Visualizing the tape of life: exploring evolutionary history with virtual reality

Emily Dolson
Program in Ecology, Evolutionary Biology, and Behavior
Department of Computer Science and Engineering
BEACON Center for the Study of Evolution in Action
East Lansing, Michigan
dolsonem@msu.edu

Charles Ofria
Program in Ecology, Evolutionary Biology, and Behavior
Department of Computer Science and Engineering
BEACON Center for the Study of Evolution in Action
East Lansing, Michigan
ofria@msu.edu

ABSTRACT

Understanding the evolutionary dynamics created by a given evolutionary algorithm is a critical step in determining which ones are most likely to produce desirable outcomes for a given problem. While it is relatively easy to come up with hypotheses that could plausibly explain observed evolutionary outcomes, we often fail to take the next step of confirming that our proposed mechanism accurately describes the underlying evolutionary dynamics. Visualization is a powerful tool for exploring evolutionary history as it actually played out. We can create visualizations that summarize the evolutionary history of a population or group of populations by drawing representative lineages on top of the fitness landscape being traversed. This approach integrates information about the adaptations that took place with information about the evolutionary pressures they were being subjected to as they evolved. However, these visualizations can be challenging to depict on a two-dimensional surface, as they integrate multiple forms of three-dimensional (or more) data. Here, we propose an alternative: taking advantage of recent advances in virtual reality to view evolutionary history in three dimensions. This technique produces an intuitive and detailed illustration of evolutionary processes. A demo of our visualization is available here: https://emilydolson.github.io/fitness_landscape_visualizations.

CCS CONCEPTS

• **Computing methodologies → Genetic algorithms**; **Scientific visualization**; **Visual analytics**;

KEYWORDS

Evolutionary dynamics, evolutionary history, virtual reality, data visualization

ACM Reference Format:
Emily Dolson and Charles Ofria. 2018. Visualizing the tape of life: exploring evolutionary history with virtual reality. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3205651.3208301

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3208301

1 INTRODUCTION

What properties of a problem cause different evolutionary algorithms to be more or less effective at solving it? If we can answer this question, we can begin to formulate a more general theory of evolutionary computation. Data visualizations are a powerful tool for developing an intuition for how to address these issues. Specifically, visualizing how populations traverse different types of search spaces under different selection regimes will provide insight into the underlying evolutionary dynamics. These results will clarify how algorithm settings interact with properties of a search space to influence the quality of solutions ultimately found. The challenge is finding a way to meaningfully summarize such a large quantity of information into a comprehensible visualization.

Attributes of a problem that may make it a better or worse target for a given algorithm can all be encapsulated as properties of that problem's fitness landscape. A fitness landscape can be thought of as a map from solution representations into the quality of that solution (its fitness) [19]. Traditionally, fitness landscapes have most commonly been visualized as three-dimensional surface plots illustrating the effects of two continuously-varying traits on fitness (e.g. [7], or Figure 1 of this paper). One of these traits is the x-axis, one is the y-axis, and the fitness conferred by that combination of values is the z-axis. Although we will focus here on fitness landscapes that can be fully depicted in three dimensions, it is important to recognize that most real-world fitness landscapes have far more dimensions and attempting to reduce them to three dimensions may be misleading [4]. In the future, we believe extensions of the approach presented here will also facilitate the visualization of higher-dimensional fitness landscapes.

In order to understand how a particular evolutionary algorithm performs on a given problem, we can observe how that algorithm moves a population across the problem's fitness landscape [6]. The most critical of this information is encapsulated by the lineages (i.e. chains of ancestors) of the population at the end of evolution, or the phylogeny (i.e. family tree) of the population as a whole. Indeed, these pieces of information have been the targets of a wealth of valuable visualization tools [2, 9, 14]. Overlaying information about an evolving population on top of a fitness landscape visualization has also already proven to be a powerful tool for understanding evolutionary dynamics on that landscape [12, 17]. However, combining these approaches is more challenging. When the entire visualization ultimately needs to be depicted in two dimensions (e.g. on a page or screen), overlaying information on a three-dimensional fitness landscape surface is only viable for simple fitness landscapes; as the ruggedness of the landscape increases, it becomes impossible to see around all of the corners.

Fortunately, we are entering an era where we no longer need to squeeze this information into two dimensions, confined to a computer screen or piece of paper. Over the past few years, virtual reality technology has advanced to the point where it is a viable tool for data visualization. Software libraries for rendering data visualizations in virtual reality have become more accessible, as have devices for viewing them. These advances open up exciting opportunities to create richer, more informative data visualizations [11, 16]. However, there is a lot to learn about how to effectively make use of these new capabilities. By taking advantage of these recent advances, we can immerse ourselves in a three-dimensional representation of our data. From this perspective, we can not only look around corners but also use depth perception to more accurately perceive the landscape.

In this paper, we present a proof-of-concept for using virtual reality to gain insight into how different selection schemes allow populations to traverse different fitness landscapes. For simplicity, we will use real-valued function optimization problems with two inputs, as such problems create fitness landscapes that can be directly visualized in three dimensions. In the future, we plan to expand this work to higher-dimensional fitness landscapes.

2 THE VISUALIZATION

The base component of our visualization is a three-dimensional surface plot of the fitness landscape (see Figure 1). In the case of the proof-of-concept presented here, we used two-input real-valued function optimization problems from the 2018 GECCO Niching Contest [7]. The fitness landscapes for these problems are the three-dimensional surfaces described by each of these functions. Our candidate solutions are pairs of floating point values: an x coordinate and a y coordinate. Setting the z-coordinate to the value of the function for those values of x and y (i.e. fitness), we can plot any candidate solution on the surface of the fitness landscape.

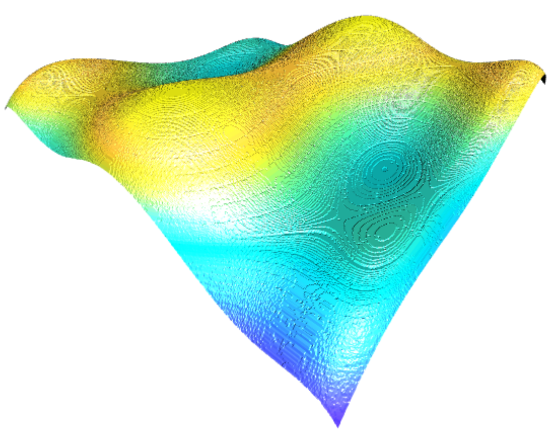

Figure 1: The base layer of our visualization: a surface plot of the fitness function. Here, we use the six-humped camel back function as our example fitness landscape.

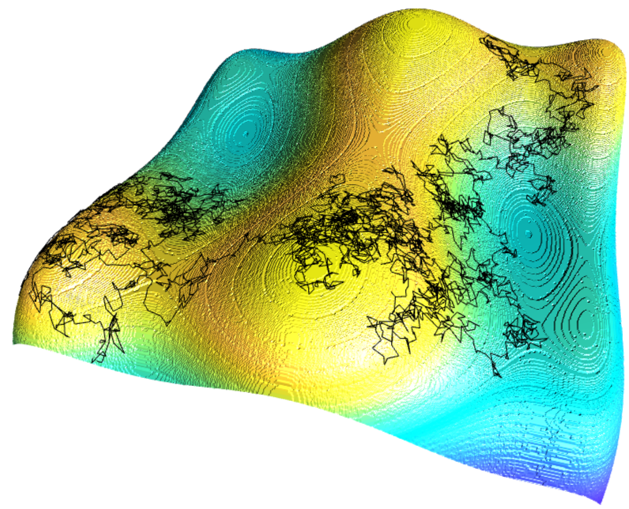

Figure 2: A single lineage drawn on top of the fitness landscape. From this visualization, we can see how the lineage crossed the landscape over evolutionary time. It is not obvious which end is the beginning, but the overall pattern is interpretable.

The next question is what information to draw on top of the fitness landscape. Since we are interested in understanding how lineages traverse the fitness landscape, the obvious approach is to draw a path from a candidate solution that existed at the end of an evolutionary run all the way back to its earliest ancestor, passing through the locations of all intermediate ancestors along the way (see Figure 2). Unfortunately, while this straightforward approach can often work, it also frequently creates an indecipherable mess on the fitness landscape as a lineage jumps around a local area, limiting its usefulness (see Figure 3).

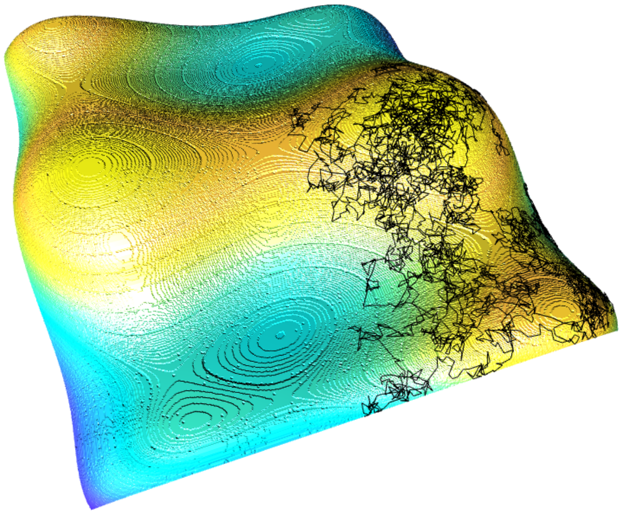

Figure 3: In this case, the lineage stayed in the same region of the fitness landscape over its entire evolutionary history. As a result, it is hard to make sense of the path.

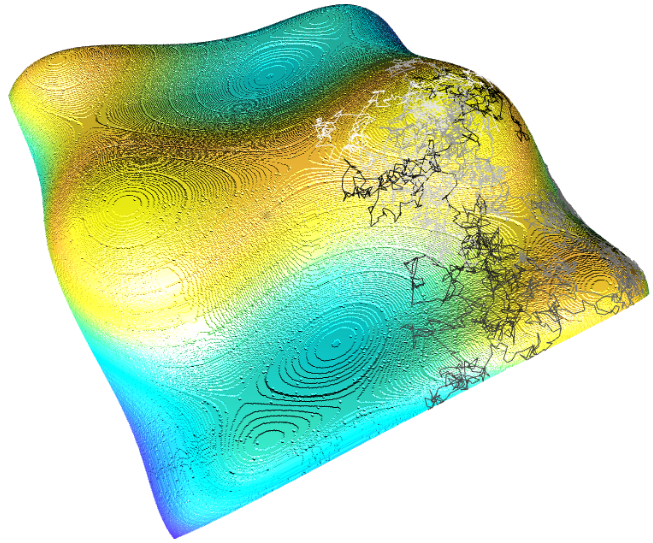

Figure 4: Changing the color of the path over evolutionary time makes the lineage's behavior interpretable; this lineage started at the peak, explored the lower adjacent peak, and then returned to the peak it started on (note: the selection regime under which this lineage evolved prioritized diversity over fitness, hence such a trajectory being viable).

To make visualizations more interpretable, we introduced a color gradient along the lineage (see Figure 4); in the visualization presented here, we use a grayscale lineage drawn onto a colorful landscape. Our lineages transition from white to black as evolutionary time progresses, indicating when each portion of the lineage existed. When drawing just a few lineages, this technique is effective. However, in order to draw large-scale inferences, we really need to be able to look at the results of multiple replicates at once. One option for distinguishing replicates from each other would be to use a different color map for each. This approach quickly becomes unwieldy, however, in light of the fact that the fitness landscape requires a color gradient of its own to serve as a visual queue for its shape. Choosing a compatible set of color maps is possible, and may be a good option in some cases, but there is an easier alternative: separating different lineages from each other along the z axis.

This z-separation approach turns out to be surprisingly intuitive (see Figure 5). In most cases, the lineage still clearly conforms to the shape of the landscape, looking almost like a partial mesh wireframe. Thus, even though the path is hovering above the actual surface, it is obvious how each part corresponds to the underlying fitness function. Moreover, it is possible to simultaneously distinguish different lineages from each other while also getting an impression of the aggregate.

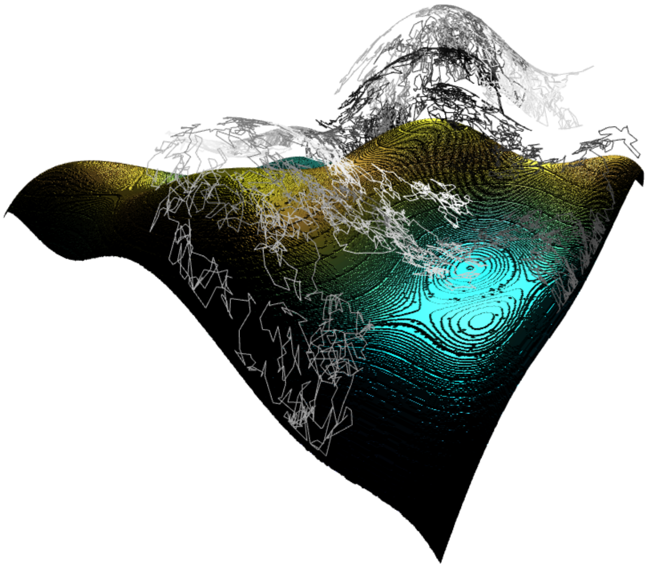

Figure 5: Example of differentiating multiple lineages by separating them along the z axis. Note that this method translates particularly badly to two dimensions and is best viewed in an interactive format.

It is also possible to add further annotations to the visualization, such as spheres marking the start and end points of lineages (see Figure 6). In cases where the lines are too messy (e.g. when the mutation rate is high), this approach can be a good alternative way to get a sense of how a population is spread across a landscape.

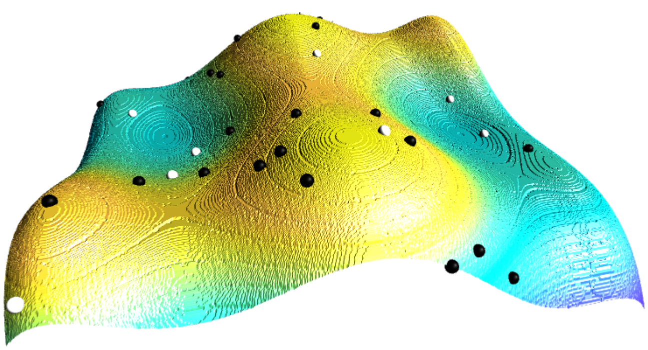

Figure 6: In some cases, spheres marking the start and end points can be easier to interpret than the full lineage path. Here, black spheres indicate the locations of extant organisms at the end of evolution. White spheres indicate the locations of their earliest ancestors (note that many extant organisms share a root ancestor).

Note that our visualization does not interpolate intermediate values between ancestors on lineage paths. At most reasonable mutation rates this has no visible effect. However, in particularly rugged landscapes or at particularly high mutation rates, lineages will sometimes fail to conform perfectly to the contours of the landscape (see Figure 7); for example if a single mutation jumps a lineage directly to the other side of a peak, the line might appear to go through the landscape under that peak. While this behavior may initially seem surprising, it is important that the user is aware of the actual mutations that occurred; having the lineage falsely travel over the peak would create a mis-impression that the peak was reached and then abandoned.

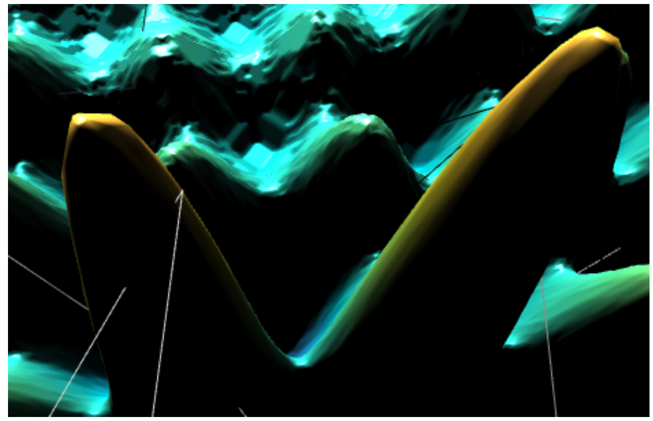

Figure 7: A high mutation rate in a rugged fitness landscape (the Shubert function) can lead to scenarios where lineages "tunnel" through peaks. Here, we can see a lineage (in white) cut straight to a point most of the way up a peak via a large mutation from elsewhere, before immediately taking another large mutational step to a different part of the fitness landscape.

3 IMPLEMENTATION

We built this data visualization using Mozilla's A-Frame framework [1]. We chose A-Frame because it is open source, easy to use, and compatible with a wide range of viewing platforms, including standard web browsers. Thus, it is able to take advantage of new technology while remaining accessible to those without it. Another huge advantage of A-Frame is that it has a large community of people developing components that can be easily plugged into it. For example, this visualization makes heavy use of the heatmap-3d component [10], which draws a three-dimensional surface based on a heat map.

Our visualization is open source and available for public use. It will accept any fitness function that can be described as a heat map. A demo version is available at https://emilydolson.github.io/fitness_landscape_visualizations. All code necessary to reproduce it is available at https://github.com/emilydolson/fitness_landscape_visualizations.

3.1 Platforms

Different platforms provide different tools for interacting with data visualizations. As such, our data visualization has different capabilities when viewed on different platforms.

3.1.1 Desktop Browser. In a normal modern web browser, A-Frame scenes render to WebGL. This capability makes them accessible to users without access to a virtual reality headset. While

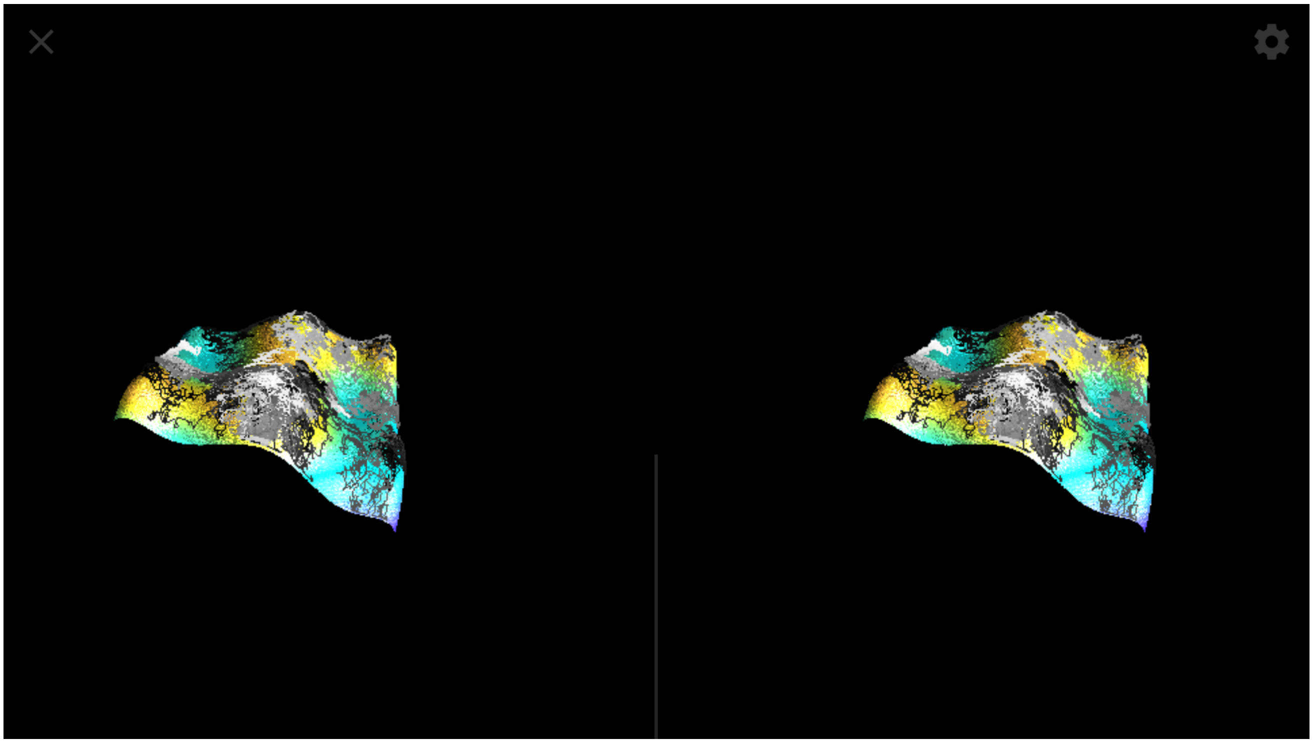

Figure 8: A screenshot of our visualization running in WebVR on a smart-phone.

WebGL visualizations do not allow the user to perceive the visualization in three dimensions, they can be manipulated by clicking and dragging. These actions will rotate the visualization, allowing the user to see it from multiple angles and providing some visual cues to help perceive all three dimensions. WebGL uses the graphics card to accelerate rendering, allowing it to support visualizations that contain large amounts of data.

3.1.2 Smartphone-based virtual reality headsets. Currently, there are a large variety of virtual reality headsets that function in conjunction with a smartphone. Essentially, the user opens an application on their phone that displays two images, one for each eye. The user then puts their phone into the headset, which positions it such that each eye sees the appropriate image. This set-up gives the user a sense of depth, allowing them to better interpret three-dimensional images. Many of these headsets, such as Google Cardboard, are inexpensive, making them a viable option for researchers and hobbyists alike.

For compatible phones, A-Frame will render the visualization to WebVR, an open-source Javascript API for web-based virtual reality. On these devices a glasses icon appears in the lower right of the visualization. Tapping this icon will put the phone into virtual reality mode, readying it to be mounted in the headset (see Figure 8).

While the improved depth perception is helpful, a downside to viewing data visualizations on these platforms is the relative lack of ability to manipulate the visualization. Because they do not have positional tracking, it is not possible to walk around the visualization. Without an additional remote, it is not possible to rotate or zoom in on the visualization. Fortunately, many of the more advanced phone-mount headsets do come with a remote.

3.1.3 Dedicated virtual reality headsets. Lastly, there are dedicated virtual reality headsets, such as the Oculus Rift and HTC Vive. These are generally the most full-featured and, consequently the least accessible (due both to the expense and to the requirement that they be attached to a relatively powerful computer). Generally, these devices feature a controller for each hand, which allows for more fine-grained manipulation of the visualization. Additionally, they support positional tracking, allowing users to walk around to see a visualization from different angles.

When used with these systems, our visualization supports pinching and spreading gestures to zoom in and out and adjust the part of the visualization being viewed (built using [13]). This form of interaction facilitates a more thorough exploration of the fitness landscape. Theoretically it could also function as a rudimentary technique for moving around the visualization on systems that do not track movement via the headset; however, this usage would require two hand controllers with six-degree-of-freedom positional tracking, which is currently a far less common feature than head position tracking.

4 CASE STUDY

To demonstrate the utility of our visualization, we provide some example insights that it has yielded. These examples all occurred in the context of creating a suite of summary metrics for quantifying evolutionary history [3]. We ran a set of experiments in which we varied selection scheme, mutation rate, and selection strength. Subsequently, we pruned this data set to a manageable size by extracting the lineage of only the fittest candidate solution from each replicate run. A subset of these data are shown in the demo visualization for this paper. Over the course of these experiments, we frequently used virtual reality visualizations of lineages overlaid on fitness landscapes to confirm our hypothesized mechanisms for various results.

One particularly helpful insight yielded by our visualization was the effect of mutation rate on the behavior of lineages evolving on the Himmelblau function (see Figure 9). This function is relatively flat, with four global maxima. Two of these maxima are closer together than the other two. As expected, lineages evolved at lower mutation rates do not wander as far from the maxima as lineages evolved at higher mutation rates. The less expected result was that, at the low mutation rate, the most successful lineage from a given run could be found on any of the four maxima. At the high mutation rate, on the other hand, all of the lineages traveled back and forth between the two maxima that are closer to each other. The mechanism behind this was result unclear when we were examining purely numerical data. By visualizing the data, however, we can quickly understand what is happening. Solutions on the maxima that are close to each other are more robust to mutations, because mutations can carry them between these two peaks. This effect, known as survival of the flattest, has been documented in many systems [18].

In the previous example, we used our visualization to look at an aggregate result across multiple lineages. In some cases, however, it is more useful to look at a single representative lineage. The Shubert function has a highly rugged fitness landscape, with periodic steep fitness peaks (see Figure 10). Most selection schemes quickly converged on a single peak without exploring much of the landscape. The exception to this pattern was a diversity-preserving selection scheme (Eco-EA [5]). Lineages evolved under this condition traversed a much larger portion of the fitness landscape, but the specifics were not immediately clear. By visualizing a single lineage, we can clarify the precise behavior of these lineages (visualizing multiple lineages at once was ineffective in this case because they usually overlap too much) (see Figure 10). Specifically, we can see that individual lineages evolved under Eco-EA traverse ridgelines to explore a high percentage of the fitness landscape. Interestingly, they often pass very close to higher fitness peaks without climbing them (presumably because those peaks are already occupied by other solutions).

5 CONCLUSIONS

We have demonstrated that using virtual reality for visualizing evolutionary computation is both useful and achievable. The visualization that we propose here, as well as other visualizations using virtual reality, can dramatically clarify the mechanisms behind outcomes that we usually observe only in the aggregate. As the field of virtual reality data visualization is still in its infancy, we predict that the power of these techniques will continue to increase.

6 FUTURE DIRECTIONS

As previously discussed, most fitness landscapes that correspond to challenging, real-world problems have far more than three dimensions. We believe that virtual reality can be a powerful tool for improving our understanding of these landscapes as well. For fitness functions with three inputs, it should be possible to use all three of the x, y, and z axes to describe the genome and depict fitness via some other variable. A promising option is to fill a three dimensional space with fog and vary the density and/or color of the fog based on the fitness corresponding to a given location in that space. More experimentation is necessary to determine whether this will actually be an intuitive depiction of the fitness landscape, and, if not, how to adjust it. For example, if fog is too challenging to visually tie to a specific location, it should at least be possible to indicate regions of fitness above a certain cut-off by depicting them as solid objects.

What about high-dimensional fitness landscapes? All attempts to visualize them thus far have used graphs to depict which genotypes or phenotypes are connected to each other [8, 15]. Fitness is typically illustrated using either color or position along the y axis. In constructing these visualizations, however, there is a tension between using a node's position in space to provide information and attempting to distinguish nodes by spreading them out. As a result, graph-based visualizations of fitness landscapes can become unwieldy and hard to interpret. Adding a third dimension cannot solve all of these problems, but it can push them farther down the road, allowing us to visualize more complex fitness landscapes. Specifically, we can use the z axis to convey information about fitness while using the x and y axes to space the nodes out. We hypothesize that this approach will produce a more intuitive representation of the fitness landscape where both the high fitness regions and the paths connecting them can be easily picked out.

All of these approaches can be further improved by incorporating additional information through interaction and animation. Interactions can allow us to provide more information about a specific element (e.g. what mutation is responsible for a given change in phenotype), enabling more efficient exploration of data. Animations allow us to use time as an additional dimension with which to convey information. In situations where lineages are too wide-ranging to comprehensibly depict as paths, an animation of individual members of a population traversing the landscape over time should be more effective. As technology continues to advance, more and more of these options will be within easy reach.

ACKNOWLEDGMENTS

This research has been supported by the National Science Foundation (NSF) through the BEACON Center under Cooperative Agreement DBI-0939454, by a Graduate Research Fellowship to ED under Grant No. DGE-1424871, and under Grant No. DEB-1655715 to CO. This work has also be supported by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research and the Digital Scholarship Lab. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

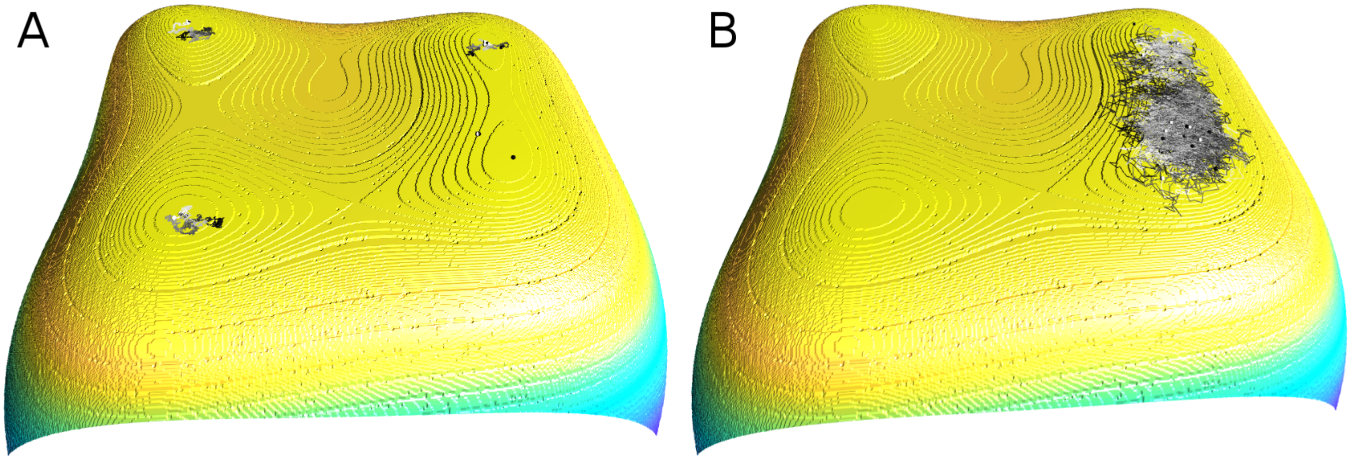

Figure 9: Lineages (start points, end points, and paths) of the fittest solution from ten replicate runs under roulette selection at two different mutation rates. A) At the lower mutation rate, lineages stay close to one of the four global maxima. B) At the higher mutation rate, lineages are not able to stay as close to a single maximum. Only two of the four maxima are occupied.

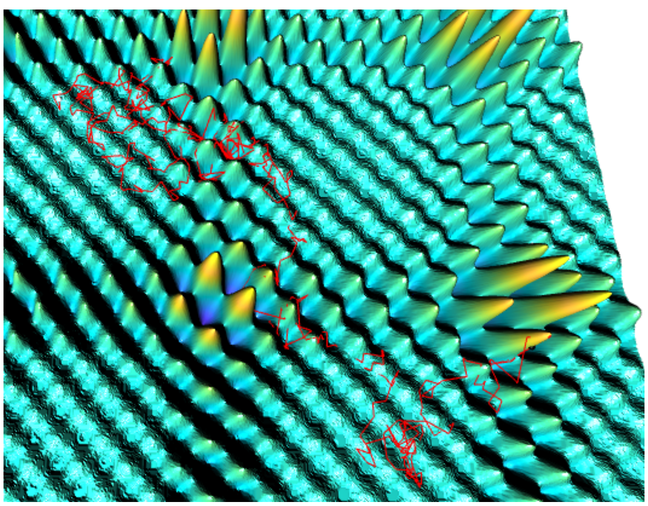

Figure 10: A single lineage traversing the Shubert function under a diversity-maintaining selection regime. The entire path is colored red to increase contrast with the fitness surface, particularly in shadowed regions.

REFERENCES

[1] 2018. aframe: :a: web framework for building virtual reality experiences. (April 2018). https://github.com/aframevr/aframe original-date: 2015-09-15T02:04:40Z.

[2] Bogdan Burlacu, Michael Affenzeller, Michael Kommenda, Stephan Winkler, and Gabriel Kronberger. 2013. Visualization of genetic lineages and inheritance information in genetic programming. ACM Press, 1351. https://doi.org/10.1145/2464576.2482714

[3] Emily Dolson, Alexander Lalejini, Steven Jorgensen, and Charles Ofria. 2018. Quantifying the tape of life: Ancestry-based metrics provide insights and intuition about evolutionary dynamics. In *In review*.

[4] Sergey Gavrilets. 2010. High-Dimensional Fitness Landscapes and Speciation. In *EvolutionâĂŤthe Extended Synthesis*, Massimo Pigliucci and Gerd B. MÃijller (Eds.). The MIT Press, 45–80. https://doi.org/10.7551/mitpress/9780262513678.003.0003

[5] S. Goings and C. Ofria. 2009. Ecological approaches to diversity maintenance in evolutionary algorithms. In *IEEE Symposium on Artificial Life, 2009. ALife '09*. 124–130. https://doi.org/10.1109/ALIFE.2009.4937703

[6] Stuart Kauffman and Simon Levin. 1987. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* 128, 1 (Sept. 1987), 11–45. https://doi.org/10.1016/S0022-5193(87)80029-2

[7] Xiaodong Li, Andries Engelbrecht, and Michael G Epitropakis. 2013. Benchmark Functions for CECâĂŹ2013 Special Session and Competition on Niching Methods for Multimodal Function Optimization. (2013), 10.

[8] David M. McCandlish. 2011. Visualizing fitness landscapes. *Evolution; international journal of organic evolution* 65, 6 (June 2011), 1544–1558. https://doi.org/10.1111/j.1558-5646.2011.01236.x

[9] Nicholas Freitag McPhee, Maggie M. Casale, Mitchell Finzel, Thomas Helmuth, and Lee Spector. 2016. Visualizing Genetic Programming Ancestries. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion (GECCO '16 Companion)*. ACM, New York, NY, USA, 1419–1426. https://doi.org/10.1145/2908961.2931741

[10] morandd. 2018. aframe-heatmap3d: Terrain-like heatmap data viz component for AFrame. (Jan. 2018). https://github.com/morandd/aframe-heatmap3d original-date: 2017-05-12T10:15:51Z.

[11] Ekaterina Olshannikova, Aleksandr Ometov, Yevgeni Koucheryavy, and Thomas Olsson. 2015. Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data* 2, 1 (Dec. 2015), 22. https://doi.org/10.1186/s40537-015-0031-2

[12] Randy Olson. 2014. Using fitness landscapes to visualize evolution in action. (2014). https://www.youtube.com/watch?v=4pdiAneMMhU

[13] Fernando Serrano. 2018. aframe-camera-transform-controls-component: Camera transform controls component for A-Frame. (March 2018). https://github.com/fernandojsg/aframe-camera-transform-controls-component original-date: 2018-02-27T12:21:12Z.

[14] Russell K. Standish and John Galloway. 2002. Visualising TierraâĂŹs tree of life using Netmap. In *ALife VIII Workshop proceedings*. 171. http://ww.molaah.com/Alife%20NetMap.pdf

[15] Barrett Steinberg and Marc Ostermeier. 2016. Environmental changes bridge evolutionary valleys. *Science Advances* 2, 1 (Jan. 2016), e1500921. https://doi.org/10.1126/sciadv.1500921

[16] Andries van Dam, David H Laidlaw, and Rosemary Michelle Simpson. 2002. Experiments in Immersive Virtual Reality for Scientific Visualization. *Computers & Graphics* 26, 4 (Aug. 2002), 535–555. https://doi.org/10.1016/S0097-8493(02)00113-9

[17] Nathaniel Virgo, Eran Agmon, and Chrisantha Fernando. 2017. Lineage selection leads to evolvability at large population sizes. MIT Press, 420–427. https://doi.org/10.7551/ecal_a_070

[18] Claus O. Wilke, Jia Lan Wang, Charles Ofria, Richard E. Lenski, and Christoph Adami. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412, 6844 (July 2001), 331–333. https://doi.org/10.1038/35085569

[19] Sewall Wright. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the Sixth International Congress of Genetics*. Brooklyn Botanic Garden, Brooklyn, NY, 356–366. http://www.esp.org/books/6th-congress/facsimile/contents/6th-cong-p356-wright.pdf