

Bioimage informatics

The QDREC web server: determining dose–response characteristics of complex macroparasites in phenotypic drug screens

Daniel Asarnow^{1,†}, Liliana Rojo-Arreola², Brian M. Suzuki^{2,3},
Conor R. Caffrey^{2,3} and Rahul Singh^{1,3,*†}

¹Department of Computer Science, San Francisco State University, San Francisco, CA, USA, ²Department of Pathology and ³Center for Discovery and Innovation in Parasitic Diseases, University of California, San Francisco, San Francisco, CA, USA

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first and last authors should be regarded as Joint First Authors.

Associate Editor: Robert F. Murphy

Received on April 30, 2014; revised on November 17, 2014; accepted on December 10, 2014

Abstract

Summary: Neglected tropical diseases (NTDs) caused by helminths constitute some of the most common infections of the world's poorest people. The etiological agents are complex and recalcitrant to standard techniques of molecular biology. Drug screening against helminths has often been phenotypic and typically involves manual description of drug effect and efficacy. A key challenge is to develop automated, quantitative approaches to drug screening against helminth diseases. The quantal dose–response calculator (QDREC) constitutes a significant step in this direction. It can be used to automatically determine quantitative dose–response characteristics and half-maximal effective concentration (EC₅₀) values using image-based readouts from phenotypic screens, thereby allowing rigorous comparisons of the efficacies of drug compounds. QDREC has been developed and validated in the context of drug screening for schistosomiasis, one of the most important NTDs. However, it is equally applicable to general phenotypic screening involving helminths and other complex parasites.

Availability and implementation: QDREC is publically available at: <http://haddock4.sfsu.edu/qdrec2/>. Source code and datasets are at: <http://tintin.sfsu.edu/projects/phenotypicAssays.html>.

Contact: rahul@sfsu.edu.

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Neglected tropical diseases caused by helminths, such as schistosomiasis, lymphatic filariasis and onchocerciasis, constitute some of the most common infections of the world's poorest people. Schistosomiasis in particular has been identified as the second most socioeconomically devastating illness after malaria. It afflicts over 200 million people worldwide, and its severe symptoms have been implicated as prime factors depriving the affected populations of their health and economic potential. Few drugs are available for

treatment of helminth diseases. For instance, just one, praziquantel, is available for treatment of schistosomiasis. Emergence of drug resistance is therefore a concern, and there is an urgent need for discovery of new therapies against helminthic illnesses ([World Health Organization, 2013](#)).

Drug screening against helminths poses unique challenges: unlike model organisms like *Caenorhabditis elegans*, clones or lines of worms are not available, the parasites tend to be non-amenable to target identification using gene knockouts or RNAi, and few

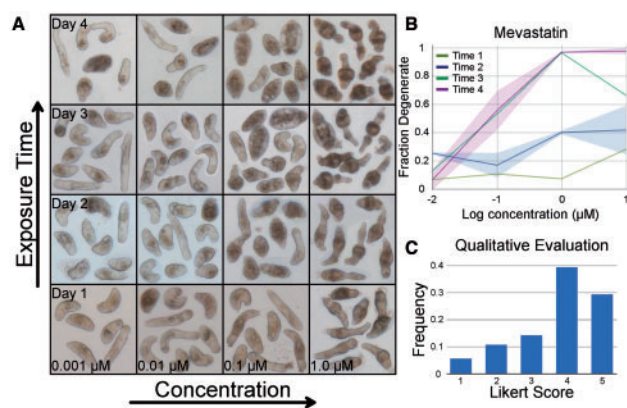


Fig. 1. QDREC results and user evaluation. **(A)** Varying responses of schistosomula to the anthelmintic drug, niclosamide, with increasing concentration and exposure time. Note the complex phenotypic changes that can occur even due to the action of a single drug. **(B)** Dose–response curves displayed in QDREC for the statin drug mevastatin over four concentrations and exposure periods. For replicates, median responses are shown with a confidence region of one standard deviation. **(C)** Histogram of Likert scores (a score of 5 indicates complete agreement) from the user evaluation, indicating that quantal response values determined by QDREC are consistent with those determined by the users themselves

molecular mechanisms of host–parasite interactions are well elucidated. Consequently, lead identification typically involves *phenotypic screening*, where the parasites are exposed to compounds and leads identified by analyzing the ensuing effects. Such analysis is non-trivial; drug action typically elicits multiple complex phenotypes that can overlap and vary over both time and concentration (see Fig. 1A). Therefore, analysis of phenotypic screening has tended to employ manual examination, making the process low-throughput, subjective and non-quantitative.

1.1 Computational challenges

The area of automatic analysis of helminth phenotypes is a complex challenge. In this area, recent progress has been made in parasite segmentation (Asarnow and Singh, 2013; Singh et al., 2009), parasite tracking from video recordings (Saha and Singh, 2012) and quantitative definitions and analysis of helminth phenotypes (Lee et al., 2012; Singh et al., 2009), including hit detection in high-throughput screens (Marcellino et al., 2012; Paveley et al., 2012). As yet, however, no automated method exists for determination of time- and dose–response characteristics, including EC₅₀ values, based on the output of image-based, phenotypic screens.

2 Quantal dose–response calculator

Quantal dose–response calculator (QDREC) is a web server for automatically determining quantal time- and dose–response characteristics from image-based, phenotypic drug screens. Given a set of images recording the time-course exposure of a parasite population, QDREC image analysis and supervised learning are used to identify parasites affected by a drug at a given concentration/time point.

The problem of determining the effect of a drug is addressed using a quantal approach: at each concentration/time point, the treated and control populations are compared based on the exhibited phenotypes, and the affected parasites are distinguished using a classification formulation. The number of affected parasites provides a quantal measure of drug effect. Details on the algorithmic underpinnings of QDREC can be found in the supporting

information, and a listing of user interaction and processing steps is included in [Supplementary Table S1](#).

2.1 Design issues and software architecture

QDREC is envisaged to be used by researchers across the world, provide ease of use and require minimal efforts on maintenance. Its web-services model was selected to meet the above goals and strike a balance between issues such as data transfer bandwidth, computational power for data analysis and the expertise needed to run and maintain complex software. In QDREC, image compression is used to mitigate possible bandwidth limitations, while remote execution relieves the need to have local processing power or software engineering expertise. The web-services model also permits seamless software upgrades for all users. Finally, QDREC includes documentation for local execution on all major OS platforms.

2.2 Method and workflow

QDREC involves four major steps: data upload and parasite segmentation, descriptor generation, classification of parasites using supervised learning and quantification of dose–/time– response. In the first step, QDREC accepts images recording the responses of parasite populations to varying experimental conditions (including compound, concentration and exposure duration), and negative control images, which record the phenotypes of parasites that have not been exposed to any compounds. Currently, QDREC works with static images, which constitutes the most common data capture mode in phenotypic screening. Next, the parasites in the images are segmented using the algorithm described in (Asarnow and Singh, 2013), which was designed to surmount difficulties presented by schistosomula in phenotypic screens, including variation in morphology, texture and behavior, as well as presence of multiple touching parasites. The segmentation algorithm avoids *a priori* shape models in favor of a signal-driven bottom-up approach, employing a novel edge classifier to split any foreground regions containing multiple parasites. Default parameters are optimized specifically for segmentation of schistosomula in unstained, bright-field micrographs of whole wells in 24-well plates at 5× magnification. Additional segmentation methods provided in the software are summarized in [Table 1](#). QDREC provides users the opportunity to review the segmented images and to repeat segmentation with user-specified parameters if so required.

Parasite phenotypes are quantified using 71 image-based features (Lee et al., 2012). These features are generally robust to changes in illumination and are not specifically tailored to schistosomula. Therefore, they may be used for describing other organisms. A fuller description of these features is provided in the supporting information.

To ensure that the controls exclude unhealthy parasites, the aforementioned feature vectors are used to exclude non-normal parasites from the control population. Further, to account for natural variation between parasite subpopulations (which can be significant), each parasite is represented both by its own feature vector as well as one representative of the control parasites from the sample it was drawn from. In the classification stage, either the feature vectors (one-stage classification) or feature vector tuples (two-stage classification) are used to assess whether or not a given parasite differs significantly from its controls. See the supporting information Section 4.5 for details on the relative merits of one- and two-stage classification in QDREC.

QDREC provides a two-stage, non-linear Support Vector Machine classifier (NL-SVM) trained using 5511 parasites

Table 1. Segmentation methods supported in QDREC

Algorithm	Brief description	OCD	MBD
Asarnow-Singh	Regions containing parasites are identified using a region-based distributing function. Phase congruency and grayscale thinning localize potential boundaries between touching parasites, and the watershed transform identifies the minimum edge set separating individual parasites, which may have been merged during the initial segmentation.	0.041	1.32
Canny	Segmentation using the Canny edge detector instead of phase congruency.	0.074	1.85
Watershed	Initial segmentation of Asarnow-Singh combined with Watershed segmentation.	0.407	9.16
Arbitrary	Users upload segmented images obtained with any algorithm of their choice.	N/A	N/A

See [Asarnow and Singh \(2013\)](#) for detailed comparisons of the first three methods. Object count discrepancy (OCD) is the proportion of parasites not correctly segmented, mean boundary deviation (MBD) is the average distance (in pixels) between true and segmented boundary contours. The Asarnow-Singh segmentation method has the best performance, identifying ~96% of the parasites and localizing their boundaries to within 1.3 pixels on average.

Table 2. Performance of QDREC using two-stage classification

Classifier	Classification accuracy		Dose–response curves		
	Cross-validated	Test	NRMSE	Correlation	Log ₁₀ (<i>P</i> -value)
SVM (RBF)	0.907	0.905	0.13 ± 0.14	0.984 ± 0.028	−2.91 ± 1.17
SVM (linear)	0.910	0.904	0.14 ± 0.15	0.982 ± 0.029	−2.92 ± 1.27
Naïve Bayes	0.747	0.710	0.79 ± 1.61	0.896 ± 0.129	−1.83 ± 1.07
Random Forest	0.901	0.895	0.16 ± 0.13	0.986 ± 0.011	−2.80 ± 0.98

Classification accuracy and the mean and standard deviation of NRMSE, correlation and log₁₀(*P*-value) are shown for 12 drugs (atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin, simvastatin, clocastel, ibandronate, K11777, niclosamide, praziquantel and sorafenib). Correlation *P*-values were determined using Student's *t*-distribution. The non-linear SVM is used in QDREC as the default classifier because it exhibited less parameter sensitivity and more reliable convergence than the linear SVM. All results are for two-stage classification using parasite and control feature vector tuples.

annotated by four human experts, whose training decisions were combined using majority voting. Alternatively, users may choose to employ a point-and-click interface to annotate individual parasites and train their own one- or two-stage classifier instances using one of the following classification methods provided in the software: linear SVM, non-linear SVM (NL-SVM), Naïve Bayes or Random Forest. The 10-fold cross-validated classification accuracy over the training set for the NL-SVM was found to be 90.7%. The cross-validated accuracies for the other classifiers are given in [Table 2](#) and [Supplementary Table S3](#), for two-stage and one-stage classification, respectively.

Finally, parasite classifications are used to compute the quantal phenotypic response for each test image, as well as the aggregate response for each unique combination of experimental conditions. Results are then displayed in both tabular and graphical form. [Figure 1B](#) shows an example dose–response plot produced by QDREC.

2.3 Experiments and validation

QDREC has been validated using a blind test set of 5798 schistosomula, which were expert annotated, but not used to train the provided classifier(s). Performance was evaluated by classification accuracy, as well as through direct comparison of manual and automatic dose–response curves using normalized root-mean-square error (NRMSE) and Pearson correlation. Classification accuracy for the provided NL-SVM classifier was 90.5%, while dose–response curves exhibited a correlation of 0.984 (*P*-value 0.0012) and an NRMSE of 0.13. These results indicate that the parasite classifications and dose–response curves produced by QDREC are comparable to those obtained by experts. Complete results, including the

other classification frameworks, are given in [Table 2](#) and [Supplementary Table S3](#), for two-stage and one-stage classification, respectively. The relative performances of one- and two-stage classification for larger, more diverse datasets is explored using additional data consisting of 23 867 schistosomula imaged with variable lighting, well density and depth of focus. The results of this analysis are given in [Supplementary Table S4](#) and [Table S5](#), for one- and two-stage classification, respectively.

As a final assessment, QDREC was used to analyze 118 images of schistosomula exposed to compounds not used in the previous experiment. A survey was conducted to determine how closely 10 experts agreed with the QDREC output on a five-point Likert scale. Participants were shown corresponding pairs of control and drug-exposed images, along with the per-image quantal-response scores obtained using QDREC and asked to quantify their assessment of QDREC on the aforementioned Likert scale. A histogram of the resulting scores is presented in [Figure 1C](#). Although some disagreement was found (due to segmentation and classification errors as well as disagreement between the human experts), the proportion of such cases was small (<30%). The distribution of scores underlines that results from QDREC are consistent with user assessments of phenotypic response.

Funding

This work was supported in part by the National Institutes of Health [grants 1R01AI089896, R21AI107390] and by the National Science Foundation [grant IIS-0644418]. L.R.-A. was supported by a UC MEXUS-CONACYT Postdoctoral Research Fellowship.

Conflict of Interest: none declared.

References

- Asarnow,D.E. and Singh,R. (2013) Segmenting the etiological agent of schistosomiasis for high-content screening. *IEEE Trans. Med. Imaging*, **32**, 1007–1018.
- Lee,H. *et al.* (2012) Quantification and clustering of phenotypic screening data using time-series analysis for chemotherapy of schistosomiasis. *BMC Genomics*, **13**, S4.
- Marcellino,C. *et al.* (2012) WormAssay: A novel computer application for whole-plate motion-based screening of macroscopic parasites. *PLoS Negl. Trop. Dis.*, **6**, e1494.
- Paveley,R.A. *et al.* (2012) Whole organism high-content screening by label-free, image-based Bayesian classification for parasitic diseases. *PLoS Negl. Trop. Dis.*, **6**, e1762.
- Saha,U. and Singh,R. (2012) Vision-based tracking of complex macroparasites for high-content phenotypic drug screening. In: Bebis,G. *et al.* (eds.), *Advances in Visual Computing/Lecture Notes in Computer Science*. Springer, Heidelberg, pp. 104–114.
- Singh,R. *et al.* (2009) Automated image-based phenotypic screening for high-throughput drug discovery. In: *Proc. 22nd IEEE International Symposium on Computer-Based Medical Systems*, 2009. CBMS 2009. IEEE, pp. 1–8.
- World Health Organization (2013) Sustaining the drive to overcome the global impact of neglected tropical diseases: second WHO report on neglected tropical diseases. World Health Organization, Geneva, Switzerland.