Aledb 1.0: A database of mutations from adaptive laboratory evolution experimentation

Phaneuf, Patrick V.; Gosting, Dennis; Palsson, Bernhard O.; Feist, Adam M.

Published in:
Nucleic acids research

Link to article, DOI:
10.1093/nar/gky983

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

D1164–D1171 *Nucleic Acids Research, 2019, Vol. 47, Database issue* *Published online 24 October 2018*
doi: 10.1093/nar/gky983

ALEdb 1.0: a database of mutations from adaptive laboratory evolution experimentation

Patrick V. Phaneuf [1], Dennis Gosting[2], Bernhard O. Palsson [1,2,3,4] and Adam M. Feist[2,4,*]

[1]Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA 92093, USA, [2]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA, [3]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA and [4]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, 2800 Kgs. Lyngby, Denmark

Received June 12, 2018; Revised September 20, 2018; Editorial Decision September 28, 2018; Accepted October 08, 2018

ABSTRACT

Adaptive Laboratory Evolution (ALE) has emerged as an experimental approach to discover causal mutations that confer desired phenotypic functions. ALE not only represents a controllable experimental approach to systematically discover genotype-phenotype relationships, but also allows for the revelation of the series of genetic alterations required to acquire the new phenotype. Numerous ALE studies have been published, providing a strong impetus for developing databases to warehouse experimental evolution information and make it retrievable for large-scale analysis. Here, the first step towards establishing this resource is presented: ALEdb (http://aledb.org). This initial release contains over 11 000 mutations that have been discovered from eleven ALE publications. ALEdb (i) is a web-based platform that comprehensively reports on ALE acquired mutations and their conditions, (ii) reports key mutations using previously established trends, (iii) enables a search-driven workflow to enhance user mutation functional analysis through mutation cross-reference, (iv) allows exporting of mutation query results for custom analysis, (v) includes a bibliome describing the databased experiment publications and (vi) contains experimental evolution mutations from multiple model organisms. Thus, ALEdb is an informative platform which will become increasingly revealing as the number of reported ALE experiments and identified mutations continue to expand.

INTRODUCTION

Adaptive Laboratory Evolution (ALE) is a tool for the study of microbial adaptation. The typical execution of an ALE experiment involves cultivating a population of microorganisms in defined conditions (i.e., in a laboratory) for a period of time that enables the selection of improved phenotypes. Standard model organisms, such as *Escherichia coli*, have proven well suited for ALE studies due to their ease of cultivation and storage, fast reproduction, well known genomes, and clear traceability of mutational events (1). With the advent of accessible whole genome resequencing, associations can be made between selected phenotypes and genotypic mutations (2).

Beginning with a starting strain, an ALE experiment can be executed by serially passing a selected culture to a fresh flask of media (Figure 1A), enabling the strain passed to continue acquiring mutations under the experimental conditions without dilution of resources. Strains propagated during ALEs are assumed to be those that outcompeted their competition due to adaptive mutations. Additional methods to perform experimental evolutions have been reviewed (2,3). Whole genome comparative sequencing, or resequencing, is used to identify mutations within evolved strains relative to the evolution's starting strain (Figure 1B). ALE experiments can additionally involve replicate evolutions: identical evolutions that are often executed in parallel. Replicate ALEs can reveal the dynamics of adaptation by enabling research into converging genotypes within an experiment (4).

ALE methods have become important scientific tools in the study of evolutionary phenomena and have contributed to research in basic discovery and applied fields. Evolutionary biologists seek to examine the dynamics and repeatability of evolution and to better understand the relationship between genotypic and phenotypic changes (5). ALE methods, along with the plummeting cost of sequencing, have greatly enabled their efforts, resulting in a variety of insights into adaptive evolution. ALE has often demonstrated that (i) increases in fitness diminish with each new adaptive mutation (6), (ii) genotypic convergence through mutations can occur on the level of functional complexes (7) and (iii) in-

*To whom correspondence should be addressed. Tel: +1 858 534 9592; Email: afeist@ucsd.edu

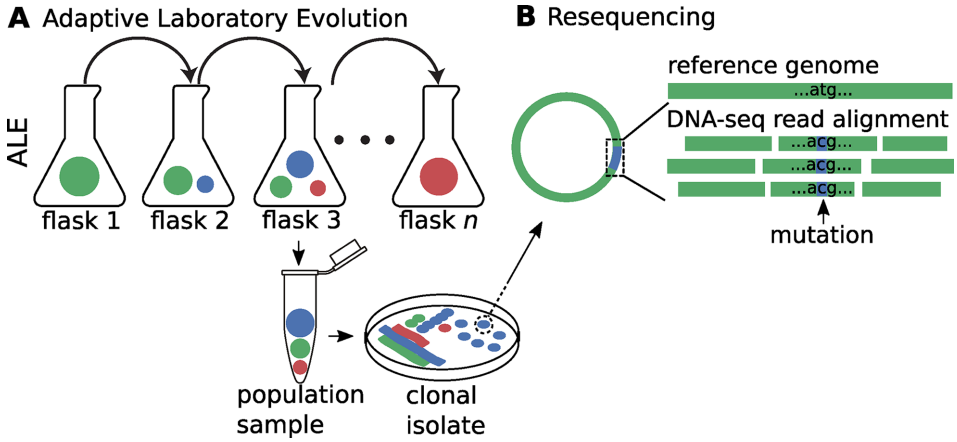


Figure 1. (**A**) An illustration of an ALE experiment where both a clonal and population sample are isolated from an intermediate (i.e., midpoint) flask. The Petri dish represents the streaking methodology for isolating a clonal colony from a population. (**B**) An illustration of how the resequencing process leverages a reference genome sequence and DNA-seq reads to identify mutations in an ALE sample.

teractions between mutations may cause nonlinear fitness effects (8).

ALE methods have also been leveraged in applications of synthetic biology to engineer microbes for commodity, industrial, and biopharmaceutical chemical synthesis (2). Comprehensive whole genome rational design is rarely achievable due to the complexity of biological systems (4,9). The inability to provide for comprehensive solutions in genome engineering can result in strains which cannot maintain homeostasis, such as strains which cannot tolerate the concentrations of products they were designed to produce. ALE has been used to produce adaptive mutations that provide solutions for the gaps left by current rational genome engineering methods (10). ALE can therefore complement rational genome engineering in the work to provide for a comprehensive whole genome solution to an application (2,9).

Accurately interpreting the results of an ALE requires the identification of causal mutations within the given observed adaptations. Identifying causal mutations requires a clear understanding of the mechanistic effects of mutations on cellular components and systems. Due to the complexity of cellular systems, interpreting the effects of mutations has proven to be a primary challenge in ALE (4,9). A common approach to mutation functional analysis is a literature search on the mutation target (e.g. a given annotated ORF). Functional studies of genetic targets have traditionally served as primary resources for interpreting mutation effects, providing information on a sequence's biological function. Published ALE results can be used to enhance efforts to identify and understand new adaptive mutations. Researchers can work to understand their ALE mutations by considering published adaptive mutations in conditions similar to their own ALEs.

A review of ALE methods (2) lists 34 separate ALE studies. Each study reports on novel combinations of selection conditions and the resulting microbial adaptive strategies. Large scale analysis of ALE results from such consolidation efforts could be a powerful tool for identifying and understanding novel adaptive mutations.

A web platform named *ALEdb* (aledb.org) has been created to meet the need for accessible consolidated ALE mutations, conditions, and publication reporting. ALEdb additionally includes features to search for specific mutations, report key mutations, and export mutation data for custom analysis. With these features, ALEdb works to fill the gap in the field of experimental evolution for an accessible resource of consolidated experimental evolution mutations.

RESULTS

A web platform to accelerate the work of ALE data to knowledge

The need for consolidated and accessible ALE experiment reporting has resulted in the generation of the web platform *ALEdb* (aledb.org). Eleven published ALE experiments, with a total of four distinct strains, 528 samples, and 11 792 observed mutations, serve as an initial data set (Figure 2).

Experimental evolution studies explore the solution space of a genome optimization problem through mutational events. This element of exploration has lead to a rich diversity of published ALE experimental conditions (2). Those experimental conditions currently represented in ALEdb are genetic perturbations (11), stress inducing environments (12), different carbon sources (13–15), and evolution duration (5). Strains can often adapt to these conditions with a variety of different evolutionary strategies, leading to different beneficial mutations. This leads to a diversity in the mutations across ALE experiments. This rich variety of databased conditions and mutations have made ALEdb an attractive research resource, and further implementation has now made this information accessible through the web.

ALEdb's feature set was developed in response to the challenge of accessible ALE mutation reporting for an ALE experiment pipeline (16). ALEdb's features enable intuitive navigation through consolidated ALE experiment data by providing two categories of features: those that describe individual ALE experiments, and those that describe all consolidated experiment data. To describe individual ALE experiments, ALEdb generates reports that detail the ALE

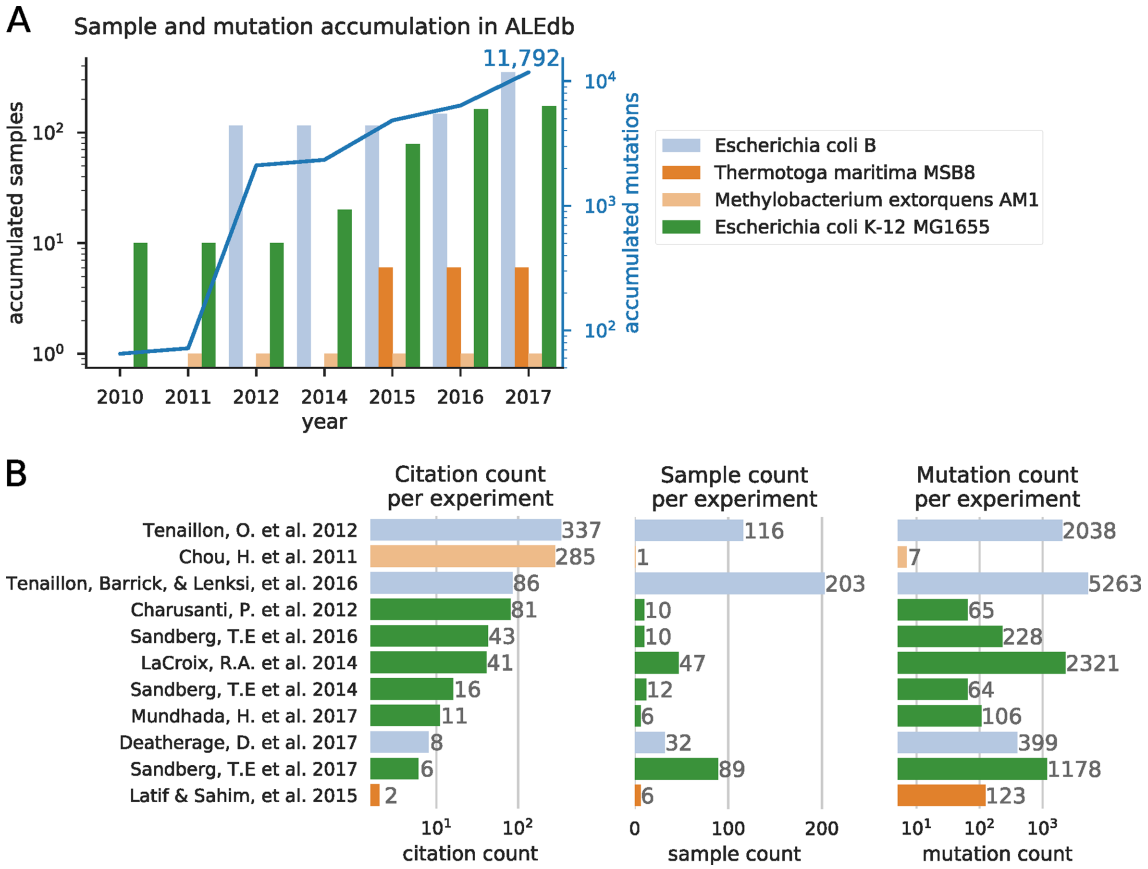


Figure 2. (**A**) A plot of the accumulated sequenced samples and mutations in ALEdb by publication year. (**B**) Each publication's sample and mutation contribution to ALEdb along with their citation count at the time of ALEdb's initial release. Citation counts were acquired from Google Scholar (scholar.google.com).

mutation lineages, key mutations, and experimental conditions per ALE sample. Each experiment is additionally described by a summary page which references the data's published work, includes summary visualizations, and annotates important details about the experiment dataset being hosted. To describe all consolidated ALE experiment data, ALEdb provides a mutation search feature, the ability to export the mutation data from one or more ALE experiments as spreadsheets, and an itemization of all publications that describe the databased mutations (Figure 3). ALEdb thus provides for a need in the experimental evolution community: a platform to search and explore consolidated experimental evolution mutation data.

Mutation functional analysis is a major challenge in experimental evolution. Besides systems biology modeling methods, this task often involves searching the literature for similar results. The ALE mutations, conditions, and publications being consolidated into ALEdb can be leveraged in this work. ALEdb can enhance a user's mutation functional analysis by using a search-driven workflow to report if mutations similar to theirs have occurred in published ALE experiments. Through ALEdb's *Search* feature, users can query for mutations of interest using multiple descriptive parameters and become aware of any databased ALE experiments that manifest similar mutations. Knowing these experiments, users can review the conditions and key mutation reports which characterize their results and refer to their associated publications through ALEdb's *Bibliome* page. These publications ultimately describe adaptive mutations and their functional analysis, which could be leveraged by users to better understand similar mutations in their own studies. ALEdb additionally includes the ability to *Export* mutation data for users interested in leveraging ALE data in applications beyond this platform (Figure 4). ALEdb's features are described in the following sections.

Mutation search and reporting

ALEdb implements a mutation *Search* to enable users to quickly find mutations of interest. Search returns a report of mutations for all databased samples according to the following mutation descriptors: gene, genome position range, mutation type, sequence change, protein change, and experiment.

Mutation search, along with most other mutation reporting mechanisms on ALEdb, present their results in the form of mutation tables (Figure 5A). Each ALE experiment can be described as a series of mutation sets relative to an ALE's starting strain. Columns represent an ALE sample and are described with the experiment name, ALE (A#), flask (F#), isolate (I#) and technical replicate (R#) value to serialize samples (Figure 5B). Rows describe the specific mutation

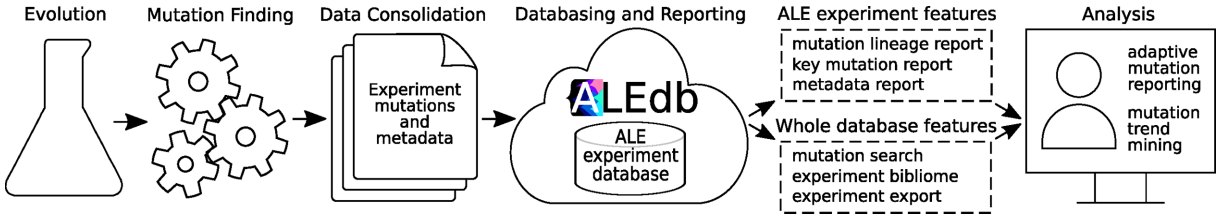


Figure 3. An illustration of the flow of ALE data from an experimental evolution to the generation of reports.

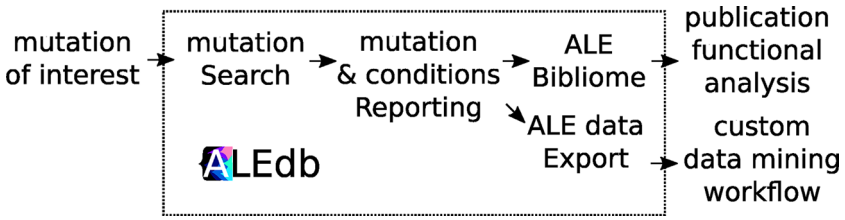
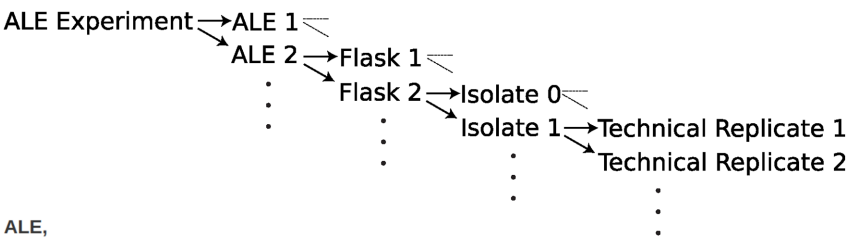


Figure 4. An illustration of the workflow for mutation functional analysis using ALEdb. Each step within the ALEdb group is the name of a user feature on the ALEdb platform.

A

Position	Mutation Type	Sequence Change	Gene	Details	GLU A4 F66 I1 R1	GLU A4 F149 I1 R1	GLU A4 F237 I1 R1	GLU A4 F403 I0 R1	GLU A4 F403 I1 R1
1,551,659	SNP	G→A	adhP	P69S (CCA→TCA)		1.00			
3,998,893	DEL	Δ5 bp	corA	coding (220-224/951 nt)		1.00	1.00	1.00	1.00
3,999,402	DEL	(GGC)2→1	corA	coding (729-731/951 nt)	1.00				
1,292,256	MOB	IS1 (–) +8 bp	hns, tdk	intergenic (-110/-488)		1.00	1.00	1.00	1.00
3,203,742	SNP	G→A	ttdA	V11V (GTG→GTA)				0.21	

B



C

ALE, Flask, Isolate, Technical Replicate	Clonal or Population	Species	Strain	Additional Strain Details	Media	Substrate	Temperature
A4 F403 I0 R1	population	E. Coli	511145	BOP27	M9	Glucose(2g/L), NH4Cl(1g/L), KH2PO4(3g/L), O2	37.0
A4 F403 I1 R1	clonal	E. Coli	511145	BOP27	M9	Glucose(2g/L), NH4Cl(1g/L), KH2PO4(3g/L), O2	37.0

Figure 5. (**A**) An example mutation lineage report where samples are represented as columns, ordered from left to right as earliest to latest in an ALE. (**B**) An illustration of the information annotated by the sample column labels within mutation tables. In the case of the first sample column in this figure's mutation table, the label describes mutations from the first technical replicate, from the first isolate, from the sixty-sixth flask, from the fourth ALE, of the GLU experiment. (**C**) An ALE experiment metadata report.

details that manifested within the samples, and values contained within cells represent the allele frequency. Ordering sample columns from the earliest to latest sample in an ALE serves to render intuitive visualizations of temporal mutation trends. This format enables researchers to intuitively identify important mutational patterns, such as the fixed mutations within the *corA* gene and *hns/tdk* intergenic region (Figure 5A). Population samples will always be described by an isolate number (I#) of 0 and is the only sample type to carry allele frequencies less than 1.0. The information describing each mutation is generated by the mutation finding stage (Figure 2) and details a mutation's type, genetic target, and potential product effects. Mutation tables therefore describe the lineage of an ALE's final sample, or endpoint, according to the mutations that manifest during an evolution.

Researchers investigating ALE experiments require reporting that enables them to quickly understand which mutations are likely causal for adaptations; the mutation tables built by ALEdb are designed to meet this need. Among the many mutations that manifest within an ALE experiment, mutation rows that describe multiple alleles of a gene will cluster together according to their positions on the genome. This is illustrated with the *corA* mutation within Figure 5A. Due to the chronological sorting of the sample columns per ALE, a mutation that fixes across samples will manifest as an unbroken sequence of cells in a mutation row annotated with an allele frequency. This is illustrated with both the *hns/tdk* and *corA* mutations in Figure 5A. These two patterns are obvious to an observer and serve well to describe the adaptive mutational trends in ALE experiments.

ALE experiment mutations cannot be completely understood without considering the experiment's conditions. ALEdb includes reports that describe an experiment sample's strain, substrate, and environment (Figure 5C). These experiment *Metadata* reports can additionally be exported as spreadsheets for analysis external to ALEdb.

Consolidated ALE knowledge

A key component in the utility of ALEdb is the per experiment knowledge built from the databased mutations. The *Bibliome* feature itemizes the publications that studied the ALE mutations databased within ALEdb. Users can leverage the mutation functional analysis within these publications toward understanding any similar mutations in their experimental evolutions.

ALE experiment mutation export

ALEdb implements an *Export* feature to give users the freedom to perform any analysis of interest on the hosted data. This feature enables users to extract one or more experiment mutation sets into comma separated value files. Users can then leverage custom analysis pipelines on the ALEdb's data towards generating novel results.

Automated ALE experiment key mutation reporting

ALEdb includes features that automate the reporting of known ALE adaptive mutation trends. These trends are termed *fixed* and *converged* key mutations, where each trend describes a unique pattern of mutations occurring within or across multiple ALEs in an experiment. These patterns have been used in published ALE studies to identify adaptive mutations (11–14). The manual consolidation of adaptive mutation evidence can be prone to human error, inconsistent between researchers, and time consuming. The automation of these common analyses contributes to more consistent analysis and more accurate results.

A *fixed* mutation is one in which first manifests in any ALE sample other than the endpoint, and is propagated to all following samples in the ALE. The propagation of a mutation from their emergence to an ALE's endpoint may describe the selection of a mutation due to its fitness benefits (13). This analysis is only possible if an ALE experiment includes midpoint samples, providing the possibility of more than one data point per ALE mutation. The identification of *fixed* mutations is accomplished by organizing mutations according to the ALE's sample chronology and identifying mutations that emerge in a midpoint and manifest in all following samples of the same ALE (Figure 6A). ALEdb's *fixed* mutation reporting automates this analysis and reports results in the format described in Figure 5A.

A *converged* mutation is one in which manifests in a genetic region seen to be mutated in multiple replicate ALEs (Figure 6B). This phenomenon describes evidence of a potential common adaptive trajectory between microbes exposed to the same conditions and has been leveraged in ALE analysis methods to more quickly identify mutations causal for adaptive phenotypes (13). ALEdb's *converged* mutation reporting automates this analysis and reports results in the format described in Figure 5A.

Design and implementation

ALEdb is implemented and deployed using a standard web application technology stack and a combination of user interface technologies. ALEdb's server-side hosts a MySQL database (https://www.mysql.com/), uses the Python based Django web framework (https://www.djangoproject.com/), and serves the content using Gunicorn (http://gunicorn.org/) and Nginx (https://www.nginx.com/). ALEdb implements its user-interface with HTML, CSS, and JavaScript, along with a combination of Javascript libraries including Bootstrap (https://getbootstrap.com/), jQuery (https://jquery.com/), DataTables (https://datatables.net/), mutation-needle-plot (http://dx.doi.org/10.5281/zenodo.14561) and D3 (https://d3js.org/).

ALEdb mutation data usage case study

To demonstrate the potential for ALEdb as an experimental evolution mutation data resource, mutations from experiments databased within ALEdb were compared to mutation data published externally. *Escherichia coli* experimental evolution mutation data was gathered from three sources for comparison: (i) experiments executed by or with the Systems Biology Research Group (SBRG) (10–15,17), (ii) the Long Term Experimental Evolution (LTEE) (5) and (iii) the experimental evolution mutation set from the Dettman and

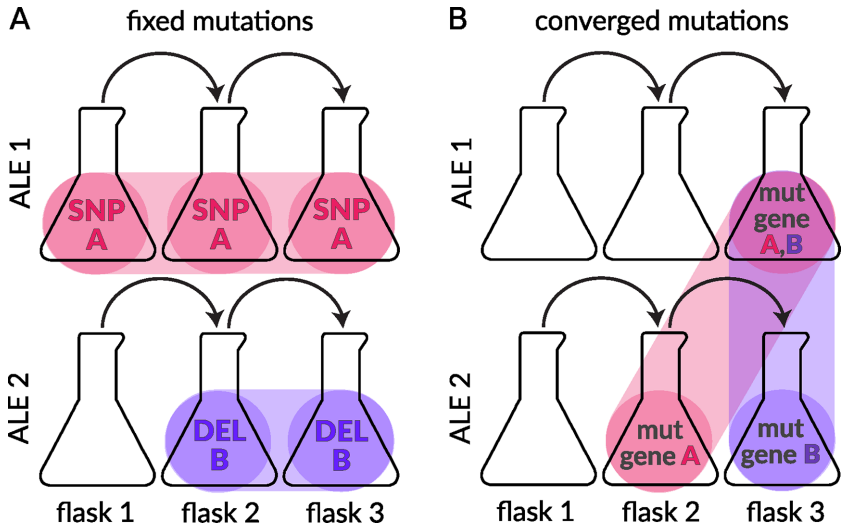


Figure 6. An illustration on the intuition utilized for fixed and converged mutation reports. (**A**) An illustration of fixed mutations. SNP A and DEL B occur in separate ALE replicates and persist through all subsequent flasks. (**B**) An illustration of converged mutations. Genetic targets A and B are seen to mutate across ALE replicates.

Kassen consolidation studies (KEE) (18). Both the SBRG and LTEE datasets were found within ALEdb while the KEE dataset was annotated within its publication. The following results were generated external to ALEdb, where ALEdb data was extracted using the *Export* feature. The LTEE hypermutator strains were not included in this case study.

Mutation type distributions were calculated across experiment sets to compare their endpoint proportions. The mutation finding software used by the SBRG and LTEE (19,20) describes mutations as single nucleotide polymorphisms (SNP), deletions (DEL), mobile element (MOB), and insertions (INS). The KEE set describes its mutations as SNPs or structural variants (SV), where SVs describe the combination of DEL, MOB and INS mutations. SNPs were the most common mutation across all datasets, with DELs, MOBs, and INS being the most common structural variants across SBRG and LTEE, in that order (Figure 7A). All experiment sets produced the same mutation type abundance and demonstrated similar distributions (Figure 7A, Supplementary Table S1).

The frequent manifestation of SNPs in experimental evolution endpoints has suggested that they play a role in adaptation. Previous work has investigated the selectivity of SNPs relative to the density of coding sequences within bacterial genomes (18). It was proposed that if coding SNPs were more causal for adaptation than non-coding SNPs, their endpoint proportions would be significantly larger than the proportion of coding nucleotides within the *E. coli* bacterial genome (i.e. larger than 0.86) (18). All of the coding SNP distributions for each of the three data sets overlap this average and do not provide evidence of being statistically different from each other (Figure 7B, Supplementary Table S2).

This case study serves as an example of the broad, multi-experiment analysis that can be readily performed with consolidated experimental evolution data. A goal of ALEdb is to provide for this need through consolidating and reporting on experimental evolution mutation data.

CONCLUSION

ALEdb works to serve the current need for a mutation database in the field of experimental evolution. It is a platform designed for the integration and reporting of ALE mutation datasets and currently integrates the mutation data and published materials of eleven published ALE experiments. Additionally, multiple features are implemented within ALEdb to enable intuitive navigation and analysis. Finally, the case study included in this work demonstrates the potential for ALEdb as a mutation data resource for broad, multi-experiment analysis.

ALEdb will continue to be developed to meet the needs of consolidating, reporting, and navigating ALE experiment data. This initial release of ALEdb considers previously generated mutation datasets. ALEdb will continue to grow with future inclusion of published ALE experiment results from currently contributing and new research organizations.

METHODS

Mutation finding pipeline

Mutation data currently hosted on ALEdb are generated by the *breseq* mutation finding pipeline (19,20). Being that these samples come from different projects, various version of breseq were used in their mutation data generation. The sequencing reads used to generate the mutation data were subjected to quality control through either *FastQC* (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the *FastX-toolkit* (http://hannonlab.cshl.edu/fastx_toolkit/) or *AfterQC* (21).

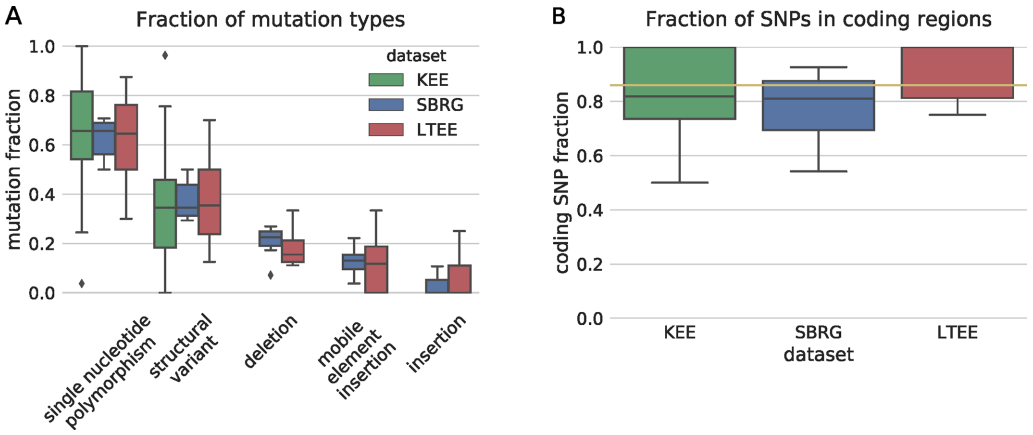


Figure 7. An example of multi-experiment analysis enabled by ALEdb's consolidated data. (**A**) Mutation type proportion distributions across endpoints for different sets of experiments. (**B**) Coding SNP proportion distributions across endpoints for different sets of experiments. The gold line represents the proportion of coding nucleotides within the E. coli bacterial genome (0.86) (18).

Experimental evolution normalization for case study

In comparing endpoint mutations between different experiments, strategies are necessary to normalize experiments which have different replicate evolution counts and lengths. For very long evolution, such as the Long-term Experimental Evolution, samples at 2000 generations were considered endpoints. To normalize for differing amounts of replicate evolutions between experiments, the average proportion of each mutation type of interest is found across an experiment's replicates to represent the experiment. Additionally, no samples containing hypermutator strains were included in the case study described in the final section.

DATA AVAILABILITY

ALEdb is freely available online at http://aledb.org and can be accessed with a JavaScript-enabled browser.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Lars Schöning, Kai Chen, and Pankaj Kumar Singh for their work on the infrastructure necessary to release ALEdb to the public. The authors would like to thank Dr. Ali Ebrahim, Dr. Ryan LaCroix, Kaiwen Zhang and Dylan Skola; their initial designs provided the foundation for ALEdb.

FUNDING

This work and the open access charge was supported by the Novo Nordisk Foundation through the Center for Biosustainability at the Technical University of Denmark [NNF10CC1016517] and by the National Institutes of Health's National Institute of Allergy and Infectious Diseases [U01AI124316].
Conflict of interest statement. None declared.

REFERENCES

1. Elena,S.F. and Lenski,R.E. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.*, **4**, 457–469.
2. Dragosits,M. and Mattanovich,D. (2013) Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Fact.*, **12**, 64.
3. Rabbers,I., van Heerden,J.H., Nordholt,N., Bachmann,H., Teusink,B. and Bruggeman,F.J. (2015) Metabolism at evolutionary optimal States. *Metabolites*, **5**, 311–343.
4. Palsson,B. (2011) Adaptive laboratory evolution this tool for studying molecular mechanisms and evolutionary dynamics follows bacteria as they adapt to environmental stress. *Microbe*, **6**, 69–74.
5. Tenaillon,O., Barrick,J.E., Ribeck,N., Deatherage,D.E., Blanchard,J.L., Dasgupta,A., Wu,G.C., Wielgoss,S., Cruveiller,S., Médigue,C. *et al.* (2016) Tempo and mode of genome evolution in a 50, 000-generation experiment. *Nature*, **536**, 165–170.
6. Barrick,J.E. and Lenski,R.E. (2009) Genome-wide mutational diversity in an evolving population of Escherichia coli. *Cold Spring Harb. Symp. Quant. Biol.*, **74**, 119–129.
7. Tenaillon,O., Rodríguez-Verdugo,A., Gaut,R.L., McDonald,P., Bennett,A.F., Long,A.D. and Gaut,B.S. (2012) The molecular diversity of adaptive convergence. *Science*, **335**, 457–461.
8. Chou,H.-H., Chiu,H.-C., Delaney,N.F., Segrè,D. and Marx,C.J. (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*, **332**, 1190–1192.
9. Williams,T.C., Pretorius,I.S. and Paulsen,I.T. (2016) Synthetic evolution of metabolic productivity using biosensors. *Trends Biotechnol.*, **34**, 371–381.
10. Mundhada,H., Seoane,J.M., Schneider,K., Koza,A., Christensen,H.B., Klein,T., Phaneuf,P.V., Herrgard,M., Feist,A.M. and Nielsen,A.T. (2017) Increased production of L-serine in Escherichia coli through adaptive laboratory evolution. *Metab. Eng.*, **39**, 141–150.
11. Charusanti,P., Conrad,T.M., Knight,E.M., Venkataraman,K., Fong,N.L., Xie,B., Gao,Y. and Palsson,B.Ø. (2010) Genetic basis of growth adaptation of escherichia coli after deletion of pgi, a major metabolic gene. *PLoS Genet.*, **6**, e1001186.
12. Sandberg,T.E., Pedersen,M., LaCroix,R.A., Ebrahim,A., Bonde,M., Herrgard,M.J., Palsson,B.O., Sommer,M. and Feist,A.M. (2014) Evolution of Escherichia coli to 42°C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Mol. Biol. Evol.*, **31**, 2647–2662.
13. LaCroix,R.A., Sandberg,T.E., O'Brien,E.J., Utrilla,J., Ebrahim,A., Guzman,G.I., Szubin,R., Palsson,B.O. and Feist,A.M. (2015) Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol.*, **81**, 17–30.

14. Sandberg,T.E., Long,C.P., Gonzalez,J.E., Feist,A.M., Antoniewicz,M.R. and Palsson,B.O. (2016) Evolution of E. coli on [U-13C]Glucose reveals a negligible isotopic influence on metabolism and physiology. *PLoS One*, **11**, e0151130.
15. Sandberg,T.E., Lloyd,C.J., Palsson,B.O. and Feist,A.M. (2017) Laboratory evolution to alternating substrate environments yields distinct phenotypic and genetic adaptive strategies. *Appl. Environ. Microbiol.*, **83**, e00410-17.
16. LaCroix,R.A., Palsson,B.O. and Feist,A.M. (2017) A model for designing adaptive laboratory evolution experiments. *Appl. Environ. Microbiol.*, **83**, e03115-16.
17. Long,C.P., Gonzalez,J.E., Feist,A.M., Palsson,B.O. and Antoniewicz,M.R. (2018) Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 222–227.
18. Dettman,J.R., Rodrigue,N., Melnyk,A.H., Wong,A., Bailey,S.F. and Kassen,R. (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol. Ecol.*, **21**, 2058–2077.
19. Deatherage,D.E. and Barrick,J.E. (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.*, **1151**, 165–188.
20. Barrick,J.E., Colburn,G., Deatherage,D.E., Traverse,C.C., Strand,M.D., Borges,J.J., Knoester,D.B., Reba,A. and Meyer,A.G. (2014) Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics*, **15**, 1039.
21. Chen,S., Huang,T., Zhou,Y., Han,Y., Xu,M. and Gu,J. (2017) AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*, **18**, 80.