ACS SyntheticBiology

Viewpoint

pubs.acs.org/synthbio

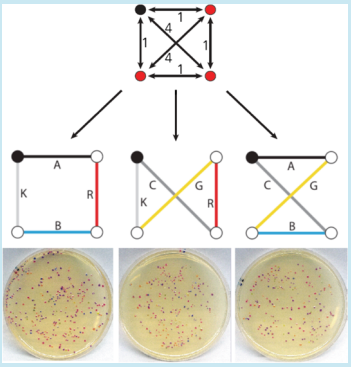
Solving a Four-Destination Traveling Salesman Problem Using *Escherichia coli* Cells As Biocomputers

Michael Esau,[†] Mark Rozema,[†] Tuo Huang Zhang,[†] Dawson Zeng,[†] Stephanie Chiu,[†] Rachel Kwan,[†] Cadence Moorhouse,[†] Cameron Murray,[†] Nien-Tsu Tseng,[†] Doug Ridgway,[‡] Dominic Sauvageau,*[,†] and Michael Ellison*[,†,‡]

[†]University of Alberta, Edmonton, Alberta T6G 2R3, Canada

[‡]Genomikon Inc., Edmonton, Alberta T6G 2H7, Canada

ABSTRACT: The Traveling Salesman Problem involves finding the shortest possible route visiting all destinations on a map only once before returning to the point of origin. The present study demonstrates a strategy for solving Traveling Salesman Problems using modified *E. coli* cells as processors for massively parallel computing. Sequential, combinatorial DNA assembly was used to generate routes, in the form of plasmids made up of marker genes, each representing a path between destinations, and short connecting linkers, each representing a given destination. Upon growth of the population of modified *E. coli*, phenotypic selection was used to eliminate invalid routes, and statistical analysis was performed to successfully identify the optimal solution. The strategy was successfully employed to solve a four-destination test problem.



The Traveling Salesman Problem (TSP) is a challenging computational problem in which, given the position of a set of destinations in a network and the distance separating them, one must solve for the shortest possible route that visits every destination exactly once and returns to the starting point. To ensure that a given route is the shortest, one must use the brute force method, that is, evaluate and then compare all possible routes. This can prove to be troublesome since as the number of destinations increases, the total number of possible routes increases geometrically according to the equation

$$N = \frac{(n-1)!}{2} \quad (1)$$

where N is the number of possible routes and n is the number of destinations in the network. This, in turn, translates into geometrically increasing computational requirements: a hypothetical computer capable of comparing 10 000 routes every second would take 3.6 days to solve a problem with 14 destinations (3.1 billion possible routes), 50 days to solve one with 15 destinations (43.6 billion possible routes), and over 2 years to solve for 16 destinations (654 billion possible routes).

One approach to this problem is using massively parallel computation, which significantly reduces the computational time required to identify the optimal route. In this case, the scale of the problem is limited by the number of processors available for computation (e.g., two computers, each assigned to half the possible routes, would take half the time to calculate all routes as a single computer). In principle, a biological system, in which a great many biomolecules could serve as processors, could massively expand the number of routes evaluated in a given time frame, allowing for more rapid solutions for TSPs and other optimization problems. This concept was pioneered by Adleman,[1] who used DNA amplification to solve a seven-point Hamiltonian Path Problem—a problem closely related to the TSP that involves proving that at least one continuous path (though not necessarily the optimal one) exists in a given network. More recently, an *E. coli* strain capable of solving four-point Hamiltonian paths was developed by the Davidson−Missouri team for the iGEM 2008 competition.[2]

We present, here, the first TSP-solving biocomputer of its kind. The method relies on the construction and transformation of a plasmid library into an *E. coli* population, where each transformed cell represents a potential route. "Routes" are assembled in the form of plasmids that contain a set of color genes that form a valid TSP route. Each gene within the plasmid corresponds to a specific path that connects a pair of destinations in the given network of destinations. Each destination is represented by a set of short-chain DNA linkers, capable of linking two specific paths, represented by colors or antibiotic resistance. Each complete plasmid thus expresses a specific color and antibiotic resistance combination that corresponds to a specific route (Figures 1 and 2A). The assembly process is biased to favor production of plasmids representing short routes so that the plasmid library includes proportionally more instances of short-route plasmids as compared to long-route plasmids. Upon growth of the

Special Issue: iGEM 2013

Received: January 17, 2014
Published: December 19, 2014

   dx.doi.org/10.1021/sb5000466 | *ACS Synth. Biol.* 2014, 3, 972−975

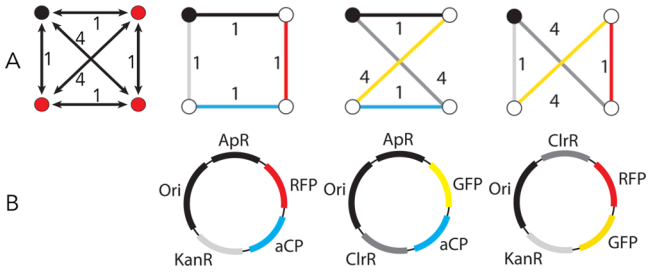


Figure 1. (A) Four-destination traveling salesman problem with path distance values and possible valid routes. The total route distances are 4, 10, and 10 distance units, respectively. (B) Plasmids corresponding to the valid routes, with matching colors indicating the gene assignments: *AmpR* in black, *ChlorR* in dark gray, *KanR* in light gray, *GFP* in yellow, *RFP* in red, and *aCP* in blue.

population, the optimal route can then be defined by the most frequently occurring phenotype. The validity of the method was demonstrated by accurately solving a four-destination TSP.

METHODS

Organism, Cultures, and Plating. *E. coli* DH5α. The bacteria was grown in 2.5% LB broth (BD, Mississauga, Canada) or LB broth supplemented with one or a combination of the following antibiotics: 100 mg/L ampicillin (Sigma-Aldrich), 25.5 mg/L chloramphenicol (Sigma-Aldrich), or 50 mg/L kanamycin (Sigma-Aldrich). Cultures were incubated overnight at 37 °C in an incubator-shaker (Innova 4300, New Brunswick Scientific). Aliquots of cultures were preserved at −80 °C in LB medium supplemented with 8% dimethyl sulfoxide.

Plating was performed by spreading cultures diluted 1:20 in fresh medium on LB agar plates (1.5% agar, BD, Mississauga, Canada) with or without antibiotics. Plates were incubated for a minimum of 24 h at 37 °C in an incubator (Model 3025, VWR Scientific).

Plasmid Library Assembly. Plasmid assembly was performed through a process of sequential, combinatorial DNA ligations based on the Genomikon technology[3] (Figure 2A). 30 μL of oligo $d(T)_{25}$ magnetic beads (New England Biolabs) were suspended with 0.3 pmoles of poly-A-tailed Origin of Replication DNA in 15 μL of wash buffer (20 mM Tris, 2 mM EDTA, 200 mM NaCl, pH 8.0) and incubated at 22 °C for 15 min to bind the DNA to the beads. The beads were then extracted and washed with 100 μL of wash buffer to remove the excess unbound DNA. The first ligation was performed by resuspending the beads in 20 μL of T4 DNA ligase buffer (New England Biolabs) with T4 DNA ligase enzyme (New England Biolabs) with a combination of "path" genes and incubating this suspension for 30 min at 4 °C. The beads were then extracted and washed with 100 μL of wash buffer in order to remove the excess nonligated DNA. Following this first gene ligation, the magnetic bead was placed in a second ligation reaction mixture, containing a short 25-base-pair linker DNA strand for ligation. This was followed by similar steps for ligation alternating between genes (paths) and linkers (destinations) until a complete solution was built (Figure 2A and B). The beads were then resuspended in 20 μL of 10 mM NaOH for 1 min at room temperature to elute the DNA from the beads, which were then extracted from the suspension. 2 μL of concentrated Tris-EDTA buffer (500 mM Tris, 50 mM EDTA) was then added to the eluted DNA solution to neutralize the pH for storage.

By introducing a mixture of genes in each gene-ligation step, an ordered library of plasmids that included every possible combination of paths was built. In order to account for the lengths of the different paths in a network, the following

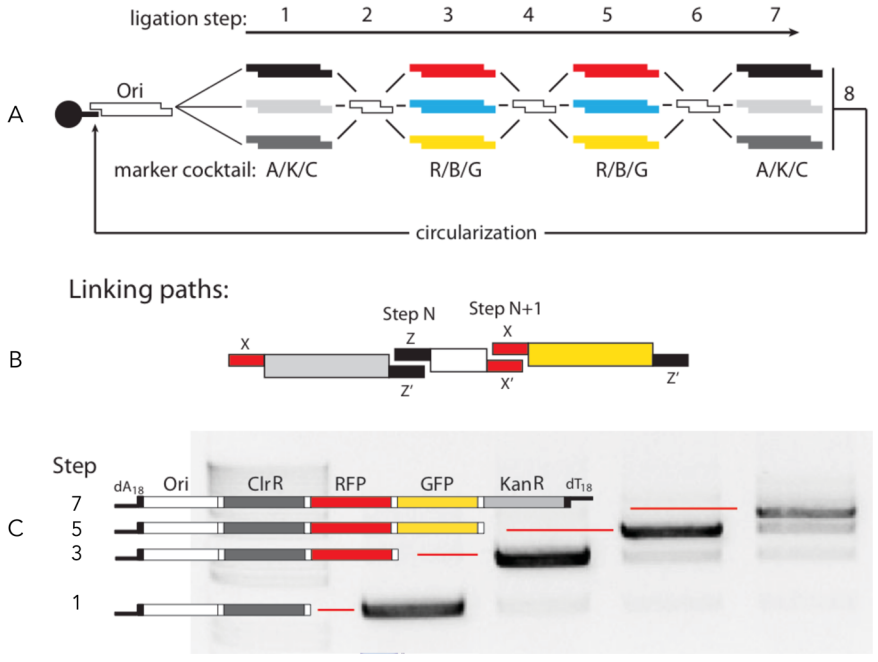


Figure 2. (A) Schematic for Genomikon assembly process. Nonpalindromic sticky ends are used to sequentially ligate combinations of gene strands alternating with short linkers to a magnetic-bead-anchored origin of replication. (B) Separate ligation reactions and unique sticky ends ensure that only one correctly oriented strand can be ligated at each step. (C) Agarose gel electrophoresis showing the progress of a sequential assembly. Lanes 1−4 correspond to samples taken after ligation steps 1, 3, 5, and 7, respectively.

dx.doi.org/10.1021/sb5000466 | *ACS Synth. Biol.* 2014, 3, 972−975

relationship was developed to determine the required concentration of gene *i* for ligation:

$$[C_i] = \alpha_i \frac{1}{d_i} \quad (2)$$

where $[C_i]$ is the concentration of gene *i* used in the reaction, d_i is the length of the corresponding path, and α_i is a calibration coefficient used to compensate for any inherent biological bias associated with the expression or ligation kinetics of gene *i*. These coefficients were based on the relative quantity of colonies expressing each gene when they were ligated to the origin at equal molar concentration, then transformed and plated. If *d* equals 1 for all paths, the adjustment from α_i results in equal quantities of resulting colonies upon plating. Calibration for gene concentrations resulted in the following α_i values used in assembly: *ChlorR* = 0.6 pmol, *GFP* = 0.6 pmol, *KanR* = 0.4 pmol, *RFP* = 0.4 pmol, *AmpR* = 0.3 pmol, *aCP* = 0.1 pmol. Using eq 2 allowed genes that represented short paths to be preferentially incorporated into the products of the assembly process. The final library of plasmids thus created featured more individual plasmids representing the optimal route of the network.

As an example, the path represented by ampicillin resistance in Figure 1A was 1 distance unit long, while the path represented by chloramphenicol resistance was 4 distance units long. As a result, in the first ligation step, the concentration of the *AmpR* gene was equal to $\alpha_{ampR} \cdot 1 = 0.3$ pmol while the concentration of the *ChlorR* gene was equal to $\alpha_{clrR} \cdot {}^1/_4 = 0.15$ pmol, such that *AmpR* was more likely to be incorporated into the product plasmid than *ChlorR*. Thus, a greater proportion of the plasmid library would be made up of ampicillin-containing plasmids.

Solution Analysis. To analyze the relative proportions of each of the unique routes assembled and determine the most abundant one, the product plasmids were used to transform a culture of chemically competent *E. coli*. Plasmids and culture were combined and incubated on ice for 30 min, then heated in a water bath at 42 °C for 20 s, returned to ice for 2 min, diluted 1:20 with prewarmed LB, and finally incubated at 37 °C for 1 h. The transformed culture was grown on three varieties of agar plates, each treated with two antibiotics (Amp/Clr, Amp/Kan, or Clr/Kan). Complete solutions were validated for colonies exhibiting phenotypes that were resistant to two antibiotics and that expressed a combination of two colors (RFP and GFP producing orange, GFP and aCP producing green, or RFP and aCP producing purple). The instances of each valid phenotype were tallied—omitting colonies whose phenotype indicated an incomplete or invalid solution (single antibiotic resistance or single color). The most frequent phenotype corresponded to the most abundant plasmid, and thus, the optimal route in the test network.

RESULTS AND DISCUSSION

Four-City TSP. As proof-of-concept, a test network consisting of four destinations arranged in a square was used (Figure 1A). This TSP had six paths connecting each possible pairing of destinations and three unique, valid routes consisting of four paths each (Figure 1A). The paths around the perimeter of the square were all 1 arbitrary distance unit, while the two paths crossing in the interior were 4 distance units each. Hence, the resulting distances are 4 units for the shortest route and 10 units for the other two routes.

Translating TSP Solutions into Plasmids. For the test network, each of the six possible paths was assigned a corresponding gene for either antibiotic resistance (ampicillin (*AmpR* or A), chloramphenicol (*ClrR* or C), or kanamycin (*KanR* or K)) or a pigment (*GFP* (G), *RFP* (R), or *aCP* (B)). The genes were assigned such that a valid route would consist of a plasmid containing two genes for antibiotic resistance and two pigments (Figure 1B) and could be identified by the phenotypic expression it induced in bacteria. For example, a possible solution plasmid would begin with an origin-of-replication strand (home city), followed by the gene *AmpR* and a linker (path to destination 1), the gene *RFP* and a linker (path to destination 2), the gene *aCP* and a linker (path to destination 3), and the gene *KanR* and a final linker (return path to home city, the final destination). The resulting phenotype would be a colony resistant to ampicilin and kanamycin with a purple color.

Plasmid Assembly. A strategy for the sequential assembly of a plasmid library on solid support that contains all solutions, is shown in Figure 2A and is described in Methods. For any given solution, seven fragments were required: an origin of replication, two antibiotic resistance genes, two color genes, and two linker sequences. The combinatorial nature of the library was created by adding a cocktail of the three resistance markers at steps 1 and 7, and a cocktail of the three color genes at steps 3 and 5.

Figure 2C shows the results of a seven-fragment test assembly for a single plasmid solution (*ClrR-GFP-RFP-KanR*). Gel electrophoresis illustrates that the major product is in fact the full-length product. The efficiency of assembly is further illustrated by the distribution of colored colonies on plates containing both chloramphenicol and kanamycin (Figure 3). Approximately 90% (or 721) of the resulting colonies produced

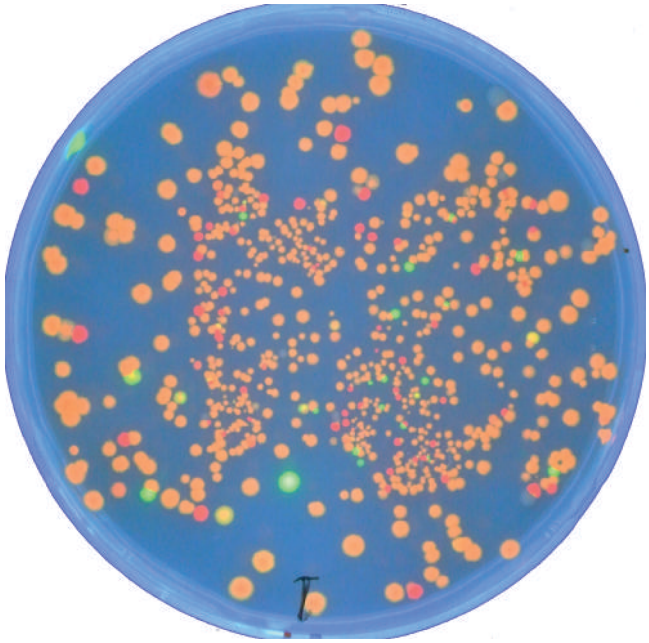


Figure 3. Photograph of UV exposed trial assembly plate treated with chloramphenicol and kanamycin. Orange-fluorescing colonies are expressing RFP and GFP, indicating successful assembly of a plasmid containing RFP, GFP, ChlorR, and KanR genes. Red-fluorescing and green-fluorescing colonies indicate incomplete plasmid assemblies, missing either GFP or RFP, respectively.

dx.doi.org/10.1021/sb5000466 | *ACS Synth. Biol.* 2014, 3, 972−975

the orange phenotype expected of a complete route (GFP +RFP). Whereas 69 colonies produced the red or green phenotype expected of an incomplete route.

We note that the ability to produce final product (or solutions) in high yield, coupled with the relatively short assembly time, translates into a minimal cost of complexity on computation. As the number of destinations is increased between problems, the number of potential solutions increases geometrically according to eq 1 but assembly time increases linearly, while growth and expression time are unaffected. This, added to massively parallel computing, is an important advantage for this type of biocomputing.

In the case of the four-destination TSP, the assembly was performed as shown in Figure 2A using the gene combinations indicated at each step. Following transformation, cells were plated on each of the antibiotic pairs Kan/Amp, Kan/Clr and Amp/Clr. Valid routes were then determined by scoring colony resistance and color (two antibiotic resistance genes and two pigment genes) (Figure 4). Sixty-one colonies presenting a valid phenotype were counted for the shortest route (purple colonies on left plate of Figure 4C), while 21 and 18 colonies were counted for the two lengthier routes (orange and green colonies on the middle and right plates of Figure 4C, respectively).

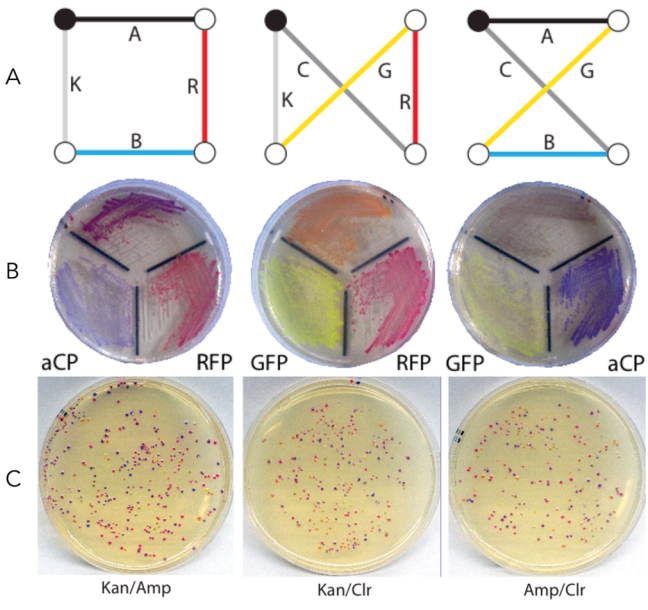


Figure 4. (A) Three valid, color-coded routes with genes/paths labeled A for *AmpR*, K for *KanR*, C for *ClrR*, G for *GFP*, R for *RFP*, and B for *aCP*. A colony on the plates in row C that is expressing four genes that match one of these routes is tallied toward that route to determine the most abundant phenotype (and, thus, the shortest route). Colonies with fewer than two antibiotic resistances cannot grow on the plates, and colonies with fewer than two pigment genes do not represent valid routes for solving a TSP and so are not counted. (B) Color standards for the various color gene combinations. The topmost segment of each plate displays a secondary color produced when the colony expressed both of the colors displayed in the lower two segments. (C) The product plates from the final trial. Each plate was treated with two antibiotics identified below it. Colonies on each plate expressed an array of all possible colors. Only phenotypes matching the three routes of row A were counted. The tallies were as follows: 61 purple Kan/Amp resistant colonies, 21 orange Kan/Clr resistant colonies, and 18 green Amp/Clr resistant colonies.

The percentage of colonies corresponding to the two nonoptimal, valid routes (*ClrR/KanR/RFP/GFP* and *ClrR/AmpR/aCP/GFP*) represented 8% of the colonies on their respective plates (middle and right plates, Figure 4C). This was very close to the theoretical prediction of 7% based on the distances for this TSP. The colonies corresponding to the optimal route amounted to 17% of the colonies on the Kan/Amp plate (Figure 4), lower than the 28% predicted. This discrepancy can explain the fact that the ratio of colonies representing the optimal solution to each nonoptimal valid solution (2:1) was slightly lower than that predicted by statistics (3:1). Nevertheless, this is a clear demonstration that the biocomputing method developed has the capacity to correctly determine the optimal solution to a TSP.

In principle, any four-destination TSP could be solved using the same system by replacing the distance values and correcting for the gene concentrations required during assembly. The system can also be expanded to solve a more complex TSP using additional unique marker genes, although expanding beyond six destinations would require direct plasmid sequencing analysis rather than phenotype identification to identify the predominant route. This is because at seven or more destinations, the quantity of possible phenotypes (upward of 360) is too large for rapid differentiation and tallying. Direct plasmid sequencing would allow for the use of arbitrary (noncoding) DNA sequences to signify the network paths, further increasing the computational capacity of the method. With additional improvements to the algorithm, assembly, and transformation protocols, the hypothetical upper complexity limit for a 3-day biocomputational method would be 17 destinations using auxotrophic strains.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: dominic.sauvageau@ualberta.ca.

*E-mail: mike.ellison@ualberta.ca.

Author Contributions

M.E. designed and performed experiments and wrote the manuscript. M.R., T.H.Z., D.Z., S.C., R.K., C.M., C.M., and N.T.T. performed experiments. D.R. designed experiments and developed the concentration algorithm. D.S. and M.E. designed and supervised experiments and wrote and revised the manuscript. All authors read and approved the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the Alberta Innovates—Technology Futures, the University of Alberta and Genome Alberta for financial support.

REFERENCES

(1) Adleman, L. M. (1994) Molecular computation of solutions to combinatorial problems. *Science. 266*, 1021−1024.

(2) iGEM 2007 Team Davidson−Missouri: Hamiltonian Pathfinders. 2007. igem.org/wiki/index.php/Davidson_Missouri_W (accessed Nov. 18, 2014).

(3) Genomikon. www.genomikon.ca (accessed Nov. 18, 2014).