Knockout Sudoku, a Method for Rapidly Curating Gene Disruption Collections

Isao A. Anzai[1], Lev Shaket[1*], Oluwakemi Adesina[1*], Michael Baym[2], Buz Barstow[1]*

[1]Department of Chemistry, Princeton University, Princeton, New Jersey, USA

[2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA

[*]These authors contributed equally to this article

Correspondence should be addressed to B.B. (buz@princeton.edu) and M.B. (baym@hms.harvard.edu)

Abstract

Knockout Sudoku is a method for the construction of whole-genome knockout collections for a wide range of microorganisms with as little as 3 weeks of dedicated labor and at a cost of approximately $10,000. The method uses manual 4-dimensional combinatorial pooling, next-generation sequencing and a Bayesian inference algorithm to rapidly process and then accurately annotate the extremely large progenitor transposon insertion mutant collections needed to achieve saturating coverage of complex microbial genomes. Here we present a protocol for the generation, combinatorial pooling and annotation of highly oversampled progenitor collections and their subsequent algorithmically guided condensation and curation into high-quality collections suitable for rapid genetic screening and gene discovery.

Introduction

Over the last decade, next-generation sequencing has dramatically improved the accessibility of genetic information. However, the percentage of genes of unknown function in a genome sequenced today remains approximately the same, 30-40%, as it did a decade ago[1,2]. This situation is particularly acute for the most esoteric microbes that offer the most unique genetic resources to genetic engineering and synthetic biology[3].

Genetic screening of arrayed collections of clonal isolates remains the preferred method of characterization for many non-fitness-related microbial phenotypes including drug targets[4], virulence factors[5-7], and secondary and cryptic metabolism[8,9]. Genetic screens conducted with whole genome knockout collections such as the Yeast Knockout Collection (YKO)[10] and the Keio Collection of *E. coli* gene deletion mutants[11] set the gold standard for gene function discovery[12]. However, construction of these collections requires extremely large technical, time, and cost investments[12]. As a result, only a small number of their type have been built to date[13]. This has motivated the development of methods that use massively parallel (next-generation) sequencing and combinatorial pooling[14] to dramatically reduce the cost and increase the ease of annotation of arrayed collections of mutants created by random transposon mutagenesis[15,16]. These exciting breakthroughs have facilitated the construction of a growing number of condensed curated gene knockout collections of pathogenic organisms and surrogates[17-19].

Despite the considerable cost- and labor-saving advantages of recent combinatorial pooling methods, their reliance upon liquid-handling robotics remains an obstacle to widespread adoption. This barrier has spurred ongoing development of rapid, easy to use, and extremely low cost methods for combinatorial pooling[20], including Knockout Sudoku[3] which we have previously introduced. Knockout Sudoku[3] uses a set of Bayesian inference algorithms that compensate for the low complexity combinatorial pools produced by manual methods during the solution of sequencing data, allowing this method to annotate the extremely large random mutant collections needed to ensure high coverage of complicated microbial genomes. The software underlying Knockout Sudoku is available in the KOSUDOKU package (**Materials**).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Overview of Procedure

Knockout Sudoku constructs a curated, condensed, non-redundant whole-genome knockout collection by annotation of a highly oversampled progenitor transposon insertion mutant collection and condensation by choosing a single representative disruption mutant for each non-essential gene. A simple 4-dimensional combinatorial pooling scheme is used to prepare a library for a massively parallel sequencing experiment (**Fig. 1** and **Fig. 2**). The sequencing dataset is deconvolved to identify and locate mutants making up the progenitor collection[3] (**Fig. 4**).

The method is optimized for use with a 96-channel pipettor and allows the user to pool a single 96-well plate in only approximately one minute, compared with almost 2 hours in a robotic scheme[15]. This means that a small team can pool and cryopreserve a 40,000 member progenitor collection in just a single day, which is sufficient to achieve high coverage of a microbial genome with approximately 3,500-4,000 non-essential genes[3]. For example, mutant with a disruption in gene *X* in well F8 on plate 10 is placed into the Plate Row (PR) 3, Plate Column (PC) 2, Row F, and Column 8 pools (**Fig. 1**). An amplicon library is then constructed from each of these pools using a semi-random nested PCR reaction that amplifies the transposon insertion site for every mutant in each pool (**Figs. 2** and **3**). The same reaction adds sequencer compatible flow-cell binding sequences and barcodes, allowing all pools to be combined and sequenced in parallel. The appearance of amplicons within the sequencing dataset containing the sequence of gene *X* and barcodes corresponding to pools PR3, PC2, F and 8 permits the mutant to be located.

As the progenitor collection size grows, the number of transposon mutants that appear at multiple locations rises in tandem. The appearance of identical mutants in multiple wells, along with cross-contamination at the pooling step, complicate mutant location by producing a large number of artifactual assignments relative to the number of real locations[3]. The likelihood of each of these location assignments is calculated by a Bayesian inference algorithm informed by internal self-consistency within the sequencing dataset, allowing us to disregard the artifacts and find the real location of these mutants[3] (**Fig. 4**). The large size of the progenitor collections that can be analyzed by Knockout Sudoku affords a wide choice of transposons insertion positions for each gene. Additionally, the assessment of cross-contamination allows us to focus our efforts on mutant isolation through colony purification[3], simultaneously reducing the construction time and increasing the quality of the final condensed collection.

The progenitor collection catalog is used to direct the construction of a non-redundant quality-controlled whole-genome knockout collection (**Fig. 5**). A single representative mutant is selected for each gene disrupted in the progenitor collection by an algorithm that balances the likelihood that the mutant will knock out the function of a gene with the ease of isolation from any well co-occupants. These selected mutants go through a purification triage where those in singly-occupied wells (i.e., a well that contains a single mutant species) are simply re-arrayed into the first portion of the condensed collection plates. Mutants that co-occupy a well but are still desirable due to the proximity of their transposon to the translation start of a gene are colony purified. For each co-occupied well, the Knockout Sudoku algorithm predicts the smallest number of colonies that must be picked to isolate the mutant of interest. These colonies are picked and then added to the condensed collection plates. The entire condensed collection is then re-pooled and validated by a second round of sequencing and alternative, orthogonal sequence analysis. One representative of each type of mutant in the colony purified set is selected for insertion into the quality-controlled collection; further, mutants for any genes still lacking representatives are added from the progenitor collection (**Fig. 5**).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Experimental Design

Prior to building a Knockout Sudoku collection, carefully adapt this protocol, calculate the time required for specific tasks and thoughtfully schedule them. Some of the planning steps listed here are essential to proceed while others will greatly reduce stress, save time, and produce a higher quality collection.

1. Calculation of Minimum Progenitor Collection Size

Prior to picking a progenitor collection, calculate the minimum number of members needed to ensure saturating coverage of the genome of the parental organism. We have found that simple analytical formulas based upon non-essential gene counts tend to underestimate the minimum progenitor collection size and instead recommend algorithmic methods. We have developed two programs, KOSUDOKU-EGENEINDEX and KOSUDOKU-COLLECTIONMC, to aid in estimation of the minimum progenitor collection size, N_p, through a Monte Carlo algorithm (**Box 1**).

2. Transposon Mutagenesis and Library Construction

The current version of Knockout Sudoku relies upon generation of a transposon insertion mutant library through mutagenesis with a modified pMiniHimar transposon delivered by conjugation with *E. coli* WM3064[3,21]. This transposon system has been reliably used to generate single-gene transposon insertion mutants in organisms such as *Azospira suillum* PS[22], *Dickeya dadantii*[23], *Francisella tularensis*[24], *Geobacter sulfurreducens*[25], *Marinobacter subterrani*[26], *Myxococcus xanthus*[27-29] and *S. oneidensis*[3,21]. In principle any transposon system can be used with Knockout Sudoku given suitable modifications to the sequencing library generation primers (**Experimental Design 10**). If the pMiniHimar system is not satisfactory, recent reviews by Barquist *et al*.[30] and Kwon e*t al*.[31] provide an extensive resource for possible alternatives.

Adapt the transposon library construction protocol detailed in **Box 3** for the parental organism of the progenitor collection. The transposon insertion mutant generation rate is particularly sensitive to the composition of the solid media used in the mating step (**Box 3**, **steps 7** and **8**). To minimize multiplicity in the progenitor collection, we recommend that the library construction protocol be modified until it yields a progenitor library (the un-plated progenitor collection in liquid media) that contains at least 10 times more members than will be picked to form the progenitor collection.

Test a pilot transposon insertion library with a range of antibiotic concentrations in both solid and liquid media to select conditions that will avoid biasing the progenitor collection towards mutants with higher antibiotic tolerance while still maintaining sterility and elimination of un-mutagenized cells. We plated an *S. oneidensis* transposon insertion library onto a series of Petri dishes with antibiotic concentrations ranging from 2.5 to 50 μg/mL kanamycin (**Supplementary Fig. 1**). Based upon the results of this titration, we decided upon a working concentration of 30 μg/mL kanamycin for creation of the *S. oneidensis* Sudoku Collection. Some organisms that are naturally tolerant of antibiotics may require a much higher working concentration (sometimes as high as 300 μg/mL).

Determine the length of time, ΔT_{solid}, over which plated mutant colonies are viable and pickable. For example, while *E. coli* colonies can be picked even after a week of storage at 4 ˚C, *S. oneidensis* colonies continue to grow at this temperature. This makes colony discrimination by colony picking robots challenging and reduces the viable lifespan of a Petri dish to only a few days. For *S. oneidensis*, we found that colonies were best picked no more than 2 days (preferably 1) after plating. Finally, determine the shelf life, ΔT_{liquid}, of transposon insertion mutants of the parental strain in liquid media as this will set the maximum time available for pooling and cryopreservation of the progenitor collection after picking.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

3. Robotic Picking

For organisms with genomes that require large progenitor collections, we recommend the use of colony picking robots to speed construction. While they are an expensive investment, some models can be rented at a low cost for several weeks and some core facilities will include them. Prior to picking a complete collection, adjust the robot to accommodate the sensitivities of the organism while maintaining a high picking rate and minimizing cross contamination events. We suggest picking into a checkerboard pattern (48 colonies per 96-well plate) to assess the effect of splatter on cross-contamination. If this is an issue, adjust the speed of picking pin descent during inoculation. If considerable growth failure during test picking occurs, increase both the inoculation time and pin cooling time after sterilization. We used a Norgren Systems CP7200 robot to pick the progenitor collection for the *S. oneidensis* Sudoku Collection[3]. After all adjustments, we achieved an average picking rate, r_{pick}, of approximately 20,000 colonies per day (or ≈ 1,500 colonies per hour). We were able to pick approximately 200 colonies from each 16 mm Petri dish (n_{dish}).

4. Growth of Collection

Use a high-speed, tight-orbit incubator shaker specifically modified for growth of cultures on micro-well plates to incubate the picked progenitor collection. Consult the shaker's documentation or contact its manufacturer to determine its maximum safe loading capacity. From this, calculate the maximum number of plates and hence colonies that can be grown in a single load, $N_{incubate}$. Use the maximum loading capacity along with the time needed for the organism to reach saturation, $\Delta T_{saturation}$, to calculate the rate of mutant incubation, $r_{incubate}$. Pay attention to special growth requirements of the organism of interest including high illumination and high CO_2 concentrations.

■ **Caution.** If the progenitor collection plates are not fully filled with colonies, for instance by picking into a checkerboard pattern, there will be almost no effect on picking rate, but the number of colonies that can be incubated in a single load will be significantly reduced.

5. Estimate Pooling Rate

Practice pooling a small test collection (using either fresh LB or water with a small amount of detergent in place of microbial culture) with a full team of assistants following the protocol detailed in **Box 4** to estimate your achievable plate pooling rate, r_{pool}. In our experience a maximum pooling rate of ≈ one 96-well plate per minute, or ≈ 400 plates per day when allowing time for cryopreservation, was achievable with a team of 5 people.

■ **Caution.** If the progenitor collection plates are not fully filled with colonies, for instance by picking into a checkerboard pattern, there will be no effect on the rate of plates that can be pooled per day but the number of colonies that can be pooled per day will be halved.

6. Calculate Maximum Pickable and Poolable Mutant Batch Size

Many microorganisms are highly perishable on agar plates or in liquid media. Use the useful lifespan of the parental organism on solid (ΔT_{solid}) and liquid (ΔT_{liquid}) media to estimate the maximum mutant batch size, N_{batch}, that can be plated, picked, and pooled without risking the viability of the collection. The maximum batch size is equal to the smallest of: the maximum batch that can be plated (N_{plate}); the maximum batch that can be picked (N_{pick}) and the maximum batch that can be pooled (N_{pool}).

If the minimum progenitor collection size needed to achieve saturating coverage of the genome (**Experimental Design 1**), $N_{p,}$ exceeds N_{batch}, the construction of the progenitor collection should be split into batches of mutants that are plated, picked, pooled, and cryopreserved before another batch is started.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

To assist in calculation of the maximum batch size and scheduling the picking and pooling of the progenitor collection we have included a program, KOSUDOKU-BATCH in the KOSUDOKU package. This program estimates the maximum batch size and provides guidelines for the maximum picking time per day, and the number of days to be dedicated to colony picking. Use of this program is described in **Box 2**.

It is safe to assume that the maximum number of colonies that can be plated onto agar in a single day is far larger than any other limiting factor. The upper limit on N_{plate} is more likely to be the maximum number of colonies that can be incubated, N_{incubate},

$$N_{\text{plate}} \leq N_{\text{incubate}}. \tag{1}$$

We estimate the upper limit on the number of colonies that can be picked, N_{pick}, as the smaller of the maximum picking rate per day and the maximum number that can be incubated per day multiplied by the useful lifetime of plated colonies, ΔT_{solid},

$$N_{\text{pick}} = \min(r_{\text{pick}}, r_{\text{incubate}})\Delta T_{\text{solid}}. \tag{2}$$

In the event that we the lifespan of the organism on liquid media is shorter than that on solid media,

$$N_{\text{pick}} = \min(r_{\text{pick}}, r_{\text{incubate}})\Delta T_{\text{liquid}}. \tag{3}$$

We strongly recommend completing picking before moving on to pooling. If the time needed for picked colonies to reach saturating density in liquid media, $\Delta T_{\text{saturation}}$, exceeds 24 hours for the organism of interest, pick for a single day only, waiting for colonies to appear and immediately moving on to pooling.

Estimate the time available for pooling by considering the maximum viability time of the progenitor collection. This is dependent upon the shelf life of the organism in liquid media; the time spent picking, ΔT_{pick}; the amount of rest taken between picking and pooling, ΔT_{rest}; and the time needed for picked colonies to reach saturating density,

$$\Delta T_{\text{pool}} \leq \left(\Delta T_{\text{liquid}} - \Delta T_{\text{pick}} - \Delta T_{\text{rest}} + \Delta T_{\text{saturation}}\right). \tag{4}$$

The number of colonies that can be pooled can be estimated as,

$$N_{\text{pool}} = r_{\text{pool}}\Delta T_{\text{pool}} n_{\text{plate}}, \tag{5}$$

where n_{plate} is the number of colonies per plate. Thus, the maximum colony batch size that can be processed without risking the viability of any member is,

$$N_{\text{batch}} = \min\left(N_{\text{plate}}, N_{\text{pick}}, N_{\text{pool}}\right). \tag{6}$$

7. Cryopreservation

Before commencing construction of the progenitor collection determine the optimal freezing storage conditions necessary for prolonged viability of the target organism. In particular, conditions should not only ensure retrieval of the wild-type but also be sufficiently mild for compromised mutants while

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

simultaneously allowing accurate and rapid pipetting of the cryoprotectant. We found that concentrations of glycerol below 20% were most suitable for fast and accurate pipetting with a 96-channel pipettor. Make sure to test storage and retrieval of mutants of the parental organism in the exact type of plate that will be used for storage of the progenitor collection. For construction of the *S. oneidensis* Sudoku collection, we tested a range of antibiotic and glycerol concentrations for growth and retrieval from frozen storage at -80 ˚C of a test set of transposon mutants and decided upon 30 µg/mL Kanamycin for both initial growth and retrieval of the collection. For cryopreservation we added an equal volume of 20% sterile glycerol to each well in the collection to reach a final concentration of 10%.

It is useful to measure the number of freeze-thaw cycles that members of the collection can withstand. If the progenitor collection can tolerate multiple freeze-thaw cycles, the need to prevent it from thawing during re-arraying to produce the condensed collection is removed, allowing either much faster manual re-arraying or even faster robotic re-arraying. We suggest picking a test set of mutants into a 96-well plate and freezing it under chosen conditions. Thaw this plate and replicate it before re-freezing. Repeat this multiple times and note how many wells are successfully replicated after each thaw-freeze cycle. Example data from an experiment of this type can be found in **Supplementary Fig. 2**.

8. Scheduling

Execution of most steps in Knockout Sudoku can typically be performed with a team of two, but pooling and re-array do benefit from additional personnel. We pre-determine a schedule prior to beginning any section of this protocol that ensures these labor-intensive steps fall on weekdays when help is most easily available.

9. Combinatorial Pooling Layout

Decide upon a combinatorial pooling layout prior to picking and pooling. Estimate how many colonies you plan to pick (**Experimental Design 1**) and how many plates will be required to store these. Assign each plate to a position in a square grid and prepare labels for each plate that include its number and coordinates in the grid. A sample plate grid is shown in **Supplementary Table 1**.

10. Amplicon Library Generation Primer Design

Primers used for the generation of the amplicon libraries in the construction of the *S. oneidensis* Sudoku Collection[3] are shown in **Supplementary Table 2**. To guide adaptation of these primers to an alternate transposon system, a detailed schematic of the amplicon library generation procedure is shown in **Fig. 3**. Should a transposon system other than pMiniHimar be used, place the defined transposon binding primers as close to the transposon-genome junction as possible to avoid creation of extremely short amplicons that contain no genomic sequence. Ensure that the primer melting temperatures are approximately 58 ˚C.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Procedure

Timing for steps up to the generation of pool amplicon libraries (**step 7**) are likely to be dependent upon choice of parental organism and certainly dependent on choice of colony picking device and available personnel. Re-estimate the effect of parental organism growth time and shelf life, equipment availability and team size on N_{batch} (**Experimental Design 6**). Timings shown here are examples from the creation of the *S. oneidensis* Sudoku Collection[3].

Calculate the Progenitor Collection Construction Timing Schedule. Timing ≈ 1 Day.

1. Calculate the minimum progenitor collection size using the programs in **Box 1**.
2. Calculate the N_{batch} along with picking sub batch size, the picking time per day, total days of picking and acceptable rest time between picking and pooling using the programs in **Box 2**.

Build a Transposon Insertion Library. Timing ≈ 4 Days.

3. Construct a transposon insertion mutant library (**Box 3)** and plate out a pickable and poolable batch of colonies, calculated in **step 2**. Consider plating an excess of colonies (25-50%) when preparing a progenitor collection batch.

Pick the Progenitor Collection. Timing ≈ 1 Day per 20,000 Colonies.

Timing estimates assumes the use of a Norgren Systems CP7200 colony picking robot adjusted to accommodate the sensitivities of an organism such as *S. oneidensis*.

4. Robotically pick colonies into 96-well plates containing 100 µL LB (or preferred growth media) with antibiotics. Seal each plate with a sterile Aeraseal membrane.
5. Grow mutant colonies for $\Delta T_{saturation}$ (21 hours for *S. oneidensis*) in a shaking incubator at 900 rpm at preferred growth temperature (30 ˚C for *S. oneidensis*).
6. If the number of picking days calculated in **step 2** is greater than 1, return to **step 4** the next day until the colony batch is complete.

■ **Pause Point.** While a break here is certainly possible, do not risk the viability of the progenitor collection by exceeding the break time determined in **step 2**.

Pool the Progenitor Collection. Timing ≈ 1 Day per 400 96-well Plates.

7. Pool the progenitor collection batch using the procedure outlined in **Box 4**.

■ **Pause Point.** Pooled, un-lysed cultures appear to be stable for at least several days without any impact on the quality of the extracted genomic DNA or mutation location. However, we strongly recommend that you be mindful of the viability time of the organism in liquid media and the time since the first micro-well plates in the batch reached saturation after picking.

8. Extract genomic DNA as outlined in **Box 5**.

■ **Pause Point.** The extracted genomic DNA can be stored for at least 1 year at -80 ˚C.

9. If more batches of colonies need to be picked to achieve saturating coverage of the genome, return to **step 3**. Otherwise, continue to **step 10**.

Generate Pool Amplicon Libraries. Timing ≈ 1 Day.

10. Generate pool amplicon libraries from the pooled progenitor collection mutants using the procedure in **Box 6**.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

■ **Caution.** Do not attempt to normalize the amount of genomic DNA template loaded into the amplicon generation reaction for the amount of each species estimated to be in each pool. Instead, add equal volumes of template to each reaction unless there is a gross discrepancy in concentration between pools of extracted genomic DNA.

11. Inspect molecular weight distributions of the pool amplicon libraries using the diagnostic gels produced in **step 10** and compare with the example images in **Supplementary Figs. 4** and **6**. If the distributions look like those in S**upplementary Fig. 3**, continue to **step 12.** If the distributions have more distinct bands like those in **Supplementary Fig. 5** go to **Troubleshooting 1**.

■ **Pause Point.** The pool amplicon libraries can be stored for several months at -20 ˚C.

Combine Pool Amplicon Libraries and Purify by Molecular Weight. Timing ≈ 6 Hours.

12. Collect the pool amplicon libraries into a single vial and purify by molecular weight using the protocol in **Box 7**.

■ **Caution.** Do not attempt to normalize the amount of PCR product added to the sequencing library for the amount of each species estimated to be in each pool. Instead add equal volumes from each library.

■ **Pause Point.** The combined and purified pool amplicon library can be stored for several months at -20 ˚C.

Sequence the Pool Amplicon Libraries by Illumina Sequencing

13. Submit the combined pool amplicon libraries for quality control by your local sequencing service. If a problem the libraries is raised consult **Troubleshooting 2.**
14. Sequence combined libraries on 2 lanes of an Illumina HiSeq in 67 bp single-end read mode.

■ **Pause Point.** As long as the progenitor collection is stable at -80 ˚C there can be an indefinite pause for data analysis.

Predict the Contents of Each Well in the Progenitor Collection. Timing ≈ 1 Day.

15. Solve the progenitor collection contents using the data analysis procedure detailed in **Box 8**.

Verify the Predictions of Progenitor Collection Contents with Sanger Sequencing

16. Test the predictions of the Knockout Sudoku algorithm by picking a random set of 10 to 94 mutants from the progenitor collection and re-arraying them into a single 96-well plate. Grow the mutants to saturation overnight.
17. Generate amplicons from each mutant culture using the procedure detailed in **Box 9**.
18. Send the amplicons for standard Sanger sequencing with PCR product clean up.
19. Test predictions of amplicon sequences against the results of Sanger sequencing using the programs listed in **Box 10.**
20. If the predictions match the results of Sanger sequencing, continue to **step 21**. If not, see **Troubleshooting 5**.

Select Mutants from the Progenitor Collection for the Condensed Collection. Timing ≈ 1 Day.

21. Use the KOSUDOKU-CONDENSE program to select a complete, non-redundant set of gene disruption mutants from the progenitor collection and generate instructions to retrieve and re-array them (**Box 11**).

Prepare the Condensed Collection

22. Re-array the selected single-occupancy wells in the progenitor collection into the first section of the condensed collection using the condensation instructions generated in **step 21** and the procedure in **Box 12**. An example re-array file is attached as **Supplementary Table 7**.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

23. Simultaneously colony purify the selected multiple-occupancy wells in the progenitor collection into the second section of the condensed collection using the condensation instructions generated in **step 21** and the procedure in **Box 12**.

■ **Caution.** The orthogonal data analysis algorithm used by KOSUDOKU-ISITINTHERE (**Box 13**) tests for the presence of a particular sequence in particular well rather than locating a sequence. Make sure to carefully follow the re-array and colony purification instructions and note any deviations from them.

24. Freeze re-arrayed plates at -80 ˚C with cryoprotectant.

■ **Pause Point.** As long as the condensed collection is stable at -80 ˚C there can be an indefinite pause prior to sequence validation.

Validate the Identities of Mutants in the Condensed Collection by Re-pooling and Re-sequencing

25. Plan out an approximately square plate grid for re-sequencing of the condensed collection. A suggested grid will have been included in the output of KOSUDOKU-CONDENSE from **step 21**.
26. Retrieve the condensed collection from the freezer if necessary.
27. Pool the condensed collection using the protocol in **Box 4**.
28. Freeze or re-freeze the condensed collection.
29. Extract genomic DNA as outlined in **Box 5**.

■ **Pause Point.** The extracted genomic DNA can be stored for at least 1 year at -80 ˚C.

30. Generate pool amplicon libraries using the protocol in **Box 6**.
31. Inspect the pool amplicon libraries by gel electrophoresis (**Box 6**). Compare the image of the gel with **Supplementary Figs. 4** and **6**. While the molecular weight distribution of each pool should look considerably less continuous than that shown in **Supplementary Fig. 3** it should not show considerably more band structure than that shown **Supplementary Fig. 5**. If the molecular weight distribution is not as expected consult **Troubleshooting 6**.

■ **Pause Point.** The pool amplicon libraries can be stored for several months at -20 ˚C.

32. Collect pool amplicon libraries into single vial and purify by molecular weight using the protocol in **Box 7**.

■ **Pause Point.** The combined and purified pool amplicon library can be stored for several months at -20 ˚C.

33. Submit the purified amplicon libraries for quality control by your local sequencing service. Should the libraries not be acceptable to the sequencing service consult **Troubleshooting 2**.
34. Sequence the libraries on 1 lane of an Illumina HiSeq 2500 in single end 67 bp read mode.
35. Validate the identities of the mutants in the condensed collection using the KOSUDOKU-ISITINTHERE program (**Box 13**).
36. Check the error rate in the condensed collection reported by KOSUDOKU-ISITINTHERE. We have found a 4-8% error rate is acceptable, but if it is higher see **Troubleshooting 7**.

Generate the Quality-controlled Collection

37. Generate instructions for further condensation of the colony purified second section of the condensed collection and addition of any missing mutants using the KOSUDOKU-QC program (**Box 14**).
38. Follow the re-array instructions generated in **step 37** to generate the quality-controlled collection.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Protocol Boxes

Box 1: Estimation of Progenitor Collection Size. Timing ≈ 1 Day.

1. Obtain a GenBank format whole genome sequence for the collection parental organism and a listing of gene essentiality data. A sample gene essentiality data file can be found in **Supplementary Table 3**. Sample input files for all programs are attached as **Supplementary Files 1** to **11** and can be found in the sample input section of the KOSUDOKU package.
2. Run the KOSUDOKU-EGENEINDEX program to generate an index of all possible transposon insertion locations in the parental organism genome and the associated locus. An example input file is attached as **Supplementary File 1**.

   ```
   > kosudoku-egeneindex ../input/kosudoku-egeneindex/kosudoku-egeneindex.inp
   ```

3. Prepare an input file for KOSUDOKU-COLLECTIONMC program that lists the location of the output of KOSUDOKU-EGENEINDEX. An example input file is attached as **Supplementary File 2**.
4. Run the KOSUDOKU-COLLECTIONMC program to generate a rarefaction curve that estimates genome coverage as a function of progenitor collection size.

   ```
   > kosudoku-collectionmc ../input/kosudoku-collectionmc/kosudoku-collectionmc.inp
   ```

Box 2: Calculation of the Progenitor Collection Construction Parameters. Timing ≈ 30 Minutes.

1. Use the KOSUDOKU-BATCH program to calculate the maximum size of a batch of transposon mutants that can be pooled without compromising sample viability. A sample input file for KOSUDOKU-BATCH is included as **Supplementary File 3**.

   ```
   > kosudoku-batch ../input/kosudoku-batch/kosudoku-batch.inp
   ```

Box 3: Construction of Transposon Insertion Library. Timing ≈ 4 Days.

Creation of transposon mutant libraries in Knockout Sudoku is largely based upon standard transposon mutagenesis procedures. However, we have found that some extra care, particularly haste in plating, may result in considerably improved diversity in the picked progenitor collection. Follow this protocol to first determine the transposon insertion mutant density by plating a dilution series and then repeat the protocol exactly for a second time to generate the progenitor collection. Timing assumes the use of microorganisms that can grow to saturation overnight.

1. Day 1: Streak out *E. coli* WM3064 carrying pMiniHimarFRT for single colonies on LB agar with 50 μg/mL kanamycin (kan) and 90 μM diaminopimelic acid (DAP). Grow overnight at 37 ˚C. Recovery of DAP auxotrophs from frozen stocks can sometimes be challenging, so be prepared to streak out more material from the frozen stock than from conventional *E. coli* frozen stocks.
2. Day 1: Streak out the parental strain for single colonies on LB agar (or preferred alternative media). Grow overnight at preferred temperature (30 ˚C for *S. oneidensis*).
3. Day 2: Pick a single colony of *E. coli* WM3064 and inoculate a 20 mL culture of LB media with 50 μg/mL kan and 90 μM DAP. Grow overnight at 37 ˚C.
4. Day 2: Pick a single colony of the parental strain and inoculate an overnight 3 mL culture of LB (or preferred alternative media).
5. Day 3: Resuspend the *E. coli* culture in 5 mL LB and then wash once in 5 mL of fresh LB. Finally, resuspend the *E. coli* cell pellet in 800 μL of LB.
6. Day 3: Resuspend the parental strain culture in 5 mL LB, spin down and then resuspend the cell pellet in 600 μL of LB.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

7. Day 3: Gently mix 400 µL of resuspended parental strain culture with 400 µL of resuspended *E. coli*. Gently spot the 800 µL of mixed cell suspension onto an LB agar (or alternative media) Petri dish and incubate overnight at the preferred growth temperature of the parental strain.
8. Day 3: As controls, also gently spot the remaining 200 µL of resuspended parental strain and 400 µL of *E. coli* onto separate LB petri dishes.
9. Day 4: Gently scrape the mating mixture and controls from their respective plates with sterile cell scraping tools and 5 mL of fresh LB each. Collect the resuspended cells into sterile disposable test tubes.
10. Day 4: Either: make a dilution series of the mating mixture and controls at dilution factors ranging from 10^{-1} to 10^{-8} dilution in LB. Plate 100 µL of cell suspension onto a series of LB-kan agar Petri dishes with a sterile spreading tool. Incubate these plates overnight at 30 ˚C.

    Alternatively: Dilute the library to 2000 colony forming units per mL using the transposon mutant density calculated in **step 11**. If plating the progenitor collection onto 16 mm Petri dishes spread 100 µL (200 colonies) with plating beads. If plating onto colony picker trays, plate 1 mL (2000 colonies) with a cell spreading tool. Incubate these plates overnight at the preferred growth temperature of the parental strain.

■ **Caution.** Although the mating mixture can be stored for several days we have found that collections that were not plated rapidly (i.e., immediately) tended to be highly repetitive (containing many more copies of identical transposon insertion mutants than random chance would suggest).

11. Day 5: Calculate the density of transposon insertion mutants in the mating mixture by counting colonies on the dilution series Petri dishes. Alternatively: pick the progenitor collection into liquid media. In both cases, check that there is no cell growth on either of the control plates.

Box 4: Pooling and Cryopreservation of Transposon Insertion Mutant Collection. Timing ≈ 1 Day per 400 96-well Plates or ≈ 1 Minute per 96-well Plate

The pooling procedure can be very fast, but it benefits from a large team. While most of the procedures in this protocol need only 1 to 2 people, pooling is greatly accelerated with a team of 4 to 5. Furthermore, this process is highly amenable to parallelization if more than one 96-channel pipettor is available.

1. Organize the pooling team. We have found that interchanging roles helps with focus, although some people may prefer to stay with a single role for the entire procedure. We recommend that one person operate the 96-channel pipettor while another acts as a spotter and hands the operator pooling plates. A third person should be responsible for organizing the progenitor collection plates and the fourth should be responsible for sealing and unsealing plates. Finally, a reserve fifth person should be resting.
2. For each plate in the progenitor collection, ensure that it is laid out in the order specified in the plate grid (**Supplementary Table 1**). We recommend stacking the progenitor collection plates in descending order by plate row (for example: plate 1 should be at the top of the plate row 1 stack and plate 21 should be at the bottom; plate 22 is on the top of the plate row 2 stack and plate 42 is on the bottom).
3. Set the vertical stop posts on the left side Liquidator pads so that the tips do not touch the destination plates.
4. Wipe down the Liquidator and surrounding workspace with 70% ethanol.
5. Label all plate row and plate column OmniTrays.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

6. Prepare a stack of plate row OmniTrays and a stack of plate column OmniTrays in descending order (plate row 1 OmniTray is on the top of the plate row stack and the plate row 20 OmniTray is on the bottom).
7. Prepare a set of row and column plates (96-well plate format trays where each row or column is an entire well). These can be reused, but we find that it is more convenient to use new ones each time a plate is almost full to avoid spillover between wells. Carefully mark the orientation of these plates.
8. Place the plate column OmniTray stack next to the Liquidator. Place the plate row OmniTray stack further away. Place the first (next) plate row OmniTray onto the back left Liquidator pad and leave it there until all plates in the plate row have been pooled.
9. Load the Liquidator with a 20 µL sterile non-filter tip box (from the back right pad). Make sure to press down firmly on the loading arm.
10. Place the current progenitor collection source plate onto the front right Liquidator Pad.
11. Place the row plate onto the front left Liquidator pad. Withdraw 10 µL from the source plate and deposit into the row plate. Make sure that the tips do not touch the destination plate.
12. Place the column plate onto the front left Liquidator pad. Withdraw 10 µL from the source plate and deposit into the column plate.
13. Place the plate column tray onto the front left Liquidator pad. Withdraw 10 µL from the source plate and deposit into the plate column tray.
14. Move the left side sliding tray on the Liquidator forward so that the plate row tray can now be accessed by the pipettor. Withdraw 10 µL from the source plate and deposit into the plate row tray.
15. Eject the Liquidator tips and discard. Return to step 8 and place the next plate on the front right Liquidator pad. Continue this loop until all plates in the plate row are pooled. Set aside the plate row tray as it is complete. Place the next plate row OmniTray onto the back left Liquidator pad and continue. After all plates for that day have been pooled, proceed to cryopreserving the plates.
16. Sample the volume of culture left in the source plates, (there should be approximately 50 µL) and add an equal volume of 20% glycerol using sterile, filter 200 µL Liquidator tips.
17. Seal each plate with aluminum sealing foil.
18. Place the sealed plates in the shaker incubator and shake for approximately 2 minutes to mix the glycerol. Freeze at -80 ˚C.
19. Transfer the contents of the pooling plates to Falcon tubes for genomic DNA extraction (**Box 5**).
20. If pooling of the progenitor is not complete, return to **step 1** the next day.

Box 5: Extraction of Genomic DNA from Pools. Timing ≈ 6 Hours.

This procedure requires use of the Zymo Research genomic DNA mini-prep kit.

1. Obtain Falcon tubes containing each pool and vortex briefly to resuspend cells.
2. Using a pipettor, extract 2 mL from each pool into a 2 mL low DNA binding microcentrifuge tube.
3. Spin down the cells at 21,000 × *g* for 30 seconds and discard the supernatant.
4. Add 500 µL genomic lysis buffer to each tube and immediately vortex. Allow to stand for 5 to 10 minutes.

■ **Pause Point.** Cell extracts can be stored at 4 ºC for several hours.

5. Transfer lysed culture to a spin column in a collection tube.
6. Centrifuge at 10,000 × *g* for 1 minute and discard the flow-through.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

7. Add 200 µL of pre-wash buffer.
8. Centrifuge at 10,000 × *g* for 1 minute and discard the flow-through.
9. Add 500 µL of wash buffer.
10. Centrifuge at 10,000 × *g* for 1 minute and discard the flow-through.
11. Place the spin column in a fresh low DNA binding microcentrifuge tube.
12. Add 50 µL of elution buffer and incubate for 2 to 5 minutes.
13. Centrifuge at maximum speed for 30 seconds to elute.

Box 6: Generation of Pool Amplicon Libraries. Timing ≈ 1 Day.

Pool amplicon library construction for Knockout Sudoku uses a 2-step nested PCR reaction[32,33]. This reaction amplifies a portion of the chromosome adjacent to the transposon present in each collection member and adds Illumina TruSeq flow-cell-binding and read-primer-binding sequences to the 3' and 5' ends of the amplicon while replacing the standard Illumina index sequence with a custom barcode sequence for each pool. Sanger verification of Knockout Sudoku predictions relies upon a similar version of this procedure (**Box 9**).

PCR program for step 1 is as follows:

1. Make a master mix of primers, reaction buffer, and enzyme. For each well use a total reaction volume of 20 µL containing 4 µL of 5× OneTaq standard reaction buffer and 2 units of OneTaq DNA Polymerase (0.4 µL). Add 200 nM of each primer: HimarSeq1.2, CEKG2C-IllR2, and CEKG2D-IllR2 (**Supplementary Table 2**) and 200 µM dNTPs. Adjust volume to 18 µL per well with deionized sterile water.
2. Dispense the master mix into a skirted 96-well PCR plate.
3. Add 2 µL per well of purified pooled genomic DNA (**Box 5**) at a concentration of ≈ 100 ng/µL as a template.

■ **Caution.** Do not attempt to normalize the amount of genomic DNA template loaded into the amplicon generation reaction for the amount of each species estimated to be in each pool. Add equal volumes of template to each reaction unless there is a gross discrepancy in concentration between pools of extracted genomic DNA.

4. Seal plate with PCR strip caps.
5. Transfer the plate to the PCR machine and run the following program: 5 minutes at 95 °C; 6 cycles of 30 seconds at 95 °C, 30 seconds at 42°C (lowering by 1 °C per cycle), 3 minutes at 68 °C; then 24 cycles of 30 seconds at 95 °C, 30 seconds at 45 °C, 3 minutes at 68 °C; and finally 5 minutes at 68 °C followed by storage at 4 ˚C.

■ **Pause Point.** The step 1 PCR products can be stored at 4 ºC overnight.

The PCR program for step 2 is as follows:

1. Make a master mix of primers, reaction buffer and enzyme. For each well use a total reaction volume of 50 µL containing 10 µL of 5× OneTaq standard reaction buffer and 5 units of OneTaq DNA Polymerase (1 µL). Add 50 nM of each primer HimarSeq2.4-4N, HimarSeq2.4-5N, HimarSeq2.4-6N and HimarSeq2.4-7N and 200 µM dNTPs. Adjust volume with deionized sterile water so that volume per well is 50 µL after addition of 0.5 µL template and barcoding primer.
2. Dispense master mix into PCR plate.
3. Add 200 nM of appropriate pool barcoding primer (**Supplementary Table 2**) to each well.
4. Add 0.5 µL of corresponding reaction 1 product as template.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

5. Seal plate with PCR strip caps.
6. Transfer the plate to the PCR cycler and run the following program: 30 cycles of 30 seconds at 95°C, 30 seconds at 56°C, 3 minutes at 68°C; 5 minutes at 68°C; and finally storage at 4 ˚C.

■ **Pause Point.** The step 2 PCR products can be stored for several months at -20 ˚C.

Inspect the pool amplicon libraries by gel electrophoresis prior to combination and purification.

1. Cast a large (≈ 200 mL) 1% agarose gel with 1× TAE and SYBR-SAFE DNA binding stain. Use a high-well-density comb to allow transfer of samples directly from a PCR plate with a multi-channel pipettor. Allow the gel to solidify for at least one hour.
2. Combine 5 µL of each pool amplicon library with 1 µL of 6× loading dye in a clean PCR plate. Mix by vortexing.
3. Load the mixed pool amplicon libraries onto the gel with a multi-channel pipettor and 1 µL of kb DNA ladder to at least one well per row on the gel.
4. Run the gel at 150 V for 30 minutes.
5. Image the gel with a gel documentation system with appropriate filters for SYBR-SAFE gels.

Box 7: Combination and Molecular Weight Selection of Pool Amplicon Libraries for Next-generation Sequencing. Timing ≈ 6 hours.

This purification step produces a sequencing library with a molecular weight distribution suitable for sequencing on an Illumina device.

1. Cast a small (≈ 8 cm long or 50 mL) 2% agarose gel with 1× TAE and SYBR-SAFE DNA binding stain. Allow to solidify for at least one hour.
2. Combine all amplicon samples generated in **Box 6** into a single microcentrifuge tube.

■ **Caution.** Do not attempt to normalize the amount of PCR product added to the sequencing library for the amount of each species estimated to be in each pool.

3. Measure concentration of sample by UV-vis spectrophotometry. The concentration should be around 200 ng/µL.
4. Combine 10 to 15 µL of combined PCR product (≈ 3 µg) with an appropriate volume of 6× loading dye (2 µL of dye for 10 µL product). Mix by vortexing.
5. Load the PCR sample onto the gel and add 1 µL of kb DNA ladder into the neighboring well.
6. Run the gel at 150 V for 30 minutes.
7. Image the gel with a gel documentation system with appropriate filters for SYBR-SAFE gels.
8. Carefully excise the section of the gel containing amplicons with lengths from 500 to 1000 bp with a scalpel.
9. Use a gel extraction kit to extract DNA from the gel fragment. Follow manufacturer's instructions and elute into a DNA low-binding microcentrifuge tube.
10. Measure concentration of sample by UV-vis spectrophotometry.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Box 8: Solution of Progenitor Collection. Timing ≈ 1 Day.

A workflow for the solution of the progenitor collection detailing input and output files is shown in **Fig. 4**. Sample input files are included as **Supplementary Tables** and **Files** with this article and in the sample input included with the KOSUDOKU package.

1. Build a BOWTIE2 index for your genome of interest.
   ```
   > bowtie2-build sample_genome.fa sample_genome
   ```
2. Generate a barcode file that contains a list of your barcode sequences (**Supplementary Table 2**) and the corresponding pool that they are used with. An example is included as **Supplementary Table 4**.
3. Prepare an input file for the sequencing dataset analysis software that will generate a pool presence table. An example file is attached as **Supplementary File 4**.
4. Run sequencing dataset analysis software to build a pool presence table.
   ```
   > kosudoku-seqanalyze ../input/kosudoku-seqanalyze/kosudoku-seqanalyze.inp
   ```
5. Examine the output of the sequence analysis program to find the percentage of reads that contain valid pool barcode and transposon sequences and that can be aligned to the reference genome. If this percentage is below ≈ 90% you should consult **Troubleshooting 3**.
6. Prepare an input file for the pool presence table analysis code. An example file is attached as **Supplementary File 5.**
7. Run the pool presence table analysis code to determine the effect of read count threshold on the taxonomy of pool presence table solution.
   ```
   > kosudoku-poolanalyze ../input/kosudoku-poolanalyze/kosudoku-poolanalyze.inp
   ```
8. Use the output of the pool presence analysis code to decide upon a read count threshold. We typically use a read count threshold of 5 reads, but this should only be decided after careful examination of the output of KOSUDOKU-POOLANALYZE. We suggest a read count threshold that maximizes the number of transposon locations that unambiguously map to locations within the progenitor collection.
9. Should the number of unique mutant sequences identified at this stage be considerably lower than the number of colonies picked to generate to the progenitor collection, consult **Troubleshooting 4**.
10. Use the read count threshold to prepare an input file for the pool presence table read count ratio fitting program. An example input file is attached as **Supplementary File 6**.
    ```
    > kosudoku-poolfit ../input/kosudoku-poolfit/kosudoku-poolfit.inp
    ```
11. Use the Bayesian inference parameters calculated in **step 10** to prepare an input file for the pool presence table solver. An example input file is attached as S**upplementary File 7** and example Bayesian inference parameters are attached as **Supplementary Table 5**.
12. Run the pool presence table solution program.
    ```
    > kosudoku-poolsolve ../input/kosudoku-poolsolve/kosudoku-poolsolve.inp
    ```

Box 9: Sanger Verification of Sudoku Predictions.

The preparation of amplicons for verification of progenitor collection content predictions by Sanger sequencing relies upon a similar procedure to that detailed in **Box 6**.

The PCR program for step 1 is as follows:

1. Make a master mix of primers, reaction buffer, and enzyme. For each well use a total reaction volume of 20 µL containing 4 µL of 5× OneTaq standard reaction buffer and 2 units of OneTaq DNA Polymerase (0.4 µL). Add 200 nM of each primer: HimarSeq2, CEKG2C, and CEKG2D

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

(**Supplementary Table 2**) and 200 μM dNTPs. Adjust volume to 19 μL per well with deionized sterile water.

2. Dispense the master mix into a 96-well skirted PCR plate.
3. Add 1 μL per well of fresh saturated bacterial culture as template.
4. Seal plate with PCR strip caps.
5. Transfer the plate to the PCR machine and run the following program: 5 minutes at 95 °C; 6 cycles of 30 seconds at 95 °C, 30 seconds at 42 °C (lowering by 1 °C per cycle), 3 minutes at 68 °C; then 24 cycles of 30 seconds at 95 °C, 30 seconds at 45 °C, 3 minutes at 68 °C; and finally 5 minutes at 68 °C followed by storage at 4 ˚C. Timing ≈ 3 hours.

■ **Pause Point.** The step 1 PCR products can be stored at 4 ºC overnight.

For step 2:

1. Make a master mix of primers, reaction buffer and enzyme. For each well use a total reaction volume of 20 μL containing 4 μL of 5× OneTaq standard reaction buffer and 2 units of OneTaq DNA Polymerase (0.4 μL). Add 200 nM of each primer HimarFRTSeq2 and CEKG4 and 200 μM dNTPs. Adjust volume with deionized sterile water so that volume per well is 19.5 μL.
2. Dispense master mix into PCR plate.
3. Add 0.5 μL of corresponding reaction 1 product as template.
4. Seal plate with PCR strip caps.
5. Transfer the plate to the PCR machine and run the following program: 30 cycles of 30 seconds at 95 °C, 30 seconds at 56 °C, 3 minutes at 68 °C; 5 minutes at 68 °C; and finally storage at 4 ˚C. Timing ≈ 3 hours.

■ **Pause Point.** The step 2 PCR products can be stored at 4 ˚C overnight or -20 ˚C for longer while awaiting Sanger sequencing.

6. Send the amplicons for standard Sanger sequencing with PCR product clean up.

Box 10: Analysis of Sanger Sequencing Reads for Verification of Knockout Sudoku Predictions. Timing ≈ 6 Hours.

1. After receiving sequencing data for verification of Knockout Sudoku predictions (**Box 9**), prepare an input file for the verification program (**Supplementary File 7**) along with a description of the sequencing results (**Supplementary Table 6**)
2. Run the prediction verification program.

```
> kosudoku-verify ../input/kosudoku-verify/kosudoku-verify.inp
```

Box 11: Selection of Mutants for the Condensed Collection. Timing ≈ 3 Hours.

1. Prepare an input file for the progenitor collection condensation program (**Supplementary File 8**).
2. Run the condensation code.

```
> kosudoku-condense ../input/kosudoku-condense/kosudoku-condense.inp
```

An example output file from KOSUDOKU-CONDENSE is attached as **Supplementary Table 7** and with the example output files in the KOSUDOKU package.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Box 12: Re-array of Progenitor Collection Single-occupancy Wells and Colony Purification of Multiple-occupancy Wells for Construction of Condensed Collection

■ **Caution.** Before taking samples from the progenitor collection, make sure to determine how many freeze-thaw cycles the organism of interest can tolerate (**Experimental Design 7**). If necessary, this procedure can be carried out manually with frozen samples but will take longer.

The re-array and colony purification instructions produced by the collection condensation program (**Box 11**) can be used to generate re-array instructions for a colony picking robot. If this option is available to you, we strongly recommend it, as it will dramatically speed re-arraying. To convert the output of KOSUDOKU-CONDENSE, consult the documentation for your colony picking robot. We plan to add automatic conversion to future versions of KOSUDOKU. At this time, colony purification of multiple-occupancy wells is done completely manually.

If manually re-arraying:

1. Organize your team. Manual re-arraying can by one person if necessary but can be readily parallelized.
2. Compile the condensed collection re-array instructions so that all wells on the same source plate, those requiring only re-array or needing colony purification, can be processed at the same time.
3. Fill and label destination plates with 100 µL per well of preferred growth media and antibiotics using sterile, filter 200 µL Liquidator tips. To make pooling of the condensed collection easier, add plate row and plate column assignments to the condensed collection plates. The first part of this collection will be used as destinations for source wells that only need re-array. The second section will be used as destinations for source wells that will be colony purified.
4. Transfer approximately 10 progenitor collection plates from the freezer to a foam cooler filled with dry ice and cover with an inverted foam cooler.
5. Remove the first (next) source plate at a time from the cooler.
6. Allow the plate to thaw, wipe the aluminum seal on the plate with 70% ethanol and wait for surface to dry.
7. Pierce the foil seal with a sharp sterile toothpick. If the well is to be re-arrayed, transfer a small amount of cells to the new destination plate determined by the re-arraying program. If the well is to be colony purified, streak out the cells onto a Petri dish clearly labelled with the source and destination wells for the sample.
8. Cover the hole created in the seal with a Tough-Spot.
9. Return to **step 7** and continue until all designated wells on the source plate have been sampled.
10. Return the plate to the cooler, and return to **step 5**.
11. Once the destination plate has been filled, cover it with an Aeraseal and grow it up in a shaking incubator at 900 rpm at appropriate temperature overnight. Continue retrieving samples from the progenitor collection until the end of the day.
12. Take petri dishes to a static incubator and grow overnight at an appropriate temperature.
13. The next day, when the cultures in the destination plate have reached saturation, sample the volume of culture left in the source plates (there should be approximately 90 µL) and add an equal volume of 20% glycerol using sterile, filter 200 µL Liquidator tips. Cover with an aluminum seal for cryopreservation and store at -80 ˚C.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

14. When colonies have appeared on the Petri dishes prepared the previous day, pick the number of colonies specified in the collection condensation instructions into the specified destinations. Be very careful to follow instructions in instruction file when doing this. If you do deviate from these instructions, be very careful to note them, and update the condensation instructions file to be used in sequence validation step (**Box 13**).
15. If more source plates need to be processed, return to **step 4** and continue.

If robotically re-arraying:

1. Organize your team. Robotic re-arraying requires a team of two people. We recommend that one person operates the colony picking robot and is responsible for re-array of single occupancy wells while the second is responsible for colony purification.
2. Convert the condensed collection re-array instructions into a format that is understood by the colony picking robot controller software. Only include wells in this file that only need re-array. Prepare a second human readable file that includes instructions for colony purification.
3. Fill and label destination plates with 100 µL per well of preferred growth media and antibiotics using sterile, filter 200 µL Liquidator tips. To make pooling of the condensed collection easier, add plate row and plate column assignments to the condensed collection plates. The first part of this collection will be used as destinations for source wells that only need re-array. The second section will be used as destinations for source wells that will be colony purified.
4. Transfer approximately 10 progenitor collection plates from the freezer to a foam cooler filled with dry ice and cover with an inverted foam cooler.
5. Remove the first (next) source plate at a time from the cooler.
6. Dust ice from the source plate and remove the aluminum seal before it thaws. Set it aside under a sterile flame to thaw.
7. Robotically re-array the source plate with the colony picking robot.

■ **Caution.** Make sure that the plate has thawed before re-arraying to avoid damage to the picking pins.

8. When re-array of the single occupancy wells on the source plate has been completed, hand the source plate to the person responsible for colony purification.
9. For each designated multiple-occupancy well on the source plate, streak out the cells onto a Petri dish clearly labelled with the source and destination wells for the sample.
10. Once all source wells on the source well have been sampled, return the plate to the cooler, and return to **step 5**.
11. Once the destination plate has been filled, cover it with an Aeraseal and grow it up in a shaking incubator at 900 rpm at appropriate temperature overnight. Continue retrieving samples from the progenitor collection until the end of the day.
12. Take petri dishes to a static incubator and grow overnight at an appropriate temperature.
13. The next day, when the cultures in the destination plate have reached saturation, sample the volume of culture left in the source plates (there should be approximately 90 µL) and add an equal volume of 20% glycerol using sterile, filter 200 µL Liquidator tips. Cover with an aluminum seal for cryopreservation and store at -80 ˚C.
14. When colonies have appeared on the Petri dishes prepared the previous day, pick the number of colonies specified in the collection condensation instructions into the specified destinations. Be very careful to follow instructions in instruction file when doing this. If you do deviate from these

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

instructions, be very careful to note them, and update the condensation instructions file to be used in sequence validation step (**Box 13**).

15. If more source plates need to be processed, return to **step 4** and continue.

Box 13: Validation of Condensed Collection by Orthogonal Sequence Analysis

To validate the predictions of the KOSUDOKU-POOLSOLVE program and that re-array and colony purification were performed correctly, one may elect to re-pool and re-sequence the condensed collection and perform orthogonal sequence analysis. This is particularly useful for selection and condensation of the colony purified section of the condensed collection.

1. Prepare an input file for the validation code (**Supplementary File 10**).
2. Run the validation code.

```
> kosudoku-isitinthere ../input/kosudoku-isitinthere/kosudoku-isitinthere.inp
```

Box 14: Generation of Instructions for Construction of Quality-controlled Collection

This code takes as input the output of collection validation program (**Box 13**) and outputs instructions for re-arraying wells in the colony purified section of the condensed collection (**Box 12**) to select only a single correct representative. Additionally, the code provides a list of any genes still missing from the condensed collection and gives instructions for their retrieval from the progenitor collection.

1. Prepare an input file for the quality control code (**Supplementary File 11**).
2. Run the quality control program.

```
> kosudoku-qc ../input/kosudoku-qc/kosudoku-qc.inp
```


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Troubleshooting

1. If the molecular weight distribution of the pool amplicon libraries generated from the progenitor collection resembles those shown in **Supplementary Fig. 3**, you are likely on the right track. However, if the distributions show more discrete band structure like those shown in **Supplementary Fig. 5** then it is possible that there is lack of diversity in the progenitor collection pool amplicon libraries. This could be due to several factors including perishment of samples prior to pooling and lack of diversity in either the plated progenitor collection or un-plated progenitor library. To address perishment, we strongly recommend carefully examining the viable lifespan of the progenitor collection in liquid media. Additionally, we suggest that you consult your notes to check the time between pooling and genomic DNA extraction (**Box 5**). To address lack of diversity in the progenitor library, carefully examine the mating protocol (**Box 3**) paying particular attention to the time between scraping the mating mix and plating. In all cases, remaking the progenitor collection is recommended.

2. When performing quality control on the Illumina-compatible amplicon library built by Knockout Sudoku (**Boxes 6** and **7**) we have often seen a discrepancy between the molecular weight distribution measured by agarose gel electrophoresis (**Supplementary Fig. 4**) and that measured by a BioAnalyzer, Genomic Tape Station or similar device. In these cases, your sequencing facility may recommend against sequencing. However, we have found that these libraries sequence very well in all cases.

3. Upon receipt of sequencing data, you may find that there are many reads that do not contain a significant amount of, if any, genomic sequence. In these cases the random primers used in library generation (**Box 6**) are likely binding inside the transposon, before the junction with the genome. If this is the case, redesign the transposon specific primers to bind at sites closer to the junction while maintaining the same melting temperature (≈ 58 ˚C). Simply re-run the pool amplicon construction reaction (**Box 6**), combine and re-purify by molecular weight (**Box 7**).

4. If analysis of the progenitor collection sequencing dataset does not find many unique transposon sites (at the very least it should contain approximately as many sites as mutants that were picked to form the progenitor collection) follow the suggestions for solving a lack of diversity in the progenitor collection listed in **Troubleshooting 1**.

5. If the results of Sanger sequencing verification of identities of mutants from the progenitor collection (**Box 10**) do not match predictions, make sure that these mutants were picked from the correct places. When picking mutants for testing, sterilize and then pierce the aluminum seal covering the plate with a toothpick and then cover the hole with a ToughTag so that it is possible to later check that the correct mutants was retrieved. To date, we have not seen a failure of prediction by Knockout Sudoku that could not be explained by a failure to retrieve the correct mutant from the progenitor collection. In all of these cases, we were able to use the progenitor collection catalog to track the mutant identified by Sanger sequencing back to a well adjacent to the one that was supposed to be picked from and were able to identify this failure thanks to the ToughTag patch.

6. To date, we have not seen a pool amplicon library produced by pooling the condensed collection that we did not find acceptable. However, if there is considerably more prominent band structure present in the molecular weight distribution of these amplicon libraries compared with that shown in **Supplementary Fig. 5**, we strongly suggest re-pooling the condensed collection.

7. If you find that the error rate in the quality-controlled collection is high (greater than 8%), it is likely due to errors in manual re-arraying and colony picking. Consider the recommendations in **Troubleshooting 5**.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Materials

Reagents

- LB medium powder; contains 10g tryptone, 5g yeast extract, 10g NaCl per liter of rehydrated media (MP Biomedicals, cat. no. MP113002042 or equivalent or alternative for chosen organism)
- Kanamycin sulfate (Sigma-Aldrich, cat. no. K1377 or equivalent or alternative for chosen transposon system)
- Isopropanol; low grade for sterilization (Sigma-Aldrich, cat. no. W292912)
- Glycerol; molecular biology grade (Sigma-Aldrich, cat. no. G5516 or equivalent)
- OneTaq DNA polymerase (New England BioLabs, cat. no. M0480S)
- Deoxyribonucleotide (dNTP) solution set; 100 mM of dNTP (New England BioLabs, cat. no. N0446S)
- OneTaq Standard Reaction Buffer (New England BioLabs, cat. no. B9022S)
- 2,6-Diaminopimelic acid (DAP) (Sigma Aldrich, cat. no. D1377-5G)
  Caution. DAP is an irritant. We recommend handling under a fume hood or wearing a dust mask and eye protection while handling.
- Genomic DNA mini-prep kit (Zymo Research, cat. no. D3006)
- Gel DNA extraction kit (Zymo Research, cat. no. D4001)
- DNA clean and concentrator kit; 5 µg capacity (Zymo Research, cat. no. D4003)
- SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, cat. no. S33102)
  Caution. while SYBR Safe is less mutagenic than ethidium bromide we still recommend caution when handling and prefer disposal as hazardous waste.
- 6× Gel loading dye; blue (New England BioLabs, cat. no. B7021S)
- 1 kb DNA ladder (New England BioLabs, cat. no. N3232S)
- Primer sequences are listed in **Supplementary Table 2**.
- Blot replicator pin cleaning solution (V&P Scientific, cat. no. VP 110)
- Agarose (Lonza, Seakem LE, cat. no. 50004 or equivalent)
- Tris Acetate-EDTA buffer concentrate (Sigma-Aldrich, cat. no. 67996-10L-F)

Consumables

- Petri dishes; sterile, 15 mm (Fisher Scientific, cat. no. FB0875712)
- 96-well storage microplates; sterile, polypropylene, U-bottom (USA Scientific, cat. no. 1830-9610)
- 96-well microplate lids; sterile, low evaporation, polystyrene (Corning, cat. no. 3931)
- 96-well format row plates; sterile (Phenix Research Products, cat. no. RRI-3028-ST)
- 96-well format column plates; sterile (Phenix Research Products, cat. no. RRI-3029-ST)
- 96-well format Omni-Tray single well microplates; sterile (Nalgene Nunc, cat. no. 242811)
- Aluminum foil seals for 96-well microplates; non-sterile (USA Scientific, cat. no. 2923-0100)
- Aeraseal membranes for 96-well microplates; sterile (Excel Scientific, cat. no. BS-25)
- Snap cap tubes; 14 mL, sterile (Corning, cat. no. 352001)
- 50 mL Serological pipettes (Genesee Scientific, cat. no. 12-107)
- 10 mL Serological pipettes (Genesee Scientific, cat. no. 12-104)
- 5 mL Serological pipettes (Genesee Scientific, cat. no. 12-102)
- Liquidator tips 20 µL; sterile, non-filter (Rainin, cat. no. LQR-20S)
- Liquidator tips 200 µL; sterile, filter (Rainin, cat. no. LQR-20F)
- Liquidator media dispensing reservoirs (Rainin, cat. no. LR-R2-PB-5-S)
- Pipette tips for gel loading; 20 µL, non-sterile (Rainin, cat. no. SS-L10)
- Pipette tips for enzyme handling; 20 µL, sterile, filter, low-retention (Rainin, cat. no. RT-L10FLR)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

- Pipette tips for PCR preparation; 20 µL, sterile, filter (Rainin, cat. no. RT-L10F)
- Pipette tips for PCR preparation; 200 µL, sterile, filter (Rainin, cat. no. RT-L200F)
- Cell scraping tools; sterile (Corning, cat. no. 353085)
- Disposable cell spreaders; sterile (Heathrow Scientific, cat. no. HS8151)
- Rattler Plating Beads, sterile (Zymo Research, cat. no. S1001)
- 96-well full skirt PCR microplates (Axygen, cat. no. PCR-96-FS-C)
- 8-cap PCR microplate sealing strips (Eppendorf, cat. no. 0030124847)
- Centrifuge tubes; 50 mL, sterile (Chemglass, cat. no. CLS-4302)
- Tough spots; 3/8 inch (Diversified Biotech, cat. no. T-SPOTS)
- Microcentrifuge tubes; low DNA binding, 2.0 mL (Eppendorf, cat. no. 022431048)

Equipment

- Colony picking robot (Norgren Systems, cat. no. CP7200 or equivalent)
- Shaking incubator; 3 mm orbit, 900 rpm maximum speed (Infors HT, Multitron Std, cat. no. I10003 or equivalent)
- Humidity control option for 3 mm orbit shaker (Infors HT, cat. no. I52000 or equivalent)
- Shaking incubator; 25 mm orbit (Infors HT, Multitron Std, cat. no. I10102 or equivalent)
- Microplate post kit for 3 mm orbit shaker; 7 inch (Infors HT)
- Microplate spacers for high-speed tight-orbit shaker incubator (Infors HT, cat. no. I20099 or equivalent)
- Tube rack for shaking incubator; 39 × 18 mm (Infors HT, ATR18, cat. no. 12305 or equivalent)
- PCR thermocycler; 96-well aluminum block (Eppendorf, Nexus Gradient cat. no. 6331000025 or equivalent)
- Blot replicator pin tool; 1.5 µL droplet (V&P Scientific, cat. no. VP 407 or equivalent)
- Library copying alignment tool (V&P Scientific, cat. no. VP 381 or equivalent)
- Blot replicator pin cleaning brush (V&P Scientific, cat. no. VP 425 or equivalent)
- 50 mL Erlenmeyer flask
- PCR capping aid tool (Eppendorf, cat. no. 951023108 or equivalent)
- Storage box; airtight, gasketed, 19 L (Sterilite, cat. no. 1932 or equivalent)
- Agarose gel electrophoresis box; wide format (Thermo Fisher Scientific, cat. no. D314 or equivalent)
- Agarose gel electrophoresis box; narrow format (Thermo Fisher Scientific, cat. no. B1A or equivalent)
- Power supply for gel electrophoresis (Thermo Fisher Scientific, cat. no. EC300XL2 or equivalent)
- 8-channel pipettor; 2-20 µL (Rainin, cat. no. L8-20XLS+ or equivalent)
- 96-channel pipettor; 5-200 µL (Rainin, cat. no. LIQ-96-200 or equivalent)
- Microcentrifuge (Eppendorf, cat. no. 5418 or equivalent)
- Infrared thermometer gun (Fluke, 62 Max Plus, cat. no. FLUKE-62MAX or equivalent)
- Ice bucket; 9 liter (Diversified Biotech, cat. no. IPAN-3000 or equivalent)
- Cooling block for PCR microplates (Eppendorf, cat. no. 022510541 or equivalent)
- Massively parallel sequencing device (Illumina, HiSeq 2500 or equivalent)
- BioAnalyzer or equivalent instrument
- Gel documentation system
- Scalpel (Elmers X-acto #2 or equivalent)
- Unix compatible computer

Reagent Setup

- 30 µg/mL Kanamycin sulfate

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

- 20% glycerol

Software

- BOWTIE2
- KOSUDOKU
  Available at https://github.com/buzbarstow/kosudoku/
- PYTHON
- IPYTHON
- MATPLOTLIB
- NUMPY
- SCIPY
- BLAST
- BIOPYTHON

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

List of Symbols and Abbreviations

Symbols

N_p	Minimum size of progenitor collection
r_{pick}	Colony picking rate
n_{dish}	Number of colonies that can be reliably identified on and picked from a Petri dish or colony picker tray.
n_{plate}	Number of colonies grown per 96-well plate. For example, if a checkerboard pattern is used, $n_{plate} = 48$.
r_{pool}	Pooling rate measured in plates per day.
$r_{incubate}$	Rate of incubation of picked colonies in liquid media.
N_{batch}	The maximum number of mutants colonies that can be processed in a single batch. Typically this is the smallest of N_{plate}, N_{pick} and N_{pool}.
N_{plate}	The maximum number of mutant colonies that can be plated in a single batch.
$N_{incubate}$	The maximum number of plated colonies that can be grown in a single batch.
N_{pick}	The maximum number of mutant colonies that can be picked in a single batch.
N_{pool}	The maximum number of mutant colonies that can be picked in a single batch.
ΔT_{solid}	Useful lifetime of plated colonies. Note that this is not the simply the time over which colonies can be used to inoculate liquid media, but the shorter time over which individual colonies can be discriminated by a robotic picking system and used to inoculate liquid media.
ΔT_{liquid}	Lifetime of cultures after reaching saturation in liquid media under optimal storage conditions.
ΔT_{pick}	Time needed to pick batch of colonies.
$\Delta T_{saturation}$	Time needed for picked colonies to reach saturating density in liquid media.

Abbreviations

YKO	Yeast Knock Out collection.
DAP	Diaminopimelic acid
kan	Kanamycin

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

Acknowledgments

We thank N. Ando, K. Davis, A. Palmer, S. Meisburger, B. Chang, E. Adler, K. Malzbender, and C. Kyauk for experimental assistance; W. Metcalf for providing *E. coli* strain WM3064; J. Gralnick for providing *Shewanella oneidensis* MR-1; L. Kovacs, J. Miller, L.R. Parsons, S. Silverman, W. Wang and J. Wiggins for assistance with next generation sequencing and media preparation; N. Ando for critical reading of this manuscript. This work was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and Princeton University startup funds (B.B.) and Fred Fox Class of 1939 funds (I.A.).

Figures

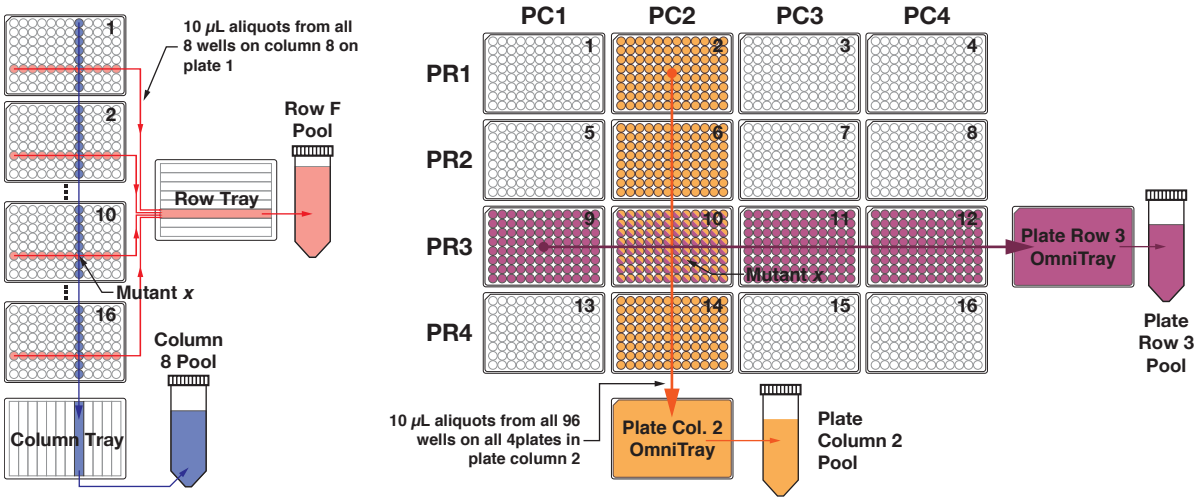

Figure 1: Combinatorial pooling scheme for Knockout Sudoku.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

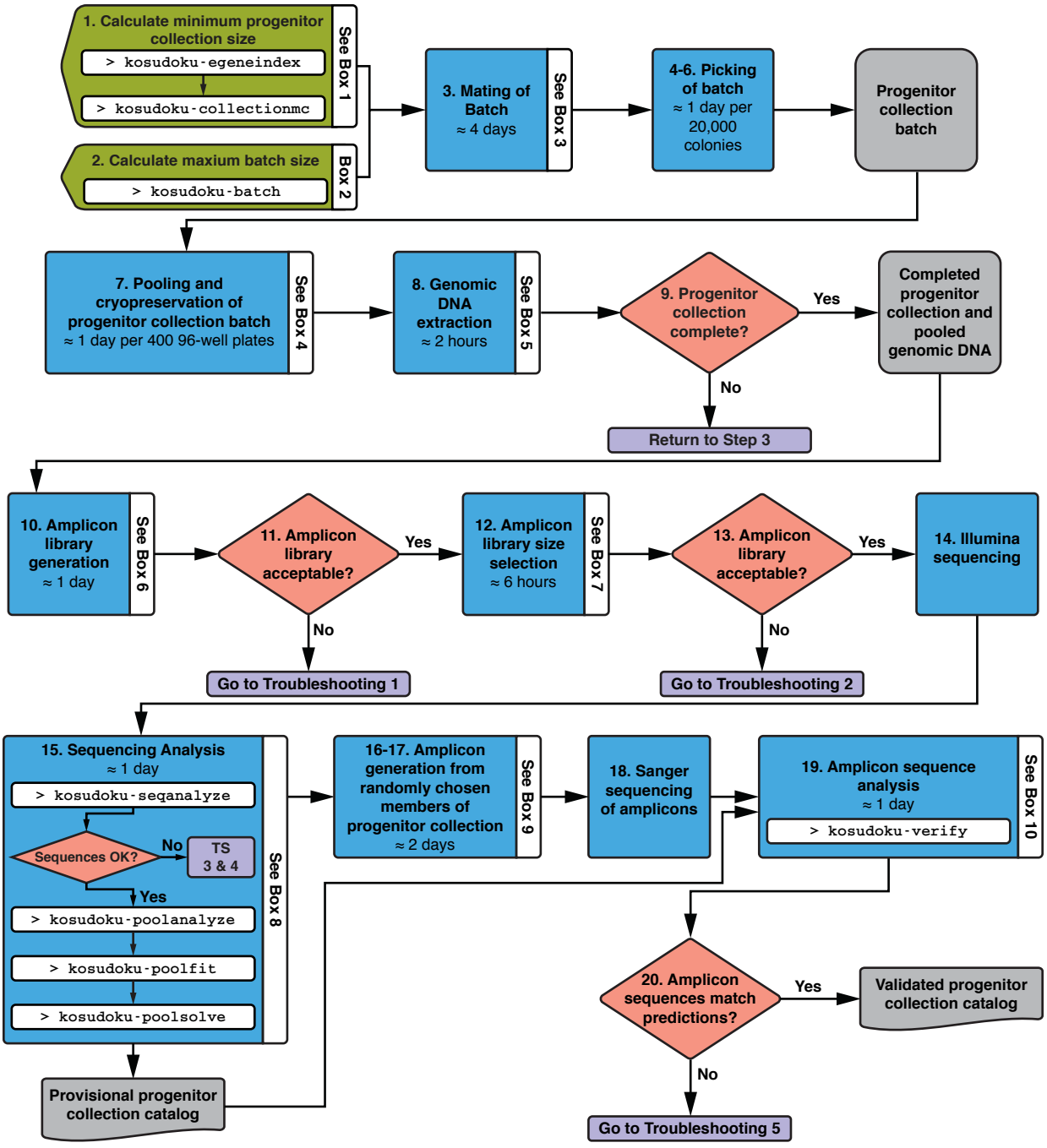

Figure 2: Workflow for creation and annotation of progenitor collection. Note in box 15, TS = Troubleshooting.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

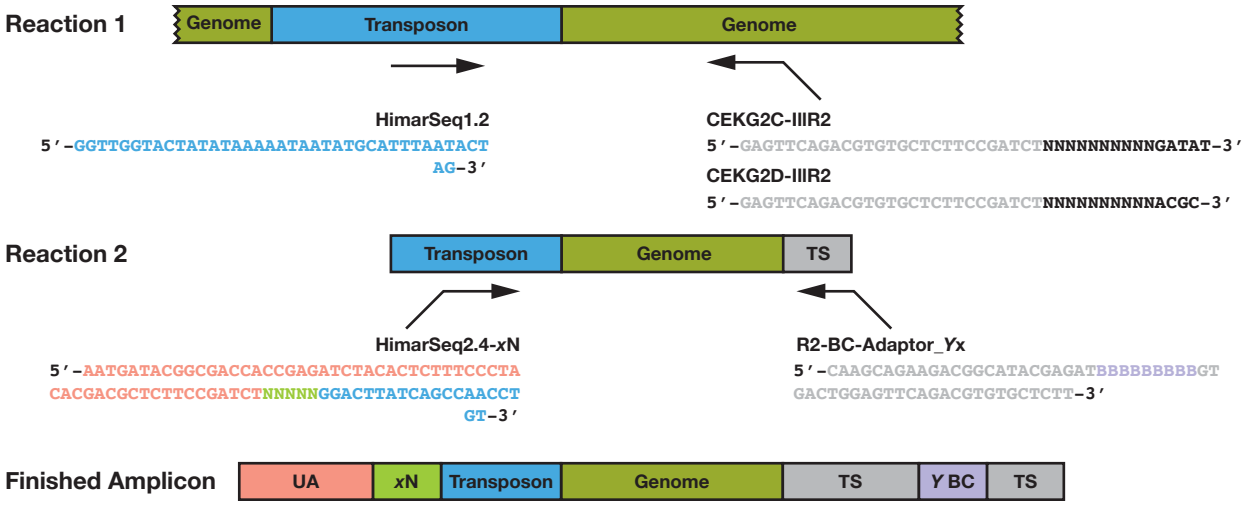

Figure 3: Detailed schematic of pool amplicon library generation. Primer sequences are color coded to match the block diagrams shown above them. TS (grey) = the Illumina TruSeq adaptor sequence, UA (coral red) = Illumina Universal Adaptor sequence, *x*N (bright green) = a 4-7 bp random sequence added to aid avoid saturation of any color channel while sequencing the transposon sequence shared by all amplicons, and *Y* BC (lavender) = an 8 bp barcode sequence unique to each pool.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

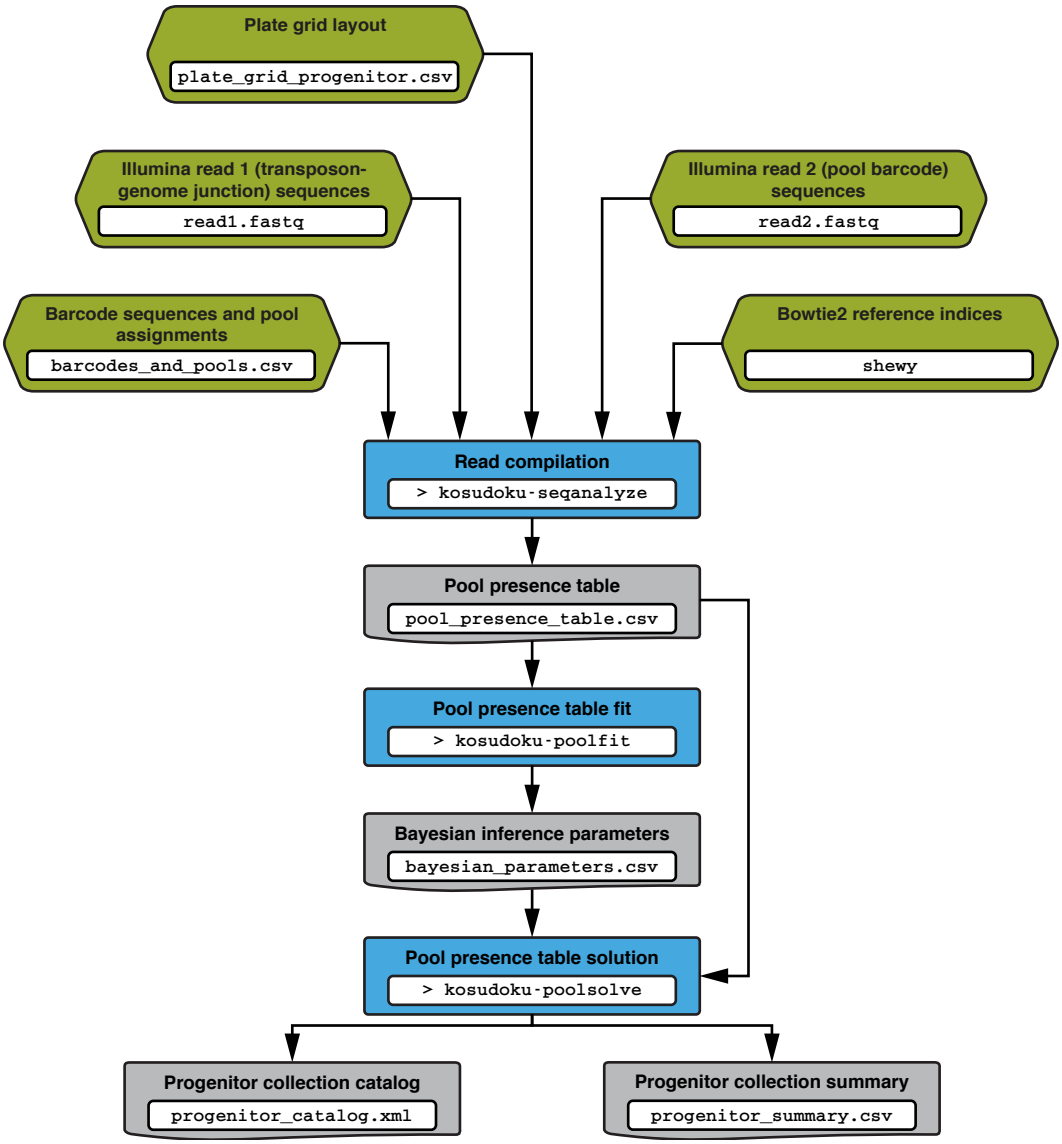

Figure 4: Software workflow and important input and output files for progenitor collection catalog solution (**Box 8**).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

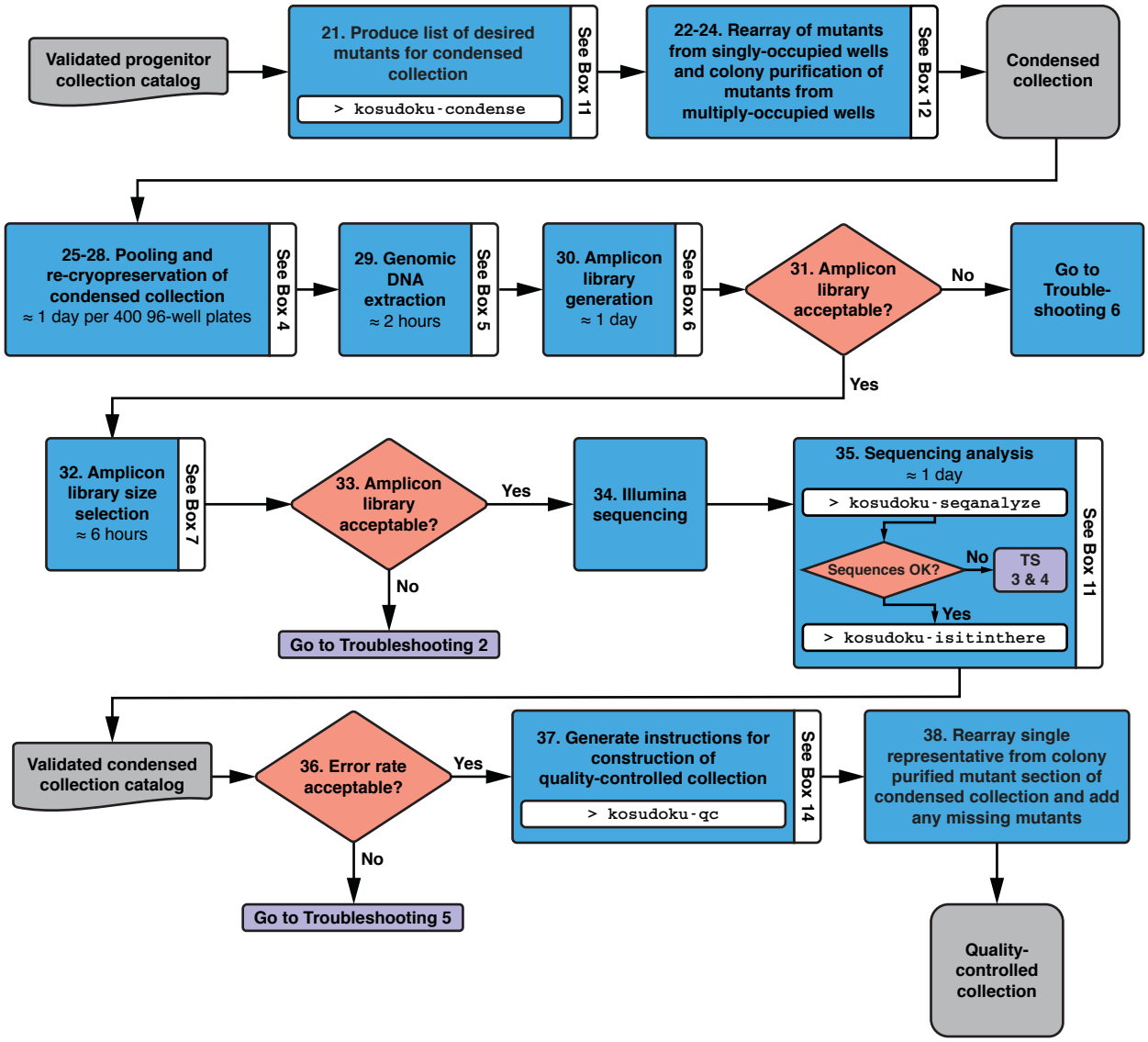

Figure 5: Workflow for creation of condensed and quality-controlled collections. Note in box 35, TS = Troubleshooting.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

References

1. van Opijnen, T. & Camilli, A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Micro* **11,** 435–442 (2013).
2. Hutchison, C. A. *et al*. Design and synthesis of a minimal bacterial genome. *Science* **351,** aad6253–aad6253 (2016).
3. Baym, M., Shaket, L., Anzai, I. A., Adesina, O. & Barstow, B. *Rapid Construction of a Whole-genome Transposon Insertion Collection for Shewanella oneidensis by Knockout Sudoku*. *biorxiv.org* (2016). doi:10.1101/044768
4. Richie, D. L. *et al*. Identification and evaluation of novel acetolactate synthase inhibitors as antifungal agents. *Antimicrob. Agents Chemother.* **57,** 2272–2280 (2013).
5. Brotcke, A. & Monack, D. M. Identification of fevR, a novel regulator of virulence gene expression in Francisella novicida. *Infection and Immunity* **76,** 3473–3480 (2008).
6. Ahlund, M. K., Rydén, P., Sjöstedt, A. & Stöven, S. Directed screen of Francisella novicida virulence determinants using Drosophila melanogaster. *Infection and Immunity* **78,** 3118–3128 (2010).
7. Fey, P. D. *et al*. A genetic resource for rapid and comprehensive phenotype screening of nonessential Staphylococcus aureus genes. *MBio* **4,** e00537–12 (2013).
8. Zerikly, M. & Challis, G. L. Strategies for the Discovery of New Natural Products by Genome Mining. *Chembiochem* **10,** 625–633 (2009).
9. Seyedsayamdost, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proceedings of the National Academy of Sciences* **111,** 7266–7271 (2014).
10. Winzeler, E. A. *et al*. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).
11. Baba, T. *et al*. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2,** 2006.0008 (2006).
12. Giaever, G. & Nislow, C. The yeast deletion collection: a decade of functional genomics. *Genetics* **197,** 451–465 (2014).
13. van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6,** 767–772 (2009).
14. Erlich, Y. *et al*. DNA Sudoku--harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Research* **19,** 1243–1253 (2009).
15. Goodman, A. L., Wu, M. & Gordon, J. I. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature Protocols* **6,** 1969–1980 (2011).
16. Goodman, A. L. *et al*. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proceedings of the National Academy of Sciences* **108,** 6252–6257 (2011).
17. Gallagher, L. A. *et al*. Sequence-Defined Transposon Mutant Library of Burkholderia thailandensis. *MBio* **4,** e00604–13–e00604–13 (2013).
18. Porwollik, S. *et al*. Defined Single-Gene and Multi-Gene Deletion Mutant Collections in Salmonella enterica sv Typhimurium. *PLoS ONE* **9,** e99820 (2014).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016

19. Gallagher, L. A. *et al*. Resources for Genetic and Genomic Analysis of Emerging Pathogen Acinetobacter baumannii. *J Bacteriol* **197,** 2027–2035 (2015).

20. Vandewalle, K. *et al*. Characterization of genome-wide ordered sequence-tagged Mycobacterium mutant libraries by Cartesian Pooling-Coordinate Sequencing. *Nat Commun* **6,** 7106– (2015).

21. Bouhenni, R., Gehrke, A. & Saffarini, D. Identification of genes involved in cytochrome c biogenesis in Shewanella oneidensis, using a modified mariner transposon. *Applied and Environmental Microbiology* **71,** 4935–4937 (2005).

22. Melnyk, R. A., Clark, I. C., Liao, A. & Coates, J. D. Transposon and Deletion Mutagenesis of Genes Involved in Perchlorate Reduction in Azospira suillum PS. *MBio* **5,** e00769–13–e00769–13 (2013).

23. Zou, L. *et al*. SlyA Regulates Type III Secretion System (T3SS) Genes in Parallel with the T3SS Master Regulator HrpL in Dickeya dadantii 3937. *Applied and Environmental Microbiology* **78,** 2888–2895 (2012).

24. Maier, T. M., Pechous, R., Casey, M., Zahrt, T. C. & Frank, D. W. In Vivo Himar1-Based Transposon Mutagenesis of Francisella tularensis. *Applied and Environmental Microbiology* **72,** 1878–1885 (2006).

25. Rollefson, J. B., Levar, C. E. & Bond, D. R. Identification of genes involved in biofilm formation and respiration via mini-Himar transposon mutagenesis of Geobacter sulfurreducens. *J Bacteriol* **191,** 4207–4217 (2009).

26. Bonis, B. M. & Gralnick, J. A. Marinobacter subterrani, a genetically tractable neutrophilic Fe(II)-oxidizing strain isolated from the Soudan Iron Mine. *Front Microbiol* **6,** 719 (2015).

27. Yu, R. & Kaiser, D. Gliding motility and polarized slime secretion. *Mol Microbiol* **63,** 454–467 (2007).

28. Dey, A. & Wall, D. A genetic screen in Myxococcus xanthus identifies mutants that uncouple outer membrane exchange from a downstream cellular response. *J Bacteriol* **196,** 4324–4332 (2014).

29. Zhu, L.-P. *et al*. Allopatric integrations selectively change host transcriptomes, leading to varied expression efficiencies of exotic genes in Myxococcus xanthus. *Microb Cell Fact* **14,** 53 (2015).

30. Barquist, L., Boinett, C. J. & Cain, A. K. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biology* **10,** 1161–1169 (2014).

31. Kwon, Y. M., Ricke, S. C. & Mandal, R. K. Transposon sequencing: methods and expanding applications. *Appl Microbiol Biotechnol* (2015). doi:10.1007/s00253-015-7037-8

32. Chun, K. T., Edenberg, H. J., Kelley, M. R. & Goebl, M. G. Rapid amplification of uncharacterized transposon-tagged DNA sequences from genomic DNA. *Yeast* **13,** 233–240 (1997).

33. Manoil, C. Tagging exported proteins using Escherichia coli alkaline phosphatase gene fusions. *Meth Enzymol* **326,** 35–47 (2000).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2294v1 | CC BY 4.0 Open Access | rec: 18 Jul 2016, publ: 18 Jul 2016