

# Harnessing the collective intelligence in response to Chalara dieback of ash

## The 21st century scientist

The evolution of the Internet and the explosion in social media portals has led to a dramatic shift in the way scientific communities relay information and ideas, from traditional unidirectional reporting to multidirectional exchanges (Darling *et al.*, 2013).

These online tools allow researchers to boost their professional visibility whilst sharing thoughts, opinions, articles and commentaries on various topics of interest. With at least 1 in 40 scientists active on the micro-blogging platform Twitter, the benefits of an online profile are far-reaching (Priem *et al.*, 2012). For instance, many scientists have found the use of social media outlets effective in broadening interest in their research articles, as reflected in spikes in download rates (Terras, 2012).

With no geographical constraints on these virtual communities and networks, scientists also often find these outlets conducive to the development of

interdisciplinary collaboration.

Beyond scientific boundaries, increasing the visibility of research provides a transparency that is necessary to unlock the ivory tower of scientific research within the public arena. By cultivating the broad interest of the general public in science, researchers are now empowered to consider ways to utilize their huge mass and collective intelligence to accelerate scientific discoveries. This method is a form of “crowdsourcing”.

## How do scientists design effective crowdsourcing projects?

Despite major advances in computational systems and algorithms, researchers often remain reliant on manual curation for fine-tuning analysis of complex datasets; with current algorithms producing a degree of false or incomplete outputs. Manual intervention is essential to repair these false outputs but this is extremely time-consuming. Crowdsourcing provides the

means to utilize people power to accelerate progress in these areas. However, to design effective crowdsourcing projects several key questions require careful consideration, such as: how can researchers motivate their target audience to participate in a crowdsourcing initiative? How do you reach the crowd? If targeting a non-specialist audience, how can researchers mould the required analysis into an interface that can be broadly understood? How will participants be recognized in future publications?

One way of capturing the interest of the general public is by integrating analysis into an interactive game format. Games provide the opportunity to seek solutions to scientific riddles in enjoyable, familiar frameworks. To date, several games have been developed by the scientific community to address specific biological questions (Table 1).

Biogame is a diagnostic game where players try to find malaria-infected cells from multiple cell pictures (Mavandadi



**Table 1.** Games released as interfaces to scientific research for the general public

Game	Year released	URL	Description
Fraxinus	2013	<a href="https://apps.facebook.com/fraxinusgame/">https://apps.facebook.com/fraxinusgame/</a>	Puzzles. Sequencing alignments of ash-dieback pathogen genome
Dizeez	2013	<a href="http://sulab.scripps.edu/dizeez/">http://sulab.scripps.edu/dizeez/</a>	Quiz. Links of genes and disease
EyeWire	2012	<a href="https://eyewire.org">https://eyewire.org</a>	Mapping of connections between retinal neurons
BioGames	2012	<a href="http://biogames.ee.ucla.edu">http://biogames.ee.ucla.edu</a>	Classification of malaria-infected cells
Phylo	2012	<a href="http://phylo.cs.mcgill.ca">http://phylo.cs.mcgill.ca</a>	Puzzles. Sequencing alignments of human genome
MalariaSpot	2012	<a href="http://www.malariaspot.com">http://www.malariaspot.com</a>	Identification of malaria parasites
GenESP	2012	<a href="http://sulab.scripps.edu/GenESP/fluid.htm">http://sulab.scripps.edu/GenESP/fluid.htm</a>	Quiz
The Cure	2012	<a href="http://www.genegames.org/cure/">http://www.genegames.org/cure/</a>	Card game. Breast cancer
Happy Match	2012	<a href="http://www.citizensort.org/web.php/happymatch">http://www.citizensort.org/web.php/happymatch</a>	Classification of species
EteRNA	2011	<a href="http://eterna.cmu.edu/web/">http://eterna.cmu.edu/web/</a>	Puzzles. RNA design and structure prediction
Foldit	2008	<a href="http://fold.it/portal/">http://fold.it/portal/</a>	Puzzles. Protein folding

*et al.*, 2012). In MalariaSpot, players mark malaria pathogens in a microscopy image including white blood cells, within a limited time (Luengo-Oroz *et al.*, 2012). Successful identifications give scores to players. The citizen sort project developed an online game “Happy match” to classify photographs of organisms (Crowston & Prestopnik, 2013). Playing EyeWire complements computational analysis for connecting neurons of the retina based on micrographs. Foldit (Cooper *et al.*, 2010) and EteRNA provide environments where players fold proteins to the correct shape or design RNAs through puzzles. In Phylo (Kawrykow *et al.*, 2012) and Fraxinus (MacLean, 2013) players highlight differences in nucleotide sequence alignments to either flag potential genetic diversity or create more accurate alignments.

Particularly where projects are of high interest to the general public, it is imperative that we harness their passion

and convert this into real contributions to scientific progress.

### The Facebook game Fraxinus

One disease that has been met with widespread anguish from the general public is Chalara dieback of ash, commonly known as ash dieback disease. The disease is caused by the fungus *Chalara fraxinea* which emerged in Poland in the early 1990s and has since spread rapidly across Europe from east to west. In September 2012, the first case of the disease in native woodland in the UK was reported just south of Norwich.

The vast media interest that ensued reflects the very real threat that is posed to our 80 million ash trees here in the UK. For instance, in Denmark, the loss of up to 90% of the ash tree population has been attributed solely to ash dieback disease.

With the ability to understand our adversary paramount, and the cost and speed of genome sequencing rapidly

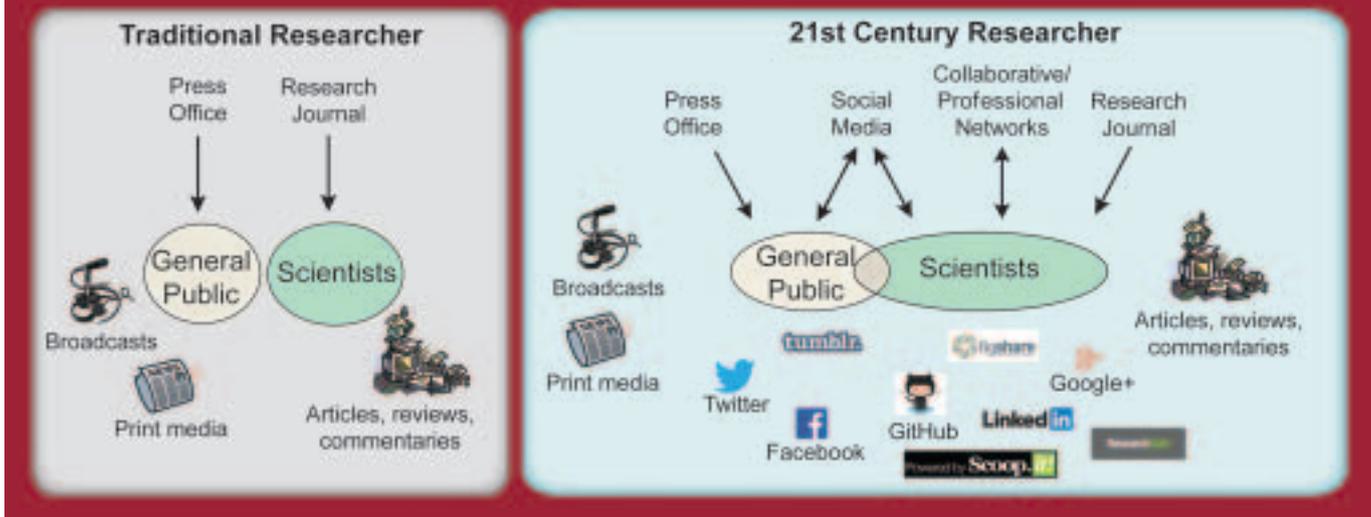
decreasing, generating and publicly releasing sequence data is key to accelerating research on emerging and re-emerging pathogens (Kamoun, 2012). In response, the OpenAshDieBack consortium was set up, led by Norwich researchers at The Sainsbury Laboratory, The John Innes Centre and The Genome Analysis Centre, and supported by BBSRC.

The central goal of this project is to generate and publicly release gene and genome sequence data for the pathogen and host that can be analysed by the worldwide scientific community (<http://oadb.tsl.ac.uk/>; accessed 23/09/2013). In addition, and in direct response to the widespread concern of the general public, the Fraxinus game was conceived as part of the crowdsourcing initiative.

By hosting Fraxinus on the Facebook platform, participants automatically send invitations across their individual social networks when playing the game, to encourage competition between

**Figure 1. The integration of social media into scientific communication**

As social media has grown in popularity numerous tools have been developed that can be utilised by scientists to promote communication both within the community and with the general public. Clip art images used with permission of Microsoft



friends and family. With more than a billion active Facebook users (<http://newsroom.fb.com/Timeline>; accessed 23/09/2013), this provides a substantial forum for the non-specialist audience to contribute biologically relevant analysis to the Fight Back Against Ash Dieback Disease.

Through the underlying OpenAshDieBack research project, an abundance of genomic and transcriptomic data was generated with the aim to assess differences between various UK pathogen isolates and host varieties at the genomic level. This genetic variation, in turn, can provide clues about how this pathogen may be causing disease and what makes a specific tree variety potentially resistant. However, computational pipelines that are commonly used to identify these types of variation are notoriously limited. In contrast, the human eye is much better at recognizing these types of patterns in large data sets.

In the Fraxinus game, players act to manually identify or curate any computationally problematic sequence variation in comparisons between pathogen isolates or host varieties. The corresponding reference genome appears as a series of coloured leaves, each representing a single nucleotide within a defined region of the pathogen or host genomes. The task is to match a series of short stretches of 21 coloured leaves from a second sample isolate against this reference genome sequence (Figure 2). Mismatches represent

sequence differences that are used to determine the genetic relatedness between the given sample and the reference isolate. As users steal patterns from their “friends” to improve the score associated with the alignment, the sequence variations are verified over and over again. This, in turn, provides robust analysis of the variation that is

identified, eventually outputting the analysis to a database that feeds back into the researchers’ analysis. With over 36,000 visits from approximately 18,000 unique visitors from 126 countries within the first week of its release, this game has really captured the enthusiasm of the non-specialist audience. The game contains 10,001 datasets that have all

**Figure 2. Fraxinus**

The aim is to match the 21 nucleotide stretches of sequence patterns from one sample against the target pattern at the top of the screen, which represents a region of the pathogen or plant genome sequences. Each nucleotide is represented by a single coloured leaf and mismatches relate to important genetic diversity between the sample and either the pathogen or tree genomes. The arrow buttons allow each pattern string to be moved left or right, while individual leaves can be moved, deleted or inserted with click-and-drag movements. As the match improves so does the “Pattern value” with players able to “Claim this Pattern” when they have the highest score



been played to some extent, with the analysis currently being transformed into real scientific results.

### The power of the crowd

The Fight Back Against Ash Dieback Disease is built on an effective crowdsourcing network, with world-renowned plant biologists and bioinformaticians at its core. To this aim, various forms of media have been used to promote engagement from a wide demographic. For instance, Twitter and traditional media have been used to spread awareness of the GitHub data repository amongst scientists to encourage their contribution to vital annotation. As a result, the analysis of genomic and transcriptomic sequences from the pathogen and ash tree has revealed many genomic features and repertoires of transcripts that are highly expressed during the infection stage ([http://oadb.tsl.ac.uk/?page\\_id=223](http://oadb.tsl.ac.uk/?page_id=223);

accessed 01/10/2013). Furthermore, in an effort to understand how the pathogen successfully infects ash trees, researchers have identified and classified genes that encode potential effector proteins (<http://oadb.tsl.ac.uk/?p=622>; accessed 01/10/2013). Effector proteins are secreted by an array of plant and animal pathogens to manipulate host physiology and help successful colonization.

Engagement with the general public has also been crucial to reporting cases of infected ash trees in the wild through the AshTag smartphone application, where users submit photographs of potentially diseased ash trees to help track its spread (<https://www.ashtag.org/>; accessed 01/10/2013). Employing a combination of social and traditional media is an extremely powerful approach that guarantees to reach the broadest of demographics.

### How will crowdsourcing be integrated into scientific research in the future?

Crowdsourcing will undoubtedly continue to increase transparency within the process of scientific discovery, whilst further promoting interdisciplinary collaboration. Computational scientists have long relied on discussion forums and data repositories to communicate. With “big data” becoming commonplace in all areas of research, we can look to our computational colleagues to continue to develop forums that include bench biologists and act as central transparent cross-disciplinary hubs.

The integration of social media has enabled scientists to start to dissolve barriers between research and the general public. However, this is dependent on scientists embracing the full power of the crowd, regardless of the perceived complexities such as issues of data “ownership”, complexity of authorship and the threat of being scooped. These issues can often be managed if addressed in the design stage of a project.

With the non-specialist audience forming a vital resource that can be utilized to accelerate scientific discovery, it is our responsibility to ensure that these issues are dealt with effectively to make the most of the power of the collective.

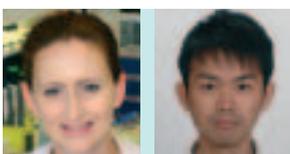
The integration of gaming technology to crowdsourcing initiatives enables the general public to help scientific progression whilst ultimately having fun!

As we look forward, we envisage a future where the largely untapped resource and collective intelligence of the general public is fully embraced in all aspects of research.

One continuing issue is how to maintain long-term engagement from the crowd. This remains a challenge and will require scientists to fully appreciate and be dedicated to crowdsourcing methodology, with active feedback to the crowd and releasing new and exciting updates to any gaming technology to maintain interest.

## references

- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovic, Z., and Players, F. (2010). Predicting protein structures with a multiplayer online game. *Nature*, **466**, pp756–760.
- Crowston, S., and Prestopnik, N. R. (2013). Motivation and data quality in a citizen science game: a design science evaluation. 46th Hawaii International Conference on System Sciences (HICSS).
- Darling, E. S., Shiffman, D., Cote, I. M., and Drew, J. A. (2013). The role of Twitter in the life cycle of a scientific publication. *Ideas in Ecology and Evolution*, **6**, pp32–43.
- Kamoun, S. (2012). Genomics of emerging plant pathogens: too little, too late. *Microbiology Today*, p140.
- Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., Zarour, E., Players, P., and Sarmenta, L. (2012). Phylo: a citizen science approach for improving multiple sequence alignment. *PLoS One*, **7**, e31362.
- Luengo-Oroz, M. A., Arranz, A., and Frean, J. (2012). Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. *Journal of Medical Internet Research*, **14**, e167.
- Maclean, D. (2013). Cutting edge: Changing the rules of the game. *eLIFE*, **2**.
- Mavandadi, S., Dimitrov, S., Feng, S., Yu, F., Sikora, U., Yaglidere, O., Padmanabhan, S., Nielsen, K., and Ozcan, A. (2012). Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study. *PLoS One*, **7**, e37245.
- Priem, J., Costello, K., and Dzuba, T. (2012). Prevalence and use of Twitter among scholars.
- Terras, M. (2012). The impact of social media on the dissemination of research: results of an experiment. *Journal of Digital Humanities*, **1**, p3. ([http://figshare.com/articles/Prevalence\\_and\\_use\\_of\\_Twitter\\_among\\_scholars/104629](http://figshare.com/articles/Prevalence_and_use_of_Twitter_among_scholars/104629))



**Diane G.O. Saunders** (left) and **Kentaro Yoshida** (right)  
The Sainsbury Laboratory, Norwich Research Park, Norwich, NR4 7UH, UK.