OPEN ACCESS Freely available online

PLOS ONE

MediaDB: A Database of Microbial Growth Conditions in Defined Media

Matthew A. Richards[1,2], Victor Cassen[2], Benjamin D. Heavner[2], Nassim E. Ajami[1], Andrea Herrmann[1], Evangelos Simeonidis[2,3], Nathan D. Price[2]*

1 Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **2** Institute for Systems Biology, Seattle, Washington, United States of America, **3** Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Abstract

Isolating pure microbial cultures and cultivating them in the laboratory on defined media is used to more fully characterize the metabolism and physiology of organisms. However, identifying an appropriate growth medium for a novel isolate remains a challenging task. Even organisms with sequenced and annotated genomes can be difficult to grow, despite our ability to build genome-scale metabolic networks that connect genomic data with metabolic function. The scientific literature is scattered with information about defined growth media used successfully for cultivating a wide variety of organisms, but to date there exists no centralized repository to inform efforts to cultivate less characterized organisms by bridging the gap between genomic data and compound composition for growth media. Here we present MediaDB, a manually curated database of defined media that have been used for cultivating organisms with sequenced genomes, with an emphasis on organisms with metabolic network models. The database is accessible online, can be queried by keyword searches or downloaded in its entirety, and can generate exportable individual media formulation files. The data assembled in MediaDB facilitate comparative studies of organism growth media, serve as a starting point for formulating novel growth media, and contribute to formulating media for *in silico* investigation of metabolic networks. MediaDB is freely available for public use at https://mediadb.systemsbiology.net.

Citation: Richards MA, Cassen V, Heavner BD, Ajami NE, Herrmann A, et al. (2014) MediaDB: A Database of Microbial Growth Conditions in Defined Media. PLoS ONE 9(8): e103548. doi:10.1371/journal.pone.0103548

Editor: Stephen S. Fong, Virginia Commonwealth University, United States of America

Received April 23, 2014; **Accepted** June 30, 2014; **Published** August 6, 2014

Copyright: © 2014 Richards et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data excepting 16S rRNA sequences are within the paper and its Supporting Information Files. All 16S rRNA sequences are available from the Silva database (http://www.arb-silva.de/) with accession numbers as follows: A.baumannii = AB109775 A.baylyi = AF509820 A.ferrireducens = AF435948 A.nidulans = JQ435097 A.niger = AM270051 A.oryzae = D63698 A.terreus = GQ338244 B.amyloliquefaciens = AB201121 B.megaterium = AB244286 B.subtilis = AB042061 B.japonicum = AB070570 C.glabrata = AB094140 C.reinhardtii = AY665726 C.salexigens = DQ789389 C.acetobutylicum = AB595131 C.beijerinckii = AB020189 C.thermocellum = CP000568 C.glutamicum = BA000036 C.necator = AJ633674 Cyanothece sp. ATCC 51142 = CP000806 D.radiodurans = AE000513 E.coli = A14565 F.tularensis = AY243028 G.metallireducens = L07834 G.sulfurreducens = FW306010 H.influenzae = AB377129 H.salinarum = AM774415 H.thermophilus = AP011112 K.vulgare = CP002018 K.pneumoniae = AB641121 K.pastoris = CABH01000078 L.plantarum = AL935263 L.lactis = AB182585 L.major = GQ332361 M.succiniciproducens = AE016827 M.maripaludis = JQ405263 M.acetivorans = AE010299 M.barkeri = CP000099 M.tuberculosis = AL123456 N.pharaonis = CR936257 N.meningitidis = AF310340 N.crassa = AABX02000116 P.gingivalis = AB547660 P.aeruginosa = AE004091 P.putida = AE015451 R.etli = AM921623 S.cerevisiae = AB278124 S.enterica = AX201177 S.oneidensis = AB447987 S.aureus = AB634830 S.thermophilus = AGFN01000376 S.coelicolor = AL939114 Synechosystis sp. PCC 6803 = AP012495 T.maritima = M21774 V.vulnificus = AB680930 Y.pestis = ADSA01000714 Z.mobilis = AF117351.

Funding: This work was funded in part by a Camille Dreyfus Teacher-Scholar Award (http://www.dreyfus.org/), a Department of Energy Genomic Sciences Program (DE-FG02-10ER64999, http://genomicscience.energy.gov/), and National Institutes of Health Center for Systems Biology (2P50 GM076547, http://www.nih.gov/). Nathan D. Price received the funding from these sources. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: nprice@systemsbiology.org

Introduction

Genomic and high-throughput sequencing technologies enable the generation of large amounts of genetic information on microorganisms without the need to grow cultures in the lab. Armed with these technologies, we can automatically generate draft metabolic network reconstructions for organisms directly from genome annotations [1] and derive metabolic network models to simulate microbial growth *in silico*. These models can be improved through an iterative curation process between experimental and computational investigations [2]. To date, this iterative process has been most successfully advanced by partnering *in silico* reconstruction with *in vitro* characterization of isolates grown in defined laboratory media—an experimental approach that remains the most comprehensive method for characterizing microbial physiology [3–9]. Techniques for building metabolic network reconstructions from genomic data have progressed sufficiently to enable the application of *in silico* models for characterizing microbes that have not been cultivated *in vitro*.

Only 0.1–1% of the estimated number of microbial species have been isolated and successfully cultivated in a laboratory environment [5,6,8]. The collection of species we can currently culture spans only 30 of over 100 established phyla and mostly contains fast-growing organisms—organisms that are not the most prevalent species in the environment [7,9]. A range of novel techniques have been applied in efforts to culture less characterized microbes, such as using diffusion chambers to mimic environmental conditions [10–13], adding growth factors or signaling compounds

secreted from other organisms [14–17], diluting media nutrients to lower concentrations [18–24], increasing incubation time [19,20,23,25–28] and running high-throughput cultures [21,29–31]. These innovations have increased the diversity and number of culturable organisms, but the large number of factors that can affect *in vitro* growth still presents a challenge for isolating and culturing microbes from environmental samples.

Recently, computational modeling has been successfully applied to support culturing efforts. Several groups have used metabolic reconstructions, which are based on organism-specific genome sequence and biochemical knowledge, to assist in media design. Applications of these networks to media design have included both direct querying of the metabolic network to identify key metabolites for growth media design [32] and simulating growth on different substrates with a genome-scale metabolic model to predict media formulation [33]. Efforts that use a metabolic network model must define an *in silico* medium to enable calculations such as Flux Balance Analysis (FBA) [34–36]. The model and simulated medium then are iteratively refined until the network successfully predicts biomass production.

Thus, simulating growth of an uncultured organism with a metabolic model requires the definition of an *in silico* growth medium or a set of candidate media, which may then be validated *in vitro*. The definition of a growth medium *in silico* often begins in the same fashion as *in vitro* attempts: by starting with a medium that has supported simulated growth in models of organisms related to the desired isolate. However, this approach is complicated by the fragmentation of information in the literature. To overcome this obstacle, we have created MediaDB: a database of experimentally determined, chemically defined growth media conditions that aims to support efforts to leverage -omics data and modeling techniques for characterizing previously uncultured isolates. MediaDB is a manually curated database of defined media formulations for organisms with fully sequenced genomes, emphasizes organisms that have existing metabolic network models, and is the first publically available electronic resource that specifically brings together organisms with genomic data and their associated growth media. MediaDB will facilitate investigation of the relationship between microbial genomes and media composition, serving as both a central repository of data linking genome sequence to media compositions, and as a resource that facilitates model-supported design of cultivation media.

Database Construction and Content

All data in MediaDB were manually curated from existing primary literature sources. We conducted organism-by-organism literature searches using standard search engines—Google Scholar, PubMed, Web of Science—on the list of *in silico* organisms maintained by the Systems Biology Research Group at UCSD [37]. Our searches were aimed at finding experimentally-verified growth data on defined media for as many organisms with curated metabolic models as possible. The search results were curated manually and the media related information was extracted and formatted in the MediaDB schema, a MySQL database consisting of 12 tables and constructed around 6 main data tables: Organisms, Compounds, Media_Names, Biomass, Sources, and Growth_Data (Figure 1). The full schema is included as supporting information (Figure S1).

Organisms

The Organisms table includes fields for genus, species, and strain, a "type" designation that specifies the organism's kingdom classification, a Boolean value denoting whether the organism has been modeled *in silico*, and, if applicable, a link to the biomass composition for that organism. As shown in Table 1, MediaDB currently contains 208 unique Organisms objects spanning 57 species and 46 genera.

Bacteria make up the majority of organisms in the database, reflecting the distribution of species that have been cultured in the laboratory and the MediaDB's emphasis on organisms with existing *in silico* metabolic reconstructions. Such reconstructions exist for 39 of the 43 bacterial species and 51 of the 57 total species in the database. The database also includes many strains for model organisms; *Escherichia coli* and *Bacillus subtilis* contribute 54 and 16 bacterial strains, respectively, to the database.

Compounds

The Compounds table includes fields to describe a chemically-defined compound in terms of its common names, chemical formula, and identifiers that can be used to cross-reference with other databases (KEGG, BiGG, Seed, ChEBI and PubChem) [38–43]. We included identifiers from these databases to enable easier exchange of information between researchers, enhance compatibility with commonly-used resources, and ease development of automated computational analyses that use data in MediaDB. Of the 14,795 compounds contained in the database, 14,785 (99.9%) have identifiers from at least one other database.

Unlike the other tables in the MediaDB schema, the Compounds table was initially curated based on the KEGG database rather than from specific literature sources and was supplemented with manual entries from other databases as necessary. Its primary purpose is to describe the composition of other data types (Media_Names, Biomass).

Media_Names

The Media_Names table consists of fields specifying a media formulation with a descriptive name, a Boolean value indicating whether or not the particular media formulation was described as minimal in its source material, and a list of names and amounts of each compound that makes up that medium in units of millimolar (mM). Due to the many-to-many nature of relating compounds to different media compositions, the relationship between media formulations and compounds are contained within the Media_-Compounds table, but can be queried to find the compounds that make up a particular media formulation. MediaDB only contains chemically defined media formulations and does not include complex formulations, such as media that use yeast extract. The focus on chemically defined media was selected to facilitate computational simulation of growth conditions and to support efforts to cultivate uncultured organisms in the laboratory. MediaDB currently contains 461 different media formulations.

Biomass

The Biomass table consists of fields describing the compounds included in the biomass objective function used in FBA of metabolic network models to simulate exponential cell growth and contains organism genus and species, the list of compounds present in the biomass composition, and the stoichiometric coefficient of each compound in relation to one "unit" of biomass. Like the MediaDB description of media, biomass is also specified by the compounds that make up its composition, resulting in a many-to-many relationship. The Biomass_Compounds table contains the links between biomass compositions and compounds and can be queried to find the compounds that make up a particular biomass composition.

As detailed in Thiele *et al.* [2], the biomass composition is an important objective function for FBA of metabolic network

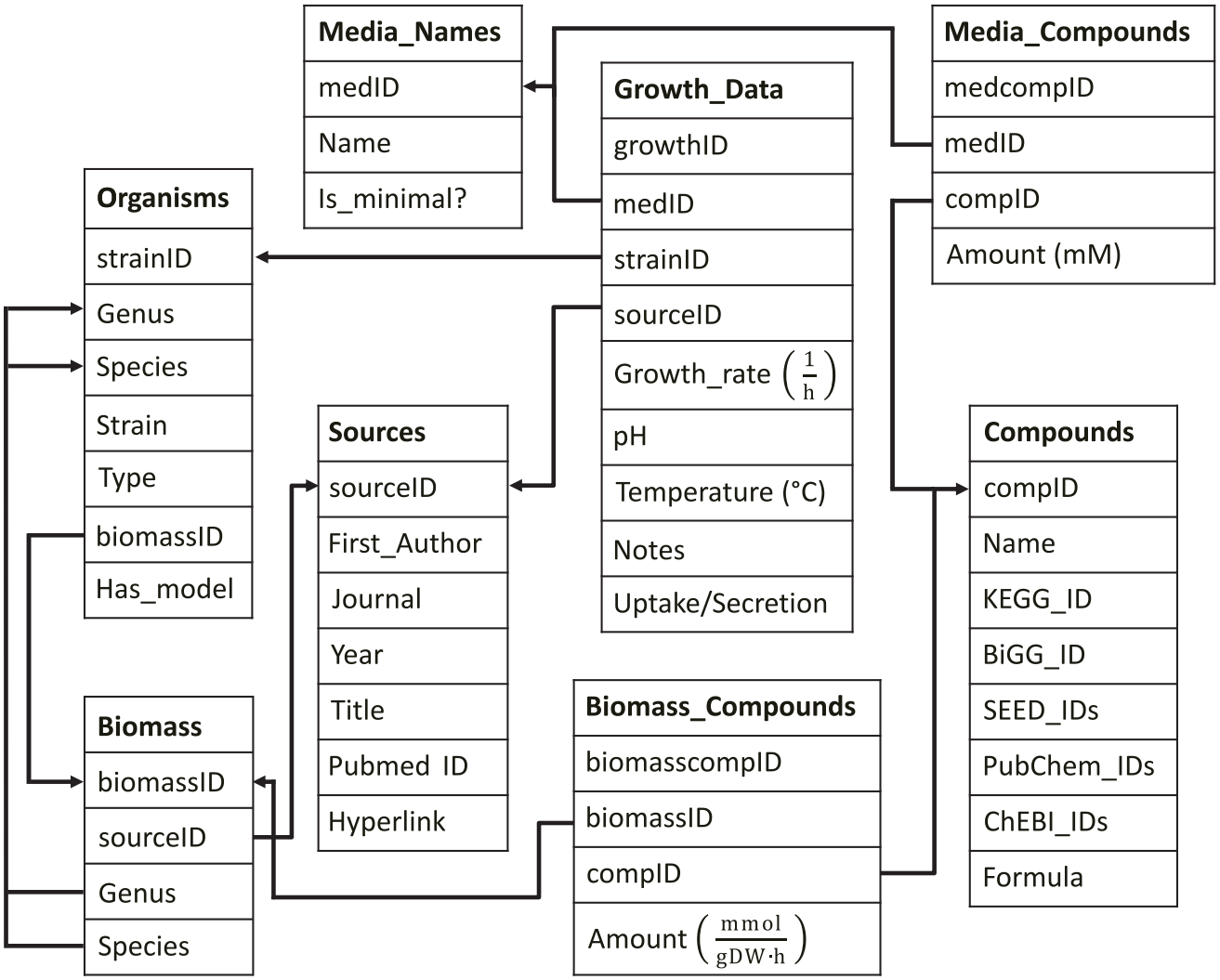

Figure 1. Simplified database schema. This graph shows the connections between the 6 main tables, Organisms, Compounds, Media_Names, Biomass, Sources, and Growth_Data. Also shown are Media_Compounds and Biomass_Compounds, linking tables that connect the Compounds table to the Media_Names and Biomass tables, respectively. Arrows indicate foreign key relationships, in which the head of the arrow points to the primary key being referenced. A full map of the MediaDB schema containing all tables and their connections can be found in Figure S1.
doi:10.1371/journal.pone.0103548.g001

models; however, it can also be difficult to experimentally determine detailed biomass composition for an organism. Thus, the biomass composition is a salient factor to consider in model construction and refinement, but we found few unique examples of this data type in existing literature sources. Instead, many models have defined the organism biomass composition by using or slightly modifying the biomass objective function from another model. We have included 4 different biomass compositions in MediaDB to provide a basis for users to construct biomass compositions for their own organisms by refining established ones.

Sources

The Sources table consists of fields describing a primary literature source (usually a book or a journal article) and is

Table 1. Taxonomy of organisms currently in MediaDB.

Kingdom	Genera	Species	Strains
Archaea	*4*	*5*	*14*
Bacteria	*36*	*43*	*172*
Eukaryote	*6*	*9*	*22*
TOTAL	**46**	**57**	**208**

doi:10.1371/journal.pone.0103548.t001

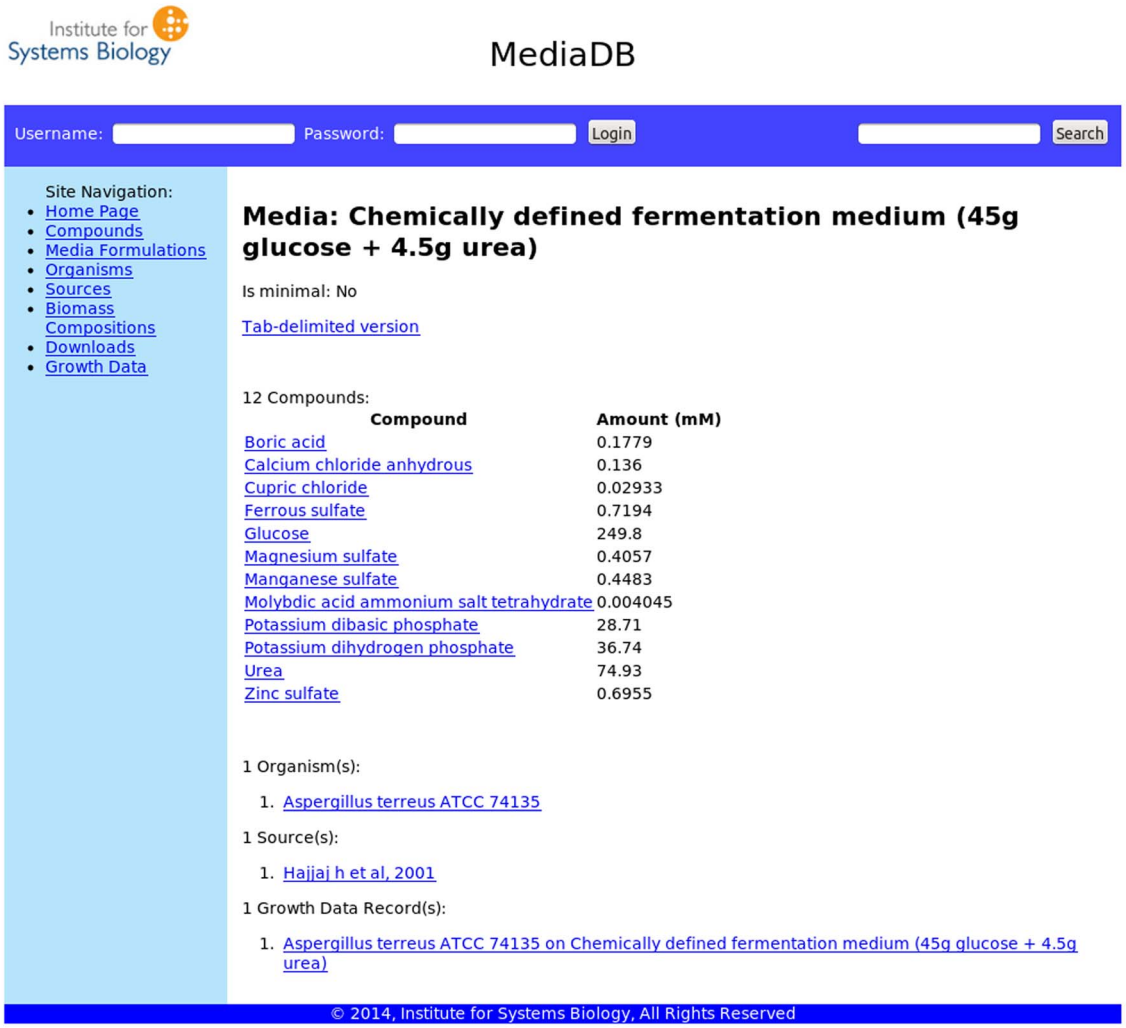

Figure 2. The MediaDB website. The database can be found at https://mediadb.systemsbiology.net. This page shows the composition of a media formulation and displays links to the organism, source, and growth record that use this medium. The "Site Navigation" panel lists the different tables that can be browsed manually and also the "Downloads" tab, where the user can export a copy of the entire MediaDB schema. The search field is at the top right of the page.
doi:10.1371/journal.pone.0103548.g002

Table 2. Highest frequency compounds in MediaDB.

Top 10 Compounds		**Top 5 Carbon Sources**	
CaCl2	86%	Glucose	47%
MgSO4	72%	Acetate	33%
KH2PO4	65%	Glycerol	19%
NaCl	65%	Pyruvate	12%
FeSO4	61%	Ethanol/Succinate	9%
ZnSO4	61%	**Top 5 Amino Acids**	
K2HPO4	58%	Cysteine	37%
NH4Cl	53%	Aspartate	33%
Biotin	51%	Arginine	33%
CuSO4	49%	Glutamate	32%
		Leucine	32%

Percentages reflect the fraction of species that contain each compound in at least one growth medium.
doi:10.1371/journal.pone.0103548.t002

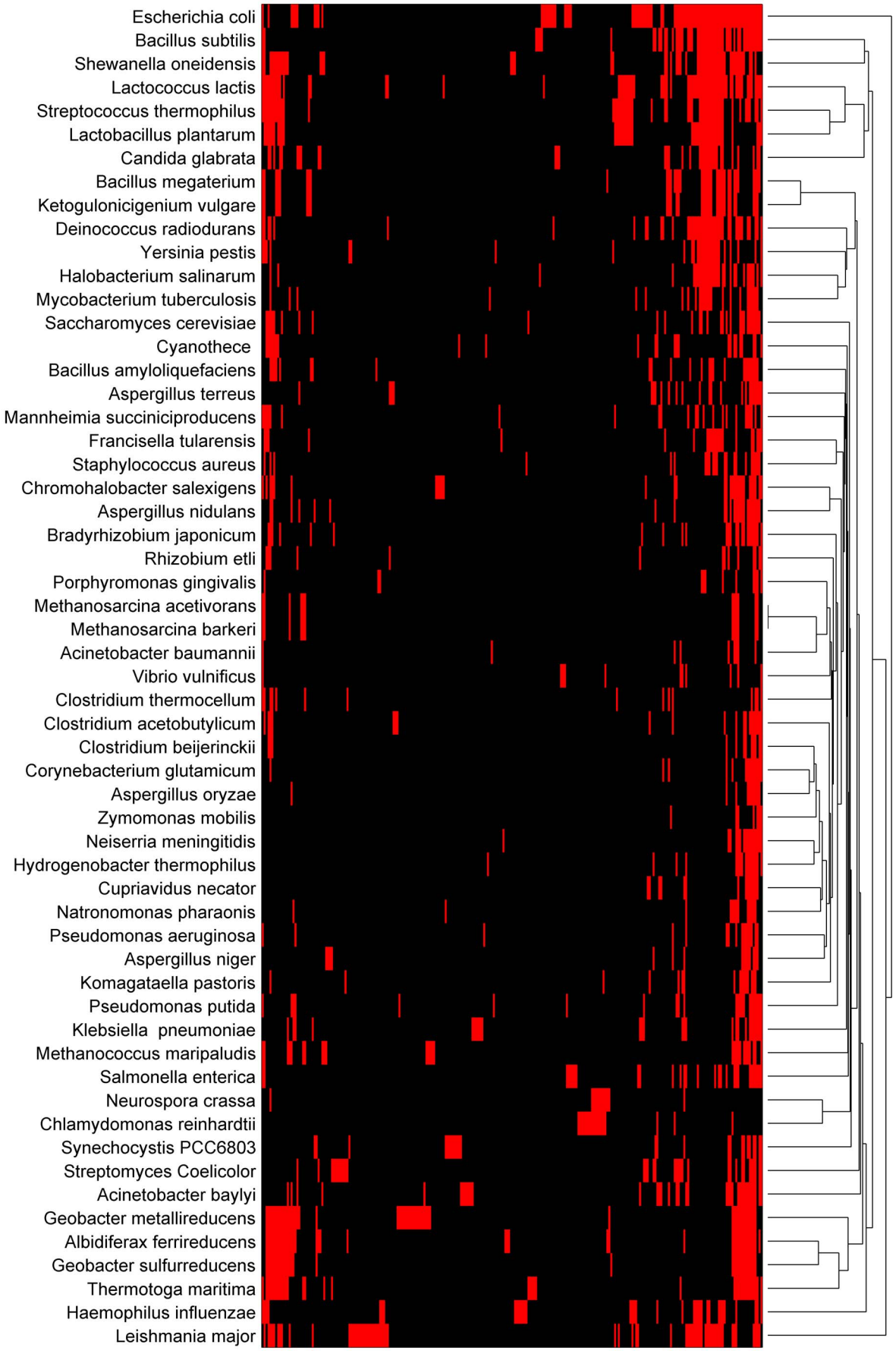
Escherichia coli
Bacillus subtilis
Shewanella oneidensis
Lactococcus lactis
Streptococcus thermophilus
Lactobacillus plantarum
Candida glabrata
Bacillus megaterium
Ketogulonicigenium vulgare
Deinococcus radiodurans
Yersinia pestis
Halobacterium salinarum
Mycobacterium tuberculosis
Saccharomyces cerevisiae
Cyanothece
Bacillus amyloliquefaciens
Aspergillus terreus
Mannheimia succiniciproducens
Francisella tularensis
Staphylococcus aureus
Chromohalobacter salexigens
Aspergillus nidulans
Bradyrhizobium japonicum
Rhizobium etli
Porphyromonas gingivalis
Methanosarcina acetivorans
Methanosarcina barkeri
Acinetobacter baumannii
Vibrio vulnificus
Clostridium thermocellum
Clostridium acetobutylicum
Clostridium beijerinckii
Corynebacterium glutamicum
Aspergillus oryzae
Zymomonas mobilis
Neiserria meningitidis
Hydrogenobacter thermophilus
Cupriavidus necator
Natronomonas pharaonis
Pseudomonas aeruginosa
Aspergillus niger
Komagataella pastoris
Pseudomonas putida
Klebsiella pneumoniae
Methanococcus maripaludis
Salmonella enterica
Neurospora crassa
Chlamydomonas reinhardtii
Synechocystis PCC6803
Streptomyces Coelicolor
Acinetobacter baylyi
Geobacter metallireducens
Albidiferax ferrireducens
Geobacter sulfurreducens
Thermotoga maritima
Haemophilus influenzae
Leishmania major

Figure 3. Heat map and dendrogram showing hierarchical clustering of species based on media compositions. Red bars indicate compounds that occur in at least one medium for that species. Black bars indicate compounds that do not appear in any media for that species. This figure was generated using the Statistics Toolbox in Matlab.
doi:10.1371/journal.pone.0103548.g003

specified using the first author's last name, the title of the work, the journal, the year of publication and, if applicable, the PubMed identifier and URL to the article. Sources are added to MediaDB if they report experimental laboratory growth of an organism in MediaDB in a medium in MediaDB. MediaDB currently contains 147 unique sources that directly link to any experimental growth media information they provided.

Growth Data

The Growth_Data table describes the combination of physical parameters reported by a literature source for *in vitro* growth of a specific organism. The Growth_Data table links the tables describing an organism, medium, and literature source, and adds information about temperature, pH, growth rate, product secretion rates, and nutrient uptake rates (whenever reported in the literature source). MediaDB currently contains 765 growth conditions.

In many instances, we found rate data associated with a particular growth condition in the form of an experimentally-measured growth rate (μ) measured in h^{-1}. We stored growth rates in the Growth_Data data field, thereby providing quantitative measures to assist in future metabolic model development. Some growth conditions were also reported with other growth-associated measurements: product secretion rates, medium compound uptake rates and product yields. Unlike growth rates, a growth condition could be associated with multiple measurements of secretion/uptake/yield; hence, we created the Secretion_Uptake table to house these rates and link them to their growth conditions. MediaDB currently contains 557 measured growth rates, 49 metabolite uptake rates, 22 product secretion rates, and 58 product yield coefficients.

Website Construction and Navigation

The MediaDB website (https://mediadb.systemsbiology.net/) provides a user-friendly interface for performing the two main functions of our database: data browsing and exporting.

Data browsing

Browsing allows the user to query MediaDB with provided data type categories, to manually search through information by navigating through the different data tables or to use keywords to search through the parameters that specify the growth condition entries (see Figure 2). The search function matches the given keyword to data entries in all tables and returns the results sorted by the table that contains the matched record.

Tables in the database are linked together on the webpage by cross-referencing to better display all pertinent information for each entry. For example, an entry in the Organisms table shows all of the related growth condition entries collected for that organism, including links to the literature source entries. Similarly, each media formulation entry links to entries for all the compounds present in that media formulation, all of the organisms reported to grow in that media formulation, and the literature source entries where the media formulation was reported. A Compounds entry displays links to all the media formulations in which the compound appears. A Source entry displays links to all the growth conditions reported in that source, as well as links to the online version of that source, when applicable.

Data export

Data can be exported from MediaDB in two different ways, allowing the user flexibility in deciding what information is important for their particular project. The most basic export, found under "Downloads" on the webpage, allows the user to download a copy of the entire MediaDB schema and all database entries to use independently of the website. This option allows the most flexibility in dealing with the data, but requires that the user be familiar enough with relational database management in MySQL to use the SQL file generated by this export.

The second export option is individual media formulation or biomass composition download, available on each media formulation or biomass composition entry page under "Tab-delimited version". This option generates a tab-delimited text file with a list of compounds and their concentrations in the chosen media formulation or biomass composition. The file also includes identifiers for the compounds in other databases. These identifiers facilitate cross-referencing of the various metabolite identifiers used in different *in silico* metabolic network models.

Database Utility

Statistics for compounds

Because the MediaDB schema provides links between organisms and the compounds in their growth media, it enables investigation of media components across organisms. For example, we compiled a list of every chemical compound that appears at least once in a growth medium for all 57 species in the database (see Table S1 for full results). Out of 260 unique compounds, the most commonly occurring compound across all species was calcium chloride ($CaCl_2$), a salt that appears in the growth media of 49 species (86% of all species in MediaDB), because it is often included in stock trace element/mineral solutions. Salts accounted for nine of the top ten most frequent compounds with the only exception being biotin, a vitamin that often appears in stock vitamin solutions and was present for 29 species (51%). Other components of media, such as the carbon source and amino acids, were less uniform across species; the most common carbon source and amino acid were glucose (47%) and cysteine (37%), respectively (a list of the most frequent compounds is shown in Table 2).

Our analysis also identified the *least* common compounds in media; 97 of the 260 compounds (37%) appeared in media for only one species and 139 (53%) appeared in media for one or two species only. These uncommon compounds generally fell into one of the following categories: 1) Trace metals included in stock solutions (e.g., nickel sulfate for *Shewanella oneidensis*); 2) Buffers for pH maintenance (e.g., ACES for *Mycobacterium tuberculosis*); 3) Antibiotics used to select for mutant strains (e.g., kanamycin for *Synechocystis PCC6803*); 4) Uncommon carbon sources (e.g., galactose for *Streptomyces coelicolor*); 5) Alternate vitamin forms (e.g., sodium pantothenate rather than calcium pantothenate for *Haemophilus influenzae*); 6) Compounds that fit niche organism metabolisms (e.g., 2-mercaptoethanesulfonate for *Methanococcus maripaludis*). Compounds in the final category were of particular interest, because they could be tied to unique portions of the known metabolism of the organism. For example, 2-mercaptoethanesulfonate (coenzyme M) only appears in media for the methanogen *M. maripaludis*, because it is a vital cofactor involved

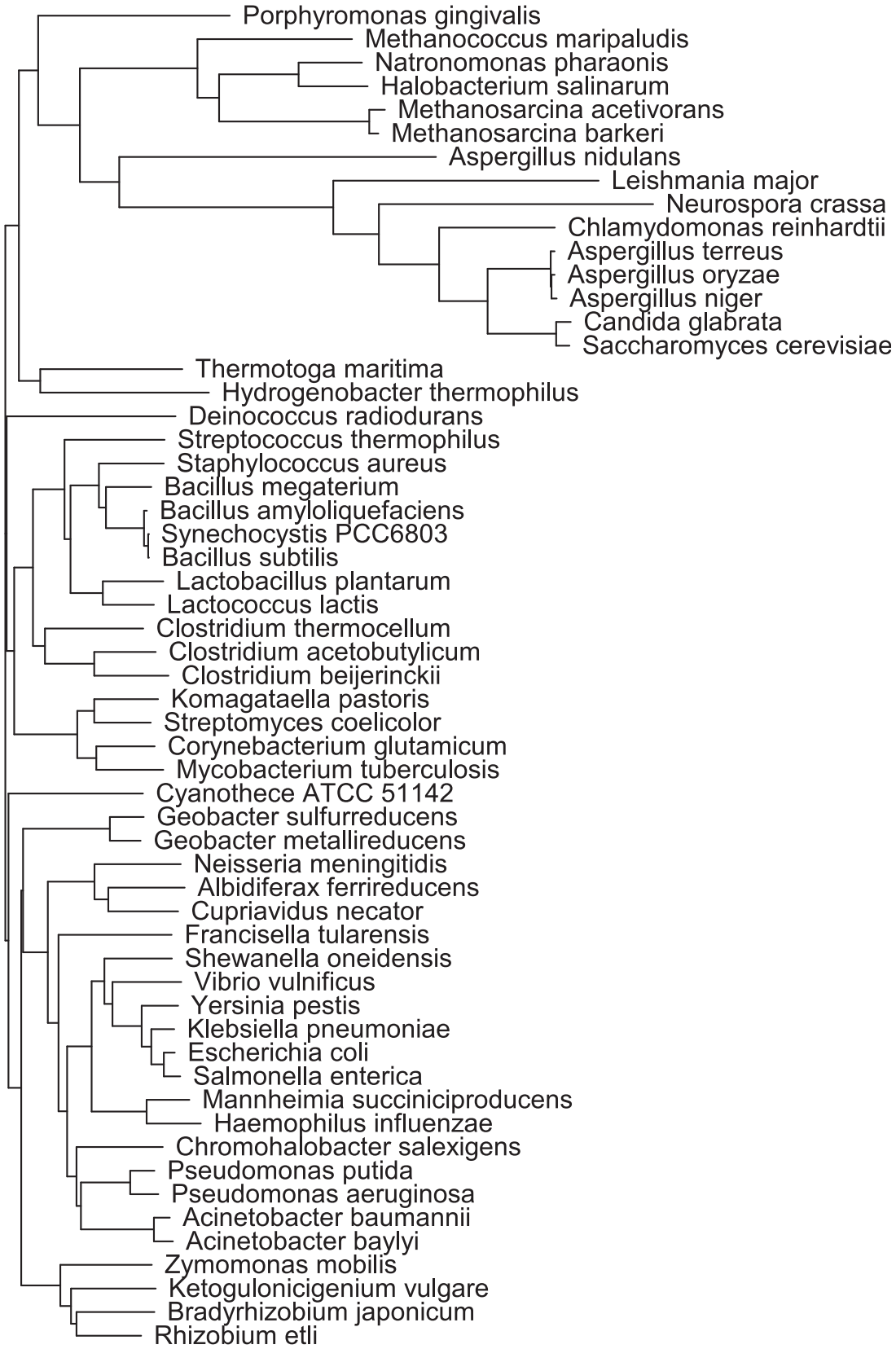

Figure 4. Phylogenetic tree of 16S rRNA sequences for species in MediaDB. Phylogeny was inferred from a CLUSTAL W alignment generated in the Biology Workbench using 16S rRNA sequences from the SILVA database.
doi:10.1371/journal.pone.0103548.g004

in methane production for that organism. As MediaDB grows, we expect that identifying such unusual compounds will play an increasingly useful role in media design.

Linking growth media to metabolism

MediaDB provides a framework for comparing the nutritional requirements of different organisms and currently includes information on a range of microbes, with a focus on organisms that have been modeled *in silico*. In order to demonstrate how MediaDB supports such comparative analysis, we compared media formulations for two organisms that have metabolic network models: *E. coli*, a model bacterium that has been grown with a wide range of compounds (81 different compounds), and *Methanosarcina acetivorans*, a model archaeon that has been grown using a smaller range of compounds (12 different compounds).

Seven compounds appeared in media formulations for both organisms: one carbon source (acetate) and six simple salts (NH_4Cl, $CaCl_2$, $MgCl_2$, KCl, KH_2PO_4, NaCl). The compounds unique to *E. coli* included multiple 5- or 6-carbon sugars (e.g., glucose, lactose, fructose, and succinate) and 19 of the 20 standard L-form amino acids (all except cysteine). The 5 compounds unique to *M. acetivorans* included methanol, a simple carbon source for methanogens that rarely appears in media for other organisms (fellow methanogen *Methanosarcina barkeri* and pathogen *Candida glabrata* are the only other species in MediaDB with media that include methanol). We also observed that, in contrast to the *E. coli* media data, cysteine was the only amino acid that appeared in growth media for *M. acetivorans*.

We expanded our comparison by using manually curated metabolic models for both *E. coli* [44] and *M. acetivorans* [45] to examine the differences found in media compounds. By examining reactions in the models, we observed that the model for *E. coli* included uptake pathways for many carbon sources that are absent in the *M. acetivorans* model, including all of the carbon sources reported in MediaDB. The *E. coli* model predicted that methanol could be produced during growth, but not consumed, whereas the *M. acetivorans* model predicted the ability to consume methanol for growth and methane production. The models also provided mechanistic justification for our media analysis that suggested differences in cysteine metabolism; the *M. acetivorans* model had the ability to both consume and secrete cysteine and the *E. coli* model predicted cysteine secretion, but not consumption. We extended this analysis by testing the models for growth on a range of experimental media from the database. We selected 11 media for *E. coli*—one for each carbon source—and the one medium for *M. acetivorans* in MediaDB, then simulated each model for growth on all 12 media (see File S1 for an example of this procedure). The *E. coli* model predicted growth on all 12 media, mirroring the organism's versatility to grow on many different carbon sources. The *M. acetivorans* model required modification to remove trace metals from the biomass objective function in order to predict growth on any medium. After the trace metals (which are not included in simulated *E. coli* media) were removed from the *M. acetivorans* model objective function, it accurately predicted growth on its own medium and on the *E. coli* medium with acetate as the carbon source, but not on any of the other media, reflecting the organism's inability to grow on complex carbon sources.

This case study illustrates the use of MediaDB as a tool for investigating the differences in nutritional requirements between organisms and as a source for *in silico* medium formulation. The differences between cultivation media for *E. coli* and *M. acetivorans* were identified using MediaDB and explained using the organisms' respective metabolic models, which include fundamental differences in carbon source and amino acid metabolism. In this example, the results of the comparisons between the media sources and metabolic models were quite parallel, as expected, because both models were manually constructed based on genomic information and information from the primary literature, including media formulation sources. In other cases, where there is disagreement between model simulation results and media information reported, MediaDB will support efforts to improve metabolic network reconstruction by providing information regarding experimentally determined media conditions.

Organism clustering by compound similarity

We used hierarchical clustering of pairwise Euclidean distance between binary vectors of compound inclusion in a medium (e.g., an entry is 1 if a given chemical is included in a medium, or 0 otherwise) to investigate the relationship between organisms in MediaDB based on published growth-supporting media. Figure 3 presents a heat map of chemical species in media, created from MediaDB data. The heat map shows bands of high-frequency compounds on the right side of the map and clusters of moderately frequent compounds on the left side; these compound groups are dominated by salts found in stock solutions and L-form amino acids, respectively. The overall sparsity of the heatmap reflects the fact that most compounds occur only once or twice across all species.

We compared this compound similarity tree (Figure 3) to a 16s rRNA phylogenetic tree constructed in the Biology Workbench [46–49] (Figure 4) and found that there was little overlap between genetic similarity and compound similarity. Aside from the two *Methanosarcina* species, which were grown in the same exact media, we observed few parallels between these two trees. Three species in the taxonomic order *Lactobacillales*—*Lactococcus lactis, Lactobacillus plantarum*, and *Streptococcus thermophilus*—clustered closely together in both trees, but the majority of organisms that formed tight clusters in one tree did not show the same closeness in the other tree. For example, the four *Aspergilli*—*A. nidulans, A. niger, A. oryzae*, and *A. terreus*—were close in terms of phylogenetic distance, but dissimilar with respect to their media compounds. On the other end of the spectrum, *Corynebacterium glutamicum*, *A. oryzae, Clostridium beijerinckii*, and *Zymomonas mobilis* show high compound similarity with one another, but are far apart phylogenetically. This observation could be an indication that phylogeny does not correlate to similarity in media formulations, but a more parsimonious explanation is that the data in MediaDB reflect the literature bias towards positive growth results. Due to this lack of negative growth results (i.e. information on what an organism *does not* grow on, which is typically omitted by researchers), we are unable to assert that any organism is incapable of growth in another's media based soley on comparisons of the collected data in MediaDB. This knowledge gap suggests a need for for futher experimental study of the relationship between phylogenetic distance and nutritional requirements for growth. Thus, information available in MediaDB describes whether a given medium has been reported to support a microbe's growth, and may be useful for generating hypotheses of possible media formulations for future experimental efforts. Our analysis also revealed clusters of organisms with high media composition similarity (Figure 3) that do not have a clear connection to observed biology. With further investigation, these similarities could reveal more complex biological relationships that do not fall under the obvious prisms of genetic or environmental

similarity. MediaDB will support such comparative studies as the resource continues to grow.

Future development

Community-contributed growth conditions. MediaDB currently contains 57 microbial species, but the scope of the fully-sequenced microbial world is much larger and continues to grow. We intend to expand the breadth of organisms and growth conditions in MediaDB by allowing users to submit their own experimentally verified, defined growth conditions. At this time, we encourage users to submit growth conditions for our review through direct contact with the authors (mediadb@systemsbiology.org), but expect to create an input form that encourages groups to add new data directly through the website.

Analysis tool development. We have demonstrated the potential for media-based comparative analysis using MediaDB with E. coli and M. acetivorans; however, we have designed MediaDB to support future development of additional tools to support research efforts. We have also made the entire database schema and its contents available for download to further facilitate tool development by MediaDB users. As such tools are developed in our group and others, we will integrate these tools into the website to assist users in their analyses.

Discussion and Conclusions

We present MediaDB, a manually curated database of defined media that have been used to cultivate organisms with sequenced genomes. Our database offers several important new capabilities for researchers through the following features: 1) brings together literature sources of experimentally verified media formulations into a centralized database; 2) contains chemically defined media, so that every compound can be linked to known metabolic pathways in metabolic network models, and so that every formulation is repeatable; 3) links with compound identifiers in existing databases for simple, repeatable and automatable cross-referencing with other sources; 4) focuses on organisms with existing *in silico* models, both encouraging researchers to use and improve such models and providing multiple media conditions to support the iterative development of *in silico* models; 5) serves as a set of organism-specific media conditions to help improve automated metabolic reconstruction methods by replacing more generic media formulations; 6) includes only species with fully-sequenced genomes to ensure that all media formulations can be tied back to genomic data; and 7) is a publically available resource that we expect will grow and increase in usage as growth conditions for more organisms are added. We anticipate that MediaDB will support the investigation of the relationship between organism growth media formulations and genomic information, and facilitate efforts to model microbial metabolism.

Availability and requirements

The MediaDB database is a publically accessible resource, available through the Institute of Systems Biology (ISB) website at https://mediadb.systemsbiology.net. The ISB infrastructure provides a stable server platform to allow for long term maintenance of MediaDB. To submit data for upload into MediaDB, or for general questions and information, please contact the authors at mediadb@systemsbiology.org.

Supporting Information

Figure S1 Full MediaDB schema. Dashed lines indicate foreign key relationships, oriented such that arrows point towards the referenced primary key. Each table is represented by a box headed by the table name and described by a list of column names and column types. This diagram was created using MySQL Workbench (www.mysql.com/products/workbench).
(PDF)

Table S1 Full compound frequency analysis results. The "Organism Compound Lists" worksheet lists the full set of compounds that appear in at least one media formulation for each organism species. The "Compound Frequencies" worksheet lists every compound that appears in at least one media formulation and the number of organism species known to utilize that compound (frequency). The "Organism Compound Numbers" lists every species and the number of compounds that appear in at least one media formulation for that species.
(XLSX)

File S1 Model simulation on known media. The compressed folder contains the *E. coli* model used for our simulations and an example Matlab script (growEcoliOnMedia.m) that demonstrates how to simulate growth of the model on media from MediaDB. This file simulates growth of *E. coli* on 11 different carbon sources corresponding to 11 different media in MediaDB.
(ZIP)

Acknowledgments

We would like to thank Nat Goodman for his input on the database schema and Zhilong Zhu and Hao Feng for assisting in curating the database during their time at the University of Illinois. We also thank Dr. Julie Bletz for critical readings of this manuscript and Denise Mauldin for her server support at ISB.

Author Contributions

Conceived and designed the experiments: MAR NDP. Performed the experiments: MAR NEA AH. Analyzed the data: MAR BDH ES. Contributed reagents/materials/analysis tools: MAR VC. Contributed to the writing of the manuscript: MAR VC BDH NEA AH ES NDP. Designed and created the database website: MAR VC.

References

1. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nature Biotechnology 28: 977–982.
2. Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5: 93–121.
3. Amann R (2000) Who is out there? Microbial aspects of biodiversity. Systematic and Applied Microbiology 23: 1–8.
4. Keller M, Zengler K (2004) Tapping into microbial diversity. Nature Reviews Microbiology 2: 141–150.
5. Alain K, Querellou J (2009) Cultivating the uncultured: limits, advances and future challenges. Extremophiles 13: 583–594.
6. Vartoukian SR, Palmer RM, Wade WG (2010) Strategies for culture of 'unculturable' bacteria. FEMS Microbiology Letters 309: 1–7.
7. Joint I, Mühling M, Querellou J (2010) Culturing marine bacteria–an essential prerequisite for biodiscovery. Microbial Biotechnology 3: 564–575.
8. Pham VH, Kim J (2012) Cultivation of unculturable soil bacteria. Trends in Biotechnology.
9. Prakash O, Shouche Y, Jangid K, Kostka JE (2013) Microbial cultivation and the role of microbial resource centers in the omics era. Applied Microbiology and Biotechnology 97: 51–62.
10. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science 296: 1127–1129.
11. Ferrari BC, Binnerup SJ, Gillings M (2005) Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. Applied and Environmental Microbiology 71: 8714–8720.

12. Yasumoto-Hirose M, Nishijima M, Ngirchechol MK, Kanoh K, Shizuri Y, et al. (2006) Isolation of marine bacteria by in situ culture on media-supplemented polyurethane foam. Marine Biotechnology 8: 227–237.
13. Bollmann A, Lewis K, Epstein SS (2007) Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates. Applied and Environmental Microbiology 73: 6386–6390.
14. Bruns A, Cypionka H, Overmann J (2002) Cyclic AMP and acyl homoserine lactones increase the cultivation efficiency of heterotrophic bacteria from the central Baltic Sea. Applied and Environmental Microbiology 68: 3978–3987.
15. Bruns A, Nübel U, Cypionka H, Overmann J (2003) Effect of signal compounds and incubation conditions on the culturability of freshwater bacterioplankton. Applied and Environmental Microbiology 69: 1980–1989.
16. Nichols D, Lewis K, Orjala J, Mo S, Ortenberg R, et al. (2008) Short peptide induces an ''uncultivable'' microorganism to grow in vitro. Applied and Environmental Microbiology 74: 4889–4897.
17. D'Onofrio A, Crawford JM, Stewart EJ, Witt K, Gavrish E, et al. (2010) Siderophores from neighboring organisms promote the growth of uncultured bacteria. Chemistry & Biology 17: 254–264.
18. Janssen PH, Schuhmann A, Mörschel E, Rainey FA (1997) Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. Applied and Environmental Microbiology 63: 1382–1388.
19. Watve M, Shejval V, Sonawane C, Rahalkar M, Matapurkar A, et al. (2000) The 'K' selected oligophilic bacteria: a key to uncultured diversity? Current Science 78: 1535–1542.
20. Janssen PH, Yates PS, Grinton BE, Taylor PM, Sait M (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. Applied and Environmental Microbiology 68: 2391–2396.
21. Connon SA, Giovannoni SJ (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Applied and Environmental Microbiology 68: 3878–3885.
22. Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418: 630–633.
23. Sangwan P, Kovac S, Davis KE, Sait M, Janssen PH (2005) Detection and cultivation of soil Verrucomicrobia. Applied and Environmental Microbiology 71: 8402–8410.
24. Button D, Schut F, Quang P, Martin R, Robertson BR (1993) Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. Applied and Environmental Microbiology 59: 881–891.
25. Sait M, Hugenholtz P, Janssen PH (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environmental Microbiology 4: 654–666.
26. Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, Breznak JA (2004) New strategies for cultivation and detection of previously uncultured microbes. Applied and Environmental Microbiology 70: 4748–4755.
27. Davis KE, Joseph SJ, Janssen PH (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. Applied and Environmental Microbiology 71: 826–834.
28. Stott MB, Crowe MA, Mountain BW, Smirnova AV, Hou S, et al. (2008) Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. Environmental Microbiology 10: 2030–2041.
29. Zengler K, Toledo G, Rappé M, Elkins J, Mathur EJ, et al. (2002) Cultivating the uncultured. Proceedings of the National Academy of Sciences 99: 15681–15686.
30. Zengler K, Walcher M, Clark G, Haller I, Toledo G, et al. (2005) High-throughput cultivation of microorganisms using microcapsules. Methods in Enzymology 397: 124–130.
31. Ingham CJ, Sprenkels A, Bomer J, Molenaar D, van den Berg A, et al. (2007) The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. Proceedings of the National Academy of Sciences 104: 18217–18222.
32. Carini P, Steindler L, Beszteri S, Giovannoni SJ (2012) Nutrient requirements for growth of the extreme oligotroph '*Candidatus* Pelagibacter ubique' HTCC1062 on a defined medium. The ISME Journal 7: 592–602.
33. Song H, Kim TY, Choi B-K, Choi SJ, Nielsen LK, et al. (2008) Development of chemically defined medium for *Mannheimia succiniciproducens* based on its genome sequence. Applied Microbiology and Biotechnology 79: 263–272.
34. Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nature Reviews Microbiology 2: 886–897.
35. Covert MW, Famili I, Palsson BO (2003) Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? Biotechnology and Bioengineering 84: 763–772.
36. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Current Opinion in Biotechnology 14: 491–496.
37. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2008) Reconstruction of biochemical networks in microorganisms. Nature Reviews Microbiology 7: 129–143.
38. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 27: 29–34.
39. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Research 33: 5691–5702.
40. Schellenberger J, Park JO, Conrad TM, Palsson BØ (2010) BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11: 213.
41. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, et al. (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Research 41: D456–D463.
42. Bolton EE, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: integrated platform of small molecules and biological activities. Annual Reports in Computational Chemistry 4: 217–241.
43. Imanishi T, Nakaoka H (2009) Hyperlink management system and ID converter system: enabling maintenance-free hyperlinks among major biological databases. Nucleic Acids Research 37: W17–W22.
44. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 4: R54.
45. Benedict MN, Gonnerman MC, Metcalf WW, Price ND (2012) Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. Journal of Bacteriology 194: 855–865.
46. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41: D590–D596.
47. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673–4680.
48. Felsenstein J (1993) {PHYLIP}: phylogenetic inference package, version 3.5 c.
49. Subramaniam S (1998) The Biology Workbench—a seamless database and analysis environment for the biologist. Proteins: Structure, Function, and Bioinformatics 32: 1–2.