

BIOLOGY: A Game for a Crowd

By Sarah C.P. Williams

The rules of Phylo are simple: drag colored blocks across rows on the computer screen until similar colors line up. Within minutes of launching the game, any average person can learn how to play and begin developing strategies to beat the current best score, which is posted prominently at the top of the page.

It can be addictive even for players who don't know that the colored blocks actually represent gene sequences submitted by scientists trying to solve real-world biological puzzles.

"I wanted a game that an average person could play when they had a couple min-

utes—like Tetris—but that would be useful for bioinformatics," says Jerome Waldispuhl, PhD, an assistant professor of computer science at McGill University who spearheaded the development of the game with a colleague. Phylo is helping Waldispuhl tackle the challenge of what—in biological jargon irrelevant to the gamers—is called multiple sequence alignment. "When a geneticist gets a new sequence, the first thing they often want to do is compare it to sequences from other species or individuals," he explains. Comparing sequences means lining them up, finding bits of the genetic code that match between the samples. Bioinformaticians typically rely on computer programs to parse the data and come up with such an alignment. But the solution provided by the computer, based solely on statistics, isn't always the best alignment, Waldispuhl says. "Multiple sequence alignment is a problem that is very difficult in computer science," he says. "But it's also one of the most used techniques in genomics studies today."

utes—like Tetris—but that would be useful for bioinformatics," says Jerome Waldispuhl, PhD, an assistant professor of computer science at McGill University who spearheaded the development of the game with a colleague.

Phylo isn't the first, or the most-played, game that aims to solve scientific puzzles that have stumped researchers and computers. But it's part of a growing trend to drive biology forward by initiating games and competitions—among scientists and non-scientists alike. Foundations are offering cash prizes to those who come up with the best solution to a scientific quandary; institutes are posing broad research questions to people across disciplines to encourage out-of-the-box thinking; and computer scientists are teaming up with life scientists to turn biological enigmas into games for the realization led to Phylo.

"I wanted a game that an average person could play when they had a couple minutes—like Tetris—but that would be useful for bioinformatics," says Waldispuhl.

utes—like Tetris—but that would be useful for bioinformatics," says Jerome Waldispuhl, PhD, an assistant professor of computer science at McGill University who spearheaded the development of the game with a colleague.

Phylo is helping Waldispuhl tackle the challenge of what—in biological jargon irrelevant to the gamers—is called multiple sequence alignment. "When a geneticist gets

To improve on what the computers do, geneticists usually sift through the data manually, looking for ways to rearrange chunks of nucleotides to match up with sequences in other samples. Waldispuhl realized that when geneticists worked through the problem in this manual way, their knowledge of genetics wasn't itself vital; they viewed the chunks of genes as they might view colored blocks in a puzzle. That

average public.

Waldispuhl—like others involved with such initiatives—is careful to make it clear that the games and competitions aren't replacing the important work being done by skilled scientists; they're supplementing this work.

"I wouldn't say that humans do better than computers at every part of this task," Waldispuhl says of multiple sequence alignment. "What we tried to do with Phylo is find a better synergy between what humans can do and what computers can do."

More than 300,000 people have played Phylo since it launched in 2010, and Waldispuhl's team reported in a 2013 *Genome Biology* paper that up to 50 percent of the time, a casual gamer can match the performance of expert players; and up to 40 percent of the time, Phylo players can improve on the solution found by a computer program. Now, the scientists have expanded the game so that researchers working on any genetic problem—from high blood pressure to can-



In the game of Phylo, players try to align colored blocks representing genetic sequence information for different species. Image captured from phylo.cs.mcgill.ca

cer—can submit their sequences to be aligned by gamers. The players' alignments don't cure disease, but they provide better starting points for the researchers who do.

Shedding Scientific Baggage

Dan MacLean, PhD, a bioinformatician at The Sainsbury Laboratory in Norwich, the United Kingdom, had a similar goal in mind when he developed the game Fraxinus. Rather than lining up blocks of colors, Fraxinus players line up green, orange, yellow, and red tinted ash leaves that represent nucleotides. The symbolism is purposeful: the genetic sequences in this game are from ash trees and the fungus (*Chalara fraxinea*) that threatens up to 95 percent of Britain's ash trees. MacLean and his colleagues are racing against the clock to understand this fungal disease—ash dieback—and save the country's trees. They want to know what parts of the fungus genome make it so effective at causing disease, as well as whether particular genetic sequences in the ash trees make them more or less susceptible to infection.

The challenge in developing a game to help compare genetic sequences of different ash trees and different fungi, MacLean says, was to understand what basic rules sci-

ing around, what were the assumptions I was making," he says. Working with a team of non-biologists helped him pare the problems down to a simple game. In the first 10 weeks Fraxinus was online, more than 10,000 sequences were analyzed.

"A computer program makes certain assumptions and there are certain limits on the number of permutations it can try,"

people successfully solve puzzles within the confines of the Fraxinus game, he can learn better ways to teach computer programs to align the sequences. The very lack of scientific baggage that the players have, he says, may help players come up with new approaches to the problem. The scientists are still reviewing the data analyzed by the players, but hope that genetic



Fraxinus, like Phylo, involves matching sequences of colored shapes to a given pattern. The fact that the pattern represents the genetic sequence of an ash tree is irrelevant to the gamer. Image captured from <https://apps.facebook.com/fraxinusgame/>

“These sequences we put into the game are not the low-hanging fruit; they’re not the easy ones to solve. They’re the ones that have multiple gaps and multiple possible arrangements. For these sorts of things, humans can do it better than computers,” MacLean says.

entists followed when they normally analyzed the data and how to convey those rules in the game.

“It was really a matter of working out what was the scientific baggage I was carry-

MacLean says. “These sequences we put into the game are not the low-hanging fruit; they’re not the easy ones to solve.

They’re the ones that have multiple gaps and multiple possible arrangements. For these sorts of things, humans can do it better than computers.”

MacLean hopes that by observing how

mutations linked to ash dieback susceptibility will be revealed.

Other online games are also advancing science with assistance from creative non-scientists. Since 2008, players of Foldit have helped determine how proteins fold; while Eyewire users trace the route of a neuron's axons through MRI slices to help build the connectome.

Open Competitions

But games aren't the only way scientists are crowdsourcing computational work to unstick the way scientists think. Universities and other entities have been launching idea challenges, often with a cash prize attached.

For example, since 2010, researchers involved in Harvard Catalyst—a cross-disciplinary effort to drive biomedical research forward—have hosted several open challenges. In February 2013, Eva Guinan, MD, a radiation oncologist and director of the Harvard Catalyst Linkages Program, and colleagues at Harvard, announced the re-

sults of a complex immuno-genomics challenge. The competition, which carried weekly \$500 prizes, sought a program that could more quickly analyze vast amounts of sequence data for the genes that make antibodies and T-cell receptors (TCRs). It's a tough problem because unlike other genes, those for antibodies and TCRs are built up combinatorially in each cell—i.e., they differ from cell to cell—making it difficult to trace the genetic origin of particular antibodies or TCRs. To issue the challenge, Guinan's team uploaded genetic data onto the website TopCoder and rephrased their problem in generic, non-biological terminology. They essentially created an information-theory and string-processing task that any computational expert could tackle.

The results astounded Guinan: The coders who entered the contest didn't converge on a single best method to solve the antibody challenge. Instead, out of more than 100 submissions, 16 different new approaches worked better than the standard algorithm. And they didn't just give better solutions; some were nearly a thousand times faster too.

"Our small community had missed not just one opportunity to improve the way we were doing things, but many different op-

portunities," says Guinan. Now, scientists can integrate the new methods into the way they study antibodies.

"When you go to that many people, you have some who are good at writing algo-

"A solution is only going to be good if the question posed is good," Guinan says. "And that's where the scientist comes in. This isn't just about throwing a bunch of data at someone and saying 'Call me when you have the answer.'"

rithms, some who are good at thinking about statistics, some who are good at string theory," says Guinan. "You get this amazing diversity of repertoire in the solver population. How could I possibly hire that many different experts here?" Guinan once again emphasizes that such competitions don't detract from the work that scientists themselves do. In fact, she says, it takes

skilled scientists to design and implement an effective competition.

"A solution is only going to be good if the question posed is good," she says. "And that's where the scientist comes in.

This isn't just about throwing a bunch of data at someone and saying 'Call me when you have the answer.'" The Harvard Catalyst group now has more competitions in the works. Up next: challenges to address the genetics of HIV and the best way to analyze colonoscopy data.

Sweetening the Pot

Some open challenges offer even bigger prizes. In 2011, for example, the Pistoia Alliance announced the Pistoia Sequence Squeeze Competition to develop new ways of compressing genetic data, with a \$15,000 prize for the best solution. **Richard Holland**, the chief business officer of Eagle

Genomics, teamed up with Pistoia to run the competition. He says that updating results constantly—with a leaderboard—was technically challenging but helped spur faster improvements.

"It encourages people to compete in order to outdo each other," he says. Leaders constantly leap-frogged each other as they improved their techniques. Holland

also learned that having strictly defined entry and judging criteria ensures that a contest runs smoothly.

And the prize money didn't hurt. For scientists and non-scientists alike, the drive to win a game or competition—whether for money or pride—combined with the natural human instinct to solve puzzles can be a powerful motivator in driving research forward.

"The wrong message here would be to say that humans are better than computers or that crowd-sourcing is better than experts," says Waldispuhl. "It's more about trying to see where we can improve different steps of the research process by doing things a little bit differently." □