



Online citizen science games: Opportunities for the biological sciences



Vickie Curtis*

Institute of Educational Technology, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom

ARTICLE INFO

Article history:

Received 16 June 2014

Received in revised form 16 June 2014

Accepted 4 July 2014

Keywords:

Citizen science

Games

Foldit

EteRNA

Phylo

ABSTRACT

Recent developments in digital technologies and the rise of the Internet have created new opportunities for citizen science. One of these has been the development of online citizen science games where complex research problems have been re-imagined as online multiplayer computer games. Some of the most successful examples of these can be found within the biological sciences, for example, Foldit, Phylo and EteRNA. These games offer scientists the opportunity to crowdsource research problems, and to engage with those outside the research community. Games also enable those without a background in science to make a valid contribution to research, and may also offer opportunities for informal science learning.

© 2014 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Citizen science is a collective term for projects that engage both professional scientists and non-scientists in the process of gathering, evaluating and/or computing various scientific data (Kostadinova, 2011). Citizen science has traditionally referred to projects which have used volunteers to collect ecological, biological or environmental data (Devictor et al., 2010). However, recent developments in digital tools and the growth of the Internet have greatly altered the citizen science landscape, increasing the number of people who are able to participate in such projects, and expanding the range of disciplines and research problems that can be addressed. A growing number of citizen science projects are conducted entirely through the Internet and participants help to analyse large sets of data that have been provided by the project scientists. These projects have been referred to as online, or virtual, citizen science (Reed et al., 2012).

2. Online citizen science

Online citizen science projects enable interested individuals to get involved with scientific research wherever there is an internet connection. Projects can take a variety of formats and the task required of the citizen scientists may vary in its level of complexity. For example, distributed computing projects simply require that participants run project software that automatically analyses 'work units' provided by the project team. For example, Folding@home (<http://folding.stanford.edu/>) is a project in which participants run algorithms that simulate

protein folding. No other active input is required. Other projects require greater cognitive involvement and participants may be asked to classify or annotate images or graphical data. For example, Cell Slider (<http://www.cellslider.net/>) asks participants to help identify cancer cells on archived cell samples. More recently, scientific research problems have been repackaged into online multi-player computer games that use stylised graphical interfaces and introduce elements of gamification such as competition between players and performance ranking.

3. Foldit

One of the first citizen science games to be developed, and perhaps one of the most successful in terms of significant results produced, is Foldit (www.fold.it). In this project, the creation of accurate protein structure models has been turned into a game, and players have been responsible for deducing the structure of proteins that have been difficult to ascertain using more conventional approaches (Khatib et al., 2011a,b). Foldit was developed at the University of Washington in Seattle by a group consisting of both biochemists and game developers. It was released to the public in May 2008. The rationale for the development of Foldit was to harness the collective problem-solving abilities of *non-experts* and to exploit a wide diversity of playing approaches to accelerate progress in understanding the three-dimensional structures of protein (Cooper, 2011). Foldit players enlist human three-dimensional problem solving skills and online manipulation tools based on computer algorithms, to produce accurate models of protein structures that have previously been unknown (Khatib et al., 2011a,b). The success of Foldit has been attributed to the diversity of playing approaches and styles, although it may also be related to the lack of expectations or biases of those who come to the game without any background knowledge of biochemistry.

* Tel.: +44 1908 653546.

E-mail address: Vickie.curtis@open.ac.uk.

Players can play individually, or within a team, and compete against one another within a points system. Protein structures that come closest to their 'natural' configuration (that is one that requires the least amount of energy) are awarded a greater number of points. Despite the competitive aspect of the game, players also work together cooperatively and collaboratively to solve the protein puzzles. Players interact through real-time internet relay chat (IRC) during the game with any other individual who is playing and individual teams have their own IRC channels where members of the same team can work together. Players can also have real-time interaction with the scientists and developers of Foldit through regular IRC 'chats' where new tools or new science puzzles can be discussed, or issues with the game interface or software can be highlighted.

Foldit is a complex game and can be difficult to learn compared to other multiple-player online games (Andersen et al., 2012). Before a player can compete and work on protein puzzles where the structure is unknown (known in Foldit as 'science puzzles'), he/she is presented with a series of tutorial or 'intro' puzzles (32 in total) that guide them through the various game tools available. These are based on proteins where the structure is already known. Once confident with the tools and the general layout of the game, a new player is then able to play the science puzzles (Fig. 1). A player does not have to complete all the tutorial puzzles before playing the science puzzles, although new players are often advised by more experienced players on the IRC to complete them if they want to play the game effectively. A number of more experienced players are active in helping those new to the game and may guide them through the learning process.

The majority of people who play Foldit have little or no background in biochemistry (Khatib et al., 2011a,b), yet some have become adept at the game and have helped to develop a novel approach to understanding the science of protein folding. Indeed, Foldit has had some success

which has resulted in a number of publications (Eiben et al., 2012; Khatib et al., 2011a; Khatib et al., 2011b). Perhaps the most significant of these is based on the efforts of Foldit players to 'solve' the structure of the Mason-Pfizer monkey virus (M-PMV) retroviral protease, a simian AIDS-causing monkey virus. While biochemists have attempted to elucidate the structure of this molecule for a number of years, two teams of Foldit players were able to construct an accurate model using the game interface in three weeks (Khatib et al., 2011b). These two teams of Foldit players were included in the list of authors on the subsequent paper.

The efforts of Foldit players have also led to a significant improvement in algorithms used to predict protein structures, and players have remodelled the backbone of a computationally designed Diels–Alderase, thus enhancing its activity and interaction with substrates (Khatib et al., 2011a; Eiben et al., 2012). Foldit players were included collectively in the list of authors in both of these publications.

Consequently, Foldit has received much attention from journalists and science communicators, and a number of articles have been published praising Foldit for its approach to collaboration with non-expert citizens, 'crowdsourcing science', and with opening up the scientific research process to 'the masses' (Hand, 2010; McGonigal, 2011).

4. Other online citizen science games: Phylo and Eterna

Phylo (<http://phylo.cs.mcgill.ca/>) was developed by computer scientists and biologists at McGill University and released in 2010. It uses comparative genomics as the basis of a Tetris-like game and utilises human pattern recognition skills to address the Multiple Sequence Alignment problem (Kawrykow et al., 2012). Players attempt to match up coloured blocks that represent nucleotide sequences from both coding and regulatory areas of genes from different animal species

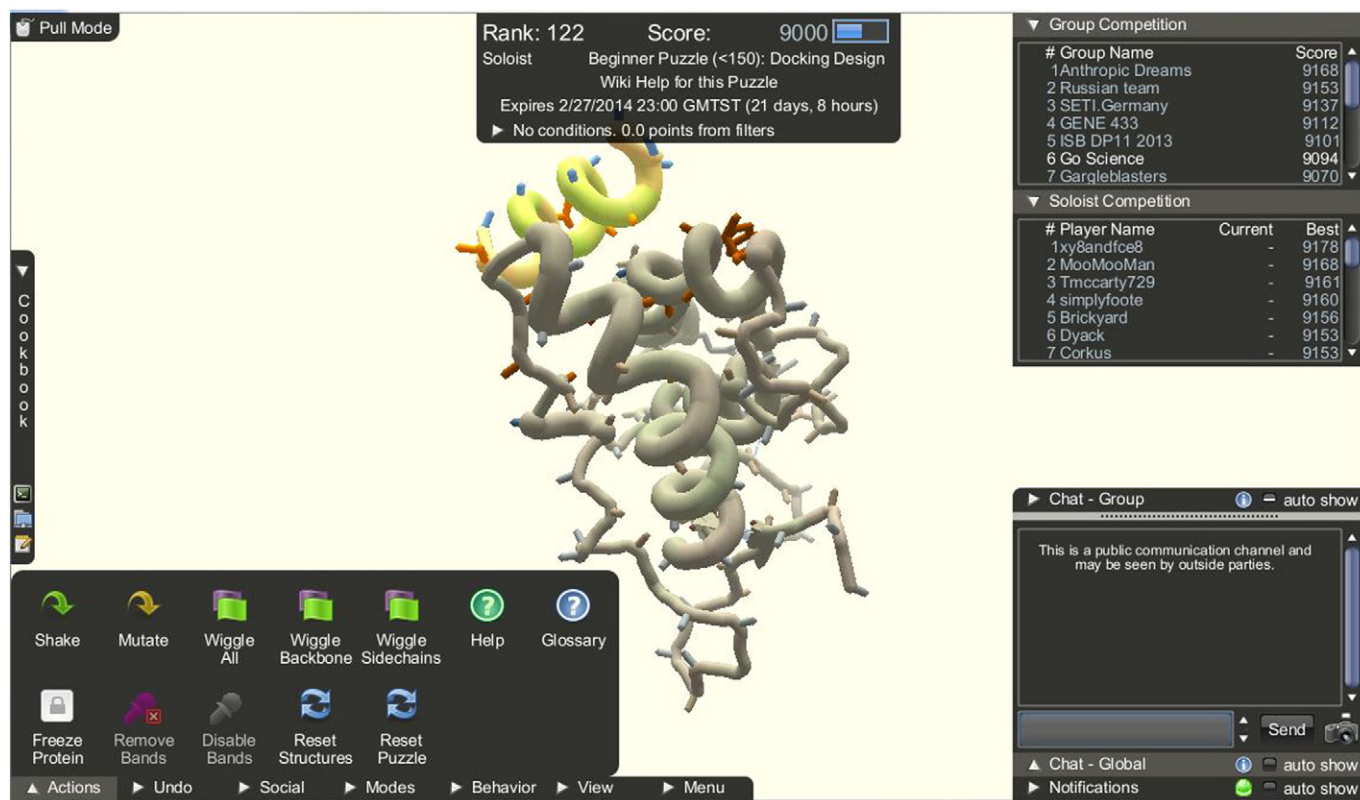


Fig. 1. Foldit science puzzle. The game tools are visible in the bottom left-hand corner, the rank of all players taking part in this puzzle is displayed in the top right-hand corner, and the internet relay chat (IRC) window is in the bottom right-hand corner.

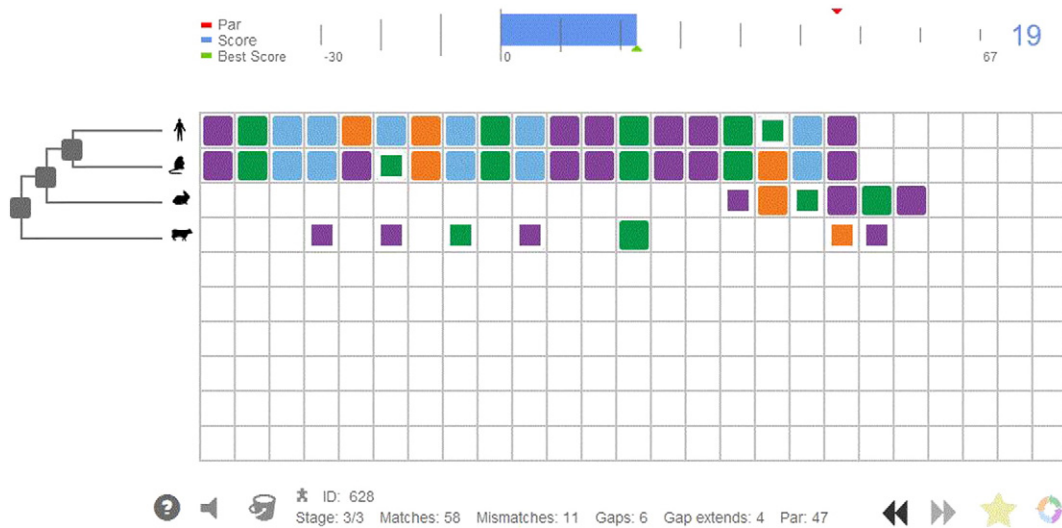


Fig. 2. Phylo puzzle. Players attempt to line up squares of the same colour as best they can minimising the spaces and gaining as many points possible. More species (shown on the left of the screen) are introduced as the level of difficulty of the puzzle increases.

in an attempt to determine phylogenetic relationships, the impact of mutations and their potential role in disease. This game is highly stylised and much of the science behind the game is hidden from the player in an attempt to broaden the spectrum of participants (Fig. 2). Players can select a puzzle based on a disease category (e.g. infectious

diseases, cancer, metabolic disease) and this may help them to understand the relevance of the research, or to contribute to areas that may be of greater personal interest or relevance.

The creators of this game have found that a citizen science approach can be used to improve the accuracy of multiple sequence alignments,

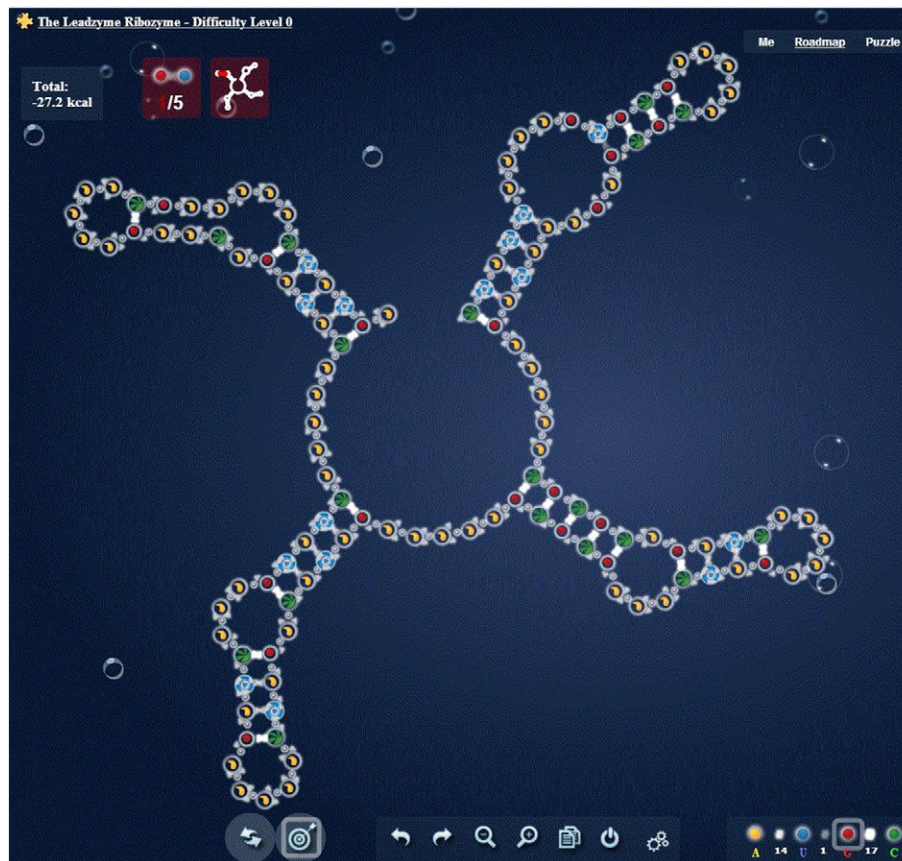


Fig. 3. EteRNA puzzle. Each of the nucleotide bases is represented by four different colours and players must produce stable linkages. Some puzzles have stipulations for numbers of specific types of base pairing (in this puzzle, there must be at least 5 uracil–guanine pairings).

and that this complex research problem can be turned into an entertaining game that can be played without any background scientific knowledge. Over 225,000 puzzles have been completed, and to date, there has been one publication based on the game output with 'Phylo players' appearing collectively on the list of authors (Kawrykow et al., 2012). This work has shown that the efforts of citizen scientists can be used in combination with existing algorithms to improve the accuracy of multiple sequence alignments, thus demonstrating the effectiveness of human–computer collaboration. The ultimate goal of this work however, is to identify the origin of genetic diseases and to identify functional patterns in DNA.

EteRNA (<http://eterna.cmu.edu/web/>) is a game that challenges players to design new ways to fold RNA. It was launched in 2010 and is a collaborative effort by computer scientists at Carnegie Mellon University and scientists at Stanford University. Like Foldit, players can work individually or in groups, and using the game interface they design two-dimensional structures that can include complex patterns such as lattices, knots and switches (Fig. 3). Players can receive feedback on their designs from the project scientists and each week, the best designs created by the players are synthesized in the lab at Stanford.

A recent publication based on the game suggests that players learn from the results of the lab-based experiments and adapt their puzzle strategies as a result (Lee et al., 2014). Puzzle designs and playing approaches have also been used to design algorithms that have outperformed prior algorithms used to design RNA structures. As well as this improvement in algorithms, the goal of EteRNA is to establish a library of synthetic RNA molecules which may eventually be used in the design of novel therapeutics.

There is an active online community of EteRNA players, and individuals can communicate with each other through an IRC while playing and share tips and strategies, as well as advice to new players. Scientists and developers also communicate regularly with players via online chats.

5. Accessibility

All three of these games can be played on modern low-cost PCs with any type of internet connection. Both Phylo and EteRNA are played within web browsers, while Foldit is a stand-alone programme. Given the greater graphical complexity and 3D nature of Foldit, the game will be slower on less powerful computers. The potential for mobile phone versions of these games has not been thoroughly explored, although there have been discussions on the Foldit forum relating to this. However, work on mobile versions of other online citizen science projects has been problematic due to the battery power required to run them.

Foldit players have constricted a game wiki where instructions for playing, help with tutorial puzzles and general playing tips and strategies are shared. This has been translated into German, Italian, Spanish, Russian, Japanese and Korean. The game itself can be played in Czech, Dutch, French, German, Hebrew, Italian, Polish, Romanian, Russian and Spanish in addition to English. Phylo has also been translated into other languages including French, German, Portuguese, Russian, Spanish, Hebrew, Romanian, Korean, Chinese and Japanese.

6. The appeal of citizen science games

Given the relatively recent appearance of online citizen science games, there has been little investigation into the motivation for playing, or more general patterns of participation. A small number of studies carried out on other online citizen science projects suggest that individuals take part because they have a background interest in the science, and because they want to make a contribution to scientific research (Nov et al., 2011; Raddick et al., 2010). Preliminary work on patterns of participation suggests that while many individuals register to take part in these projects, only a small percentage will become active

participants and that even smaller groups of 'super participants' perform the majority of the tasks (Ponciano et al., 2014). This pattern of participation, also known as a 'power law' distribution or the 'Pareto Principle', is observed in other online communities, for example Wikipedia, where small numbers of committed participants provide the majority of the content (Cooper, 2013).

Observations of these three citizen science games appear to confirm a similar pattern of participation. For example, Foldit has tens of thousands of registered participants, yet a closer examination of their user profiles has revealed that most of these individuals have never actually played the game, or may have only completed one or two tutorial puzzles. Observations of the online forum, the IRC, transcripts of online scientist and developer chats, and the presence of the same individuals in the game header boards suggest that active playing community is in the hundreds rather than the thousands with a much smaller number of players dominating the game. While this pattern of participation may be characteristic of other types of online contributory projects, small numbers of players may also be the result of the lengthy tutorial process, and the level of difficulty of the game.

7. Conclusions: the potential benefits of online citizen science games

Citizen science games enable those without a background or formal qualification in science to become involved in authentic scientific research wherever there is access to the Internet. Like other types of citizen science projects, games may offer opportunities for informal science learning, particularly if the project scientists actively communicate with the community of participants or provide background knowledge and relevant links on project websites. Observations of the interactions between Foldit players suggest that some individuals acquire a high degree of scientific knowledge and are comfortable using scientific and technical terms.

In addition to the benefits for participants, there are also benefits for scientists in that these projects enable some research problems to be effectively crowdsourced, and allow scientists to take advantage of differing approaches to problem solving and human pattern recognition skills. In some cases (e.g. Phylo and Foldit) novel approaches have been identified through the work of citizen scientists (Khatib et al., 2011a,b; Kawrykow et al., 2012). Online citizen science games, as well as other types of online citizen science projects, also present an opportunity for public engagement with those external to the research community and a chance to increase the transparency of the research process (Curtis, 2014; Nielsen, 2012). Ultimately, projects such as Foldit, Phylo and EteRNA demonstrate that there are valid research outcomes. The games described here have the potential to greatly improve our understanding of the genetic processes underlying important diseases, and may well play a future role in the design of effective therapeutics.

While online citizen science projects may offer a number of opportunities, they also require an investment of resources to develop and maintain the technical infrastructure. This will inevitably be determined by the scope and complexity of the game, and scientists will need to work alongside software developers and/or computer game specialists. In addition to investing in a technical infrastructure, scientists must also be prepared to invest time in a project and interact with potentially large groups of online citizen scientists. Preliminary research in this area which has included interviews with scientists and developers involved in online citizen science projects, suggests that some scientists underestimate the level of commitment that this side of citizen science requires. Indeed, Foldit (as well as other projects) has hired 'community support managers' whose primary role is to liaise between project scientists and the community of volunteers. However, feedback from scientists also suggests that this interaction between scientists and citizen scientists can be highly rewarding, offering opportunities for dialogue and collaboration with those outside the traditional research community.

Acknowledgements

I would like to thank Dr Carol Isaacson Barash for her helpful comments on this manuscript.

References

- Andersen, E., O'Rourke, E., et al., 2012. The impact of tutorials on games of varying complexity. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems*, Austin, Texas, USA, ACM.
- Cooper, C., 2013. Zen in the art of citizen science: apps for collective discovery and the 1 percent rule of the web. *Sci. Am.* (<http://blogs.scientificamerican.com/guest-blog/2013/09/11/zen-in-the-art-of-citizen-science-apps-for-collective-discovery-and-the-1-rule-of-the-web/>, accessed 5/4/14).
- Cooper, S., 2011. *A Framework for Scientific Discovery Through Video Games*. Doctor of Philosophy, University of Washington.
- Curtis, V., 2014. Public engagement through the development of science-based computer games: the Wellcome Trust's "Gamify Your PhD" initiative. *Sci. Commun.* 36 (3), 379–387.
- Devictor, V., Whittaker, R.J., et al., 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Divers. Distrib.* 16, 354–362.
- Eiben, C.B., Siegel, J.B., et al., 2012. Increased Diels–Alderase activity through backbone remodelling guided by Foldit players. *Nat. Biotechnol.* 30 (2), 190–192.
- Hand, E., 2010. People power. *Nature* 466, 685–687 (5 August 2010).
- Kawrykow, A., Roumanis, G., et al., 2012. Phylo: a citizen science approach for improving multiple sequence alignment. *PLoS ONE* 7 (3).
- Khatib, F., Cooper, S., et al., 2011a. Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci.* 108.
- Khatib, F., DiMaio, F., et al., 2011b. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat. Struct. Mol. Biol.* 18, 1175–1177.
- Kostadinova, I., 2011. Citizen science – the new helping hand for scientists. *Curr. Sci.* 100 (7), 973–976.
- Lee, J., Kladwang, W., et al., 2014. RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci.* <http://dx.doi.org/10.1073/pnas.1313039111>.
- McGonigal, J., 2011. Be a gamer, save the world. *Wall Street J.* (22/01/11), URL: <http://online.wsj.com/article/SB10001424052748704590704576092460302990884.html> (accessed 5/4/14).
- Nielsen, M., 2012. *Reinventing Discovery. The New Era of Networked Science*. Princeton University Press.
- Nov, O., Arazy, O., et al., 2011. Dusting for Science: Motivation and Participation of Digital Citizen Science Volunteers. *iConference 2011*, (Seattle, Washington).
- Ponciano, L., B. F., et al., 2014. Volunteers' Engagement in Human Computation Astronomy Projects. *Computing in Science & Engineering* in press.
- Raddick, J., Bracey, G., et al., 2010. Galaxy Zoo: exploring the motivations of citizen science volunteers. *Astron. Educ. Rev.* 9 (1).
- Reed, J., Rodriguez, W., et al., 2012. A framework for defining and describing key design features of virtual citizen science projects. *Proceedings of the 2012 iConference*. Toronto, Ontario, Canada, ACM, pp. 623–625.