A. J., McLaughlin, J. A., Myers, C. J., P, U., Pocock, M., Rodriguez, C., Soldatova, L., Stan, G. B., Swainston, N., Wipat, A., and Sauro, H. M. (2015) SBOL Visual: A Graphical Language for Genetic Designs, *PLoS Biol* 13, e1002310.
- [5] Bhatia, S., and Densmore, D. (2013) Pigeon: a design visualizer for synthetic biology, *ACS Synth Biol* 2, 348-350.
- [6] Alonso-Gutierrez, J., Kim, E. M., Batth, T. S., Cho, N., Hu, Q., Chan, L. J., Petzold, C. J., Hillson, N. J., Adams, P. D., Keasling, J. D., Garcia Martin, H., and Lee, T. S. (2015) Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering, *Metab Eng* 28, 123-133.
- [7] Hillson, N. J., Plahar, H. A., Beal, J., and Prithviraj, R. (2016) Improving Synthetic Biology Communication: Recommended Practices for Visual Depiction and Digital Submission of Genetic Designs, *ACS Synth Biol* 5, 449-451.
- [8] Chen, J., Densmore, D., Ham, T. S., Keasling, J. D., and Hillson, N. J. (2012) DeviceEditor visual biological CAD canvas, *J Biol Eng* 6, 1.
- [9] Oberortner, E., and Densmore, D. (2015) Web-based software tool for constraint-based design specification of synthetic biological systems, *ACS Synth Biol* 4, 757-760.
- [10] Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D., and Bader, J. S. (2010) GeneDesign 3.0 is an updated synthetic biology toolkit, *Nucleic Acids Res* 38, 2603-2606.
- [11] Linshiz, G., Stawski, N., Goyal, G., Bi, C., Poust, S., Sharma, M., Mutalik, V., Keasling, J. D., and Hillson, N. J. (2014) PR-PR: cross-platform laboratory automation system, *ACS Synth Biol* 3, 515-524.
- [12] Hillson, N. J., Rosengarten, R. D., and Keasling, J. D. (2012) j5 DNA assembly design automation software, *ACS Synth Biol* 1, 14-21.
- [13] Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J. D., and Hillson, N. J. (2013) PaR-PaR laboratory automation platform, *ACS Synth Biol* 2, 216-222.

- [14] Gach, P. C., Shih, S. C., Sustarich, J., Keasling, J. D., Hillson, N. J., Adams, P. D., and Singh, A. K. (2016) A Droplet Microfluidic Platform for Automating Genetic Engineering, *ACS Synth Biol* 5, 426-433.
- [15] Linshiz, G., Jensen, E., Stawski, N., Bi, C., Elsbree, N., Jiao, H., Kim, J., Mathies, R., Keasling, J. D., and Hillson, N. J. (2016) End-to-end automated microfluidic platform for synthetic biology: from design to functional analysis, *J Biol Eng* 10, 3.
- [16] Shih, S. C., Goyal, G., Kim, P. W., Koutsoubelis, N., Keasling, J. D., Adams, P. D., Hillson, N. J., and Singh, A. K. (2015) A Versatile Microfluidic Device for Automating Synthetic Biology, *ACS Synth Biol* 4, 1151-1164.
- [17] Hillson, N. J. (2012) Scar-Less Multi-Part DNA Assembly Design Automation, U.S. Patent 9,361,427. Issued June 7, 2016.
- [18] Martin, H. G., Kumar, V. S., Weaver, D., Ghosh, A., Chubukov, V., Mukhopadhyay, A., Arkin, A., and Keasling, J. D. (2015) A Method to Constrain Genome-Scale Models with ¹³C Labeling Data, *PLoS Comput Biol* 11, e1004363.
- [19] Russo, E. (2003) The birth of biotechnology, *Nature* 421, 456-457.
- [20] Tang, N., S. Y. Ma, and J. D. Tian. (2013) New tools for cost-effective DNA synthesis, *Synthetic biology: tools and applications* 3-21.
- [21] Gardner, T. S. (2013) Synthetic biology: from hype to impact, *Trends Biotechnol* 31, 123-125.
- [22] Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martín, H. G. (2016) Synthetic and systems biology for microbial production of commodity chemicals, *Npj Systems Biology And Applications* 2, 16009.
- [23] Hyduke, D. R., Lewis, N. E., and Palsson, B. O. (2013) Analysis of omics data with genome-scale models of metabolism, *Mol Biosyst* 9, 167-174.
- [24] Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Jr., Assad-Garcia, N., Glass, J. I., and Covert, M. W. (2012) A whole-cell computational model predicts phenotype from genotype, *Cell* 150, 389-401.
- [25] Nelli, F. (2015) Machine Learning with scikit-learn, In *Python Data Analytics: Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language*, pp 237-264, Apress, Berkeley, CA.
- [26] Davidsohn, N., Beal, J., Kiani, S., Adler, A., Yaman, F., Li, Y., Xie, Z., and Weiss, R. (2015) Accurate predictions of genetic circuit behavior from part characterization and modular composition, *ACS Synth Biol* 4, 673-681.