# An Interactive Visualisation for Selecting PCR Primers

**Paul Rutherford, Clare Churcher and John McCallum**[†]

Applied Computing Group
P.O. Box 84, Lincoln University
Canterbury, New Zealand

Rutherp3/Churcher@Lincoln.ac.nz

[†] Crop and Food Research
Private Bag 4704, Christchurch
New Zealand

McCallumJ@crop.cri.nz

## Abstract

[1] Biological scientists often need to design primers. These are short pieces of DNA used for copying sections of a DNA sequence using the polymerase chain reaction (PCR). Primers have various properties that influence the likelihood of success and scientists wish to quickly and easily select the best primers for amplification of a target sequence. Current web-based primer design software (Primer3) requires the user to specify ranges for the primer properties and then generates and returns suitable primers. Depending on the range of the criteria this can often result in no suitable primers being found or too many for easy scrutiny. The only option for the user is to adjust the criteria, resubmit the query and hope this will result in manageable number of suitable primers being returned.

In this paper we describe a visualisation tool that allows the user to interactively explore the primers suggested by Primer3. The user can soften the criteria ranges submitted to Primer3 to ensure a result and then experiment with adjusting the criteria in order to quickly find the most suitable primers. Initial user trials indicate that the prototype tool is successful in facilitating the choice of suitable primers for a given sequence.

*Keywords*: Information Visualisation, Primer design, PCR, Dynamic Exploration

## 1 Introduction

Biological scientists often need to design primers for copying sections of a DNA sequence using the polymerase chain reaction (PCR) process (Klug and Cummings, 2000). These primers are short pieces of DNA and have various properties that influence the likelihood of success in the reaction. In order to ensure success the primers must be carefully chosen with respect to these properties.

Primer3 (Rozen and Skaletsky, 1996, 1997, 1998) is a web based tool that can be queried to suggest primers that meet particular criteria. Primer3 allows the user to specify boundaries for several criteria including the temperature at which the primer is operational and sensitivity of the primers in the PCR process. One of the input windows to Primer3 is shown in Figure 1.

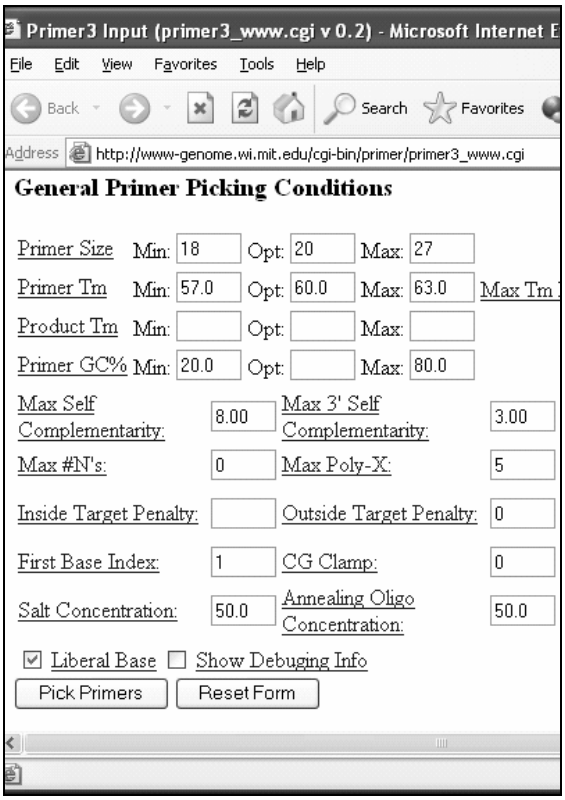


**Figure 1: Part of the Primer3 web interface showing one of the sets of criteria which can be set. Others can also be specified**

Primer3 then generates a list of possible primers. Primers that do not meet all the criteria are disregarded while those that satisfy the criteria have a score calculated and are displayed in textual format in an output window. Typically, no primer is consistently better than all others (Kampke *et al.*, 2001) and the scientist's domain knowledge is used to determine a final choice.

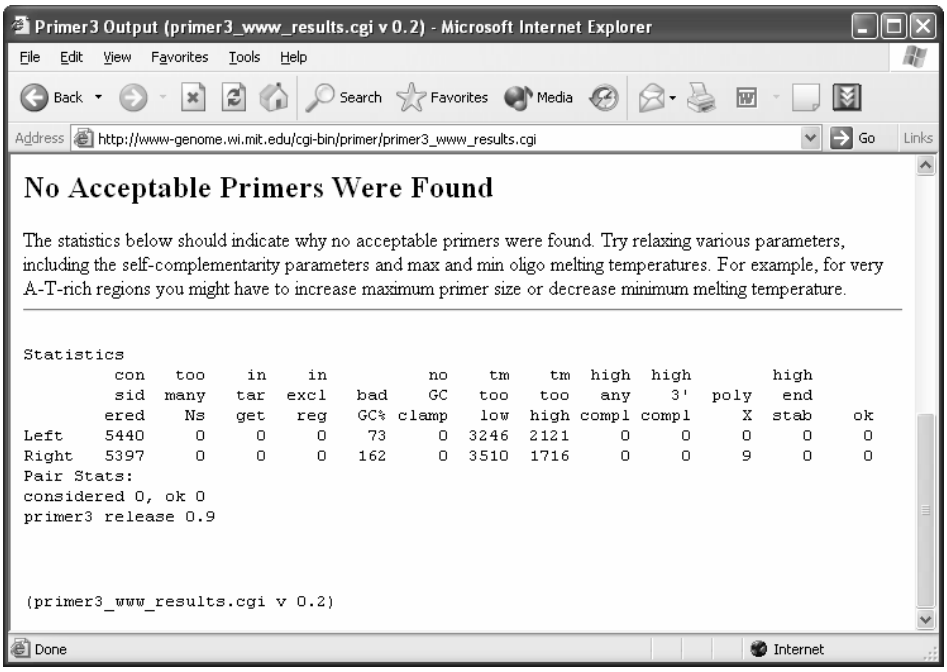


**Figure 2 – Results page from Primer3 web interface, showing no acceptable primers were found.**

Typically when users formulate criteria for Primer3 they will specify a standard set of criteria that is idealistic. This is to limit the number of primers that will be suggested and to keep them within the user's ideal range of values. However, often these restrictions are too strict and no primers are suggested. The user must then reformulate the query and try again. There is little guidance on how to reformulate the query (see Figure 2), so it can take many attempts to adjust criteria to find acceptable primers.

Even in the case where a selection of primers are returned, the user has no way of knowing if a slight adjustment of one criterion might have produced a more satisfactory result.

In this paper we discuss: issues involved with information retrieval from databases; describe Primer3 Explorer our application to aid dynamic exploration of primer data; and report user feedback.

## 2 Querying Databases

Traditional information retrieval assumes that the user is familiar with the database being queried, will come with a 'precisely formulated' query and a satisfactory answer will be provided (Spence and Tweedie, 1998). Primer3 takes the target sequence and uses this to dynamically determine possible primers. It then compares the properties of these generated primers with the criteria provided by the user. While the user can specify his criteria precisely, he is unable to predict the range of property values of the primers generated.

The optimal outcome is that a small number of acceptable primers are suggested. The scientist can then review the textual output and use his domain knowledge to choose the most suitable primer. If too few or too many primers are returned, the user will need to reformulate the query. If the query returned too many suggestions the user may have some idea on how to reform the query, but if no responses are returned the user can only guess what modifications to make.

Reformulating the query is difficult because the constraints specified are hard constraints, i.e. possible primers either meet the criteria or they do not. If the possible primers do not meet the criteria, they are not presented to the user. It would be more useful if soft criteria could be used so that possible primers not strictly meeting the criteria are still presented to the user, possibly identified in some way.

### 2.1 Browsing

Browsing is described by Spence as an assessment of content where the goal is not clearly defined (1999, cited by McCarthy and Bergel, 2003). It would be useful for Primer3 users to be able start with broad criteria and then explore the larger set of results returned. The user could then dynamically adjust the criteria to produce smaller sets of items for detailed consideration.

### 2.2 Dynamic Exploration

For dynamic exploration an interface is required that allows the user to see the complete set of results and to interact with it. The user needs to be able to filter the results according to various criteria and immediately see the results of the changes. Slider controls are good for this as they allow the user to quickly and easily specify the limits they'd like to place on a range of values.

When any adjustment is made to the filter parameters (through manipulations of the filter controls) the response needs to be immediate. That is, the effect must occur within 0.1s of its cause. This supports dynamic exploration, often called the "What if?" activity (Spence and Tweedie, 1998). The advantages of the dynamic exploration method are that the filter adjustment process will be:

- Incremental: users can see the effects of small adjustments to a parameter
- Rapid: users see the results without returning to a previous form, adjusting it, resubmitting, and waiting for a response
- Reversible: if the user adjusts a parameter too far it is easily undone.

This will allow effective and efficient experimentation and investigation to occur.

Another objective of dynamic exploration is to help the user know how to refine subsequent queries (Spence and Tweedie, 1998). This can be achieved by maintaining an overview of all data items while highlighting those satisfying the filter criteria in some way. In this way the user can see the 'near misses'

This is similar to the Dynamic HomeFinder application developed by Christopher Ahlberg and Ben Shneiderman (1994).

## 3 Primer3 Explorer

In this section we describe our prototype application, Primer3 Explorer, which has been developed using VTK (Schroeder *et al.*, 1998) and Tcl/TK (Ousterhout, 1994). Primer3 Explorer assists the user to interactively determine suitable primers for a given sequence. The data to be explored is retrieved by submitting a query with broad criteria to Primer3.

### 3.1 Displaying multi-dimensional data

In order to visually explore the different properties of the primers returned from Primer3 we need to display the items in some type of space. Many applications use a physical spatial mapping to display data, but there is no obvious spatial layout for this data that will be both useful and meaningful. Instead, we create a display by allowing the user to select different properties of the primers to be used as axes (Ahlberg and Shneiderman, 1994).

Figure 3 shows a display where Minimum and Maximum temperature form the axes for displaying the primers. Each primer is represented by a box. Boxes are a convenient representation because they are small, scalable, highly visible, can be colour-coded, and can be displayed rapidly (Ahlberg and Shneiderman, 1994).

Each point's location encodes two properties. A third property can be represented on the plot by modifying another property of the box. We chose to use colour, as a colour sequence may be used to represent a continuously varying range of values and is easily separable from location (Ware, 1999).

To facilitate reading of colour 'values' a scalar bar displaying the mapping from colour to value is displayed next to the plot (see Figure 3).

We allow the user to choose the properties that are mapped to axes and colour.

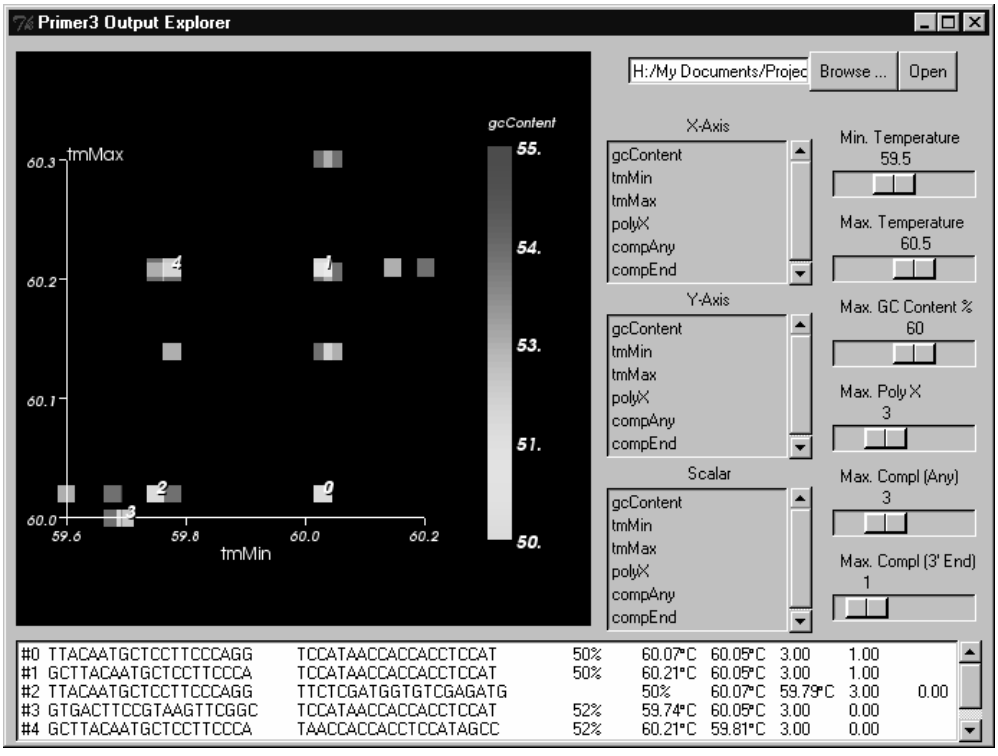


**Figure 3 – The plot from the Primer3 Explorer. Each box represents a primer. The user can choose attributes to be represented by the axes and by colour (middle panel). As the ranges for the attributes are varied (with the sliders) those items which are excluded are greyed out. Details of the acceptable items are shown in the scrollable text window at the bottom.**

### 3.2 Filtering the Data

The query submitted to Primer3 has deliberately broad criteria to allow a number of primers to be returned for further exploration. Sliders allow the user to refine the query by adjusting the ranges of those properties most commonly specified when submitting a query to Primer3. Each slider presents either the minimum or maximum value for the range of values for that property. The query is the result of ANDing all these conditions together. The program differentiates those primers not conforming to the new set of criteria by greying them out. Retaining them in the visualisation helps guide the user in subsequent adjustments.

The interactions provided support dynamic exploration of the data set by allowing rapid, iterative and reversible changes to filter criteria. As the user adjusts the filter controls, the group of primers that meet the criteria immediately changes in the display. From this, the user can see how many primers are selected and the values of those properties represented by the axes and colour. If no primers meet the criteria, the presence of unselected points will guide the user in adjusting the filters to find some.

When the user has narrowed the data down to a small set of primers he will want to select one of them. Detailed information about selected primers is displayed in a text window for visual inspection and to facilitate the final choice.

## 4 User experience

In order to get some feedback on our prototype applications we informally observed two scientists who are experienced at using the Primer3 web interface (Rozen and Skaletsky, 2000) as shown in Figure 1.

Primer3 is well used by the bioinformatics community. The scientists we spoke with were pleased with the success of primers they have selected from those suggested by Primer3. However, they reported having to return to the query form frequently to reformulate queries that had returned no primers or primers that were not acceptable for various reasons. A lot of time is spent "going backwards and forwards [adjusting and re-querying Primer3]" in order to receive any suggestions. When they have a number of possible primers, the primer sequence and properties are visually inspected to determine acceptability. This was described as requiring "eye-balling vast amounts of text" to make a final decision. Often the first acceptable primer found in the reported set was selected without exploring the remainder for a possibly better candidate.

In our informal trial we observed the scientists searching for three different primers: these included one sequence for which we knew primers would easily be found and one that we expected would be quite difficult. We asked them to first use the Primer3 web interface to find a candidate primer and then to use our application which had been set up with results from a Primer3 query with broad criteria.

The interface of our visualisation was briefly explained to the users, and they then used it for themselves. They quite quickly warmed to the idea of exploring the data, and felt free to manipulate the sliders in order to reduce the set of suggested primers to a small number. The process was described as "easy" and they liked the idea of dynamically manipulating the criteria to narrow choices down to a small set.

The users liked the fact that they did not need to focus on the detail (the actual values) of the criteria as they manipulated the sliders. Being able to see data visually as opposed to tables of numerical data was described as advantageous.

Interestingly the users did not change the properties being mapped to the axes or colour to any extent. They found that simply adjusting the sliders to reduce the number of coloured boxes was sufficient to quickly find a result. Indeed the position of the box on the plane and its colour did not seem to affect the decision making. This may be due to the fact primers were relatively easily found. Once a small number of boxes had been isolated the user's attention turned immediately to the details in the text window for a final choice. Further investigation is needed to determine whether the placement of the boxes and the use of colour could be utilised more effectively, perhaps for indicating the degree to which the overall criteria are met.

In using the web interface to Primer3 the users had to resubmit their query more than once in half the cases in order to find any satisfactory primers. In one case the user was not particularly happy with the choices offered but said they would make do rather than going to the effort of resubmitting a query.

Using our application the users were able to quickly find an appropriate primer in every case. One user repeatedly stated they found primers they preferred to the ones they had chosen using the web interface to Primer3.

The users commented some additional primer properties such as primer length could be usefully included in the display and suggested some modifications to the layout for the text in the detail window. One user was "convinced it's an idea worth pursuing further". Further work could integrate the application with Primer3.

## 5 Conclusion

We have developed a prototype application to allow users to interactively explore primers returned from Primer3. Users provide broad criteria to Primer3 and then dynamically adjust the criteria boundaries in order to reduce the number of valid primers to a manageable size for closer investigation. Users found they were successful in finding suitable primers simply by adjusting the criteria to reduce the number of primers selected and then investigating the textual detail. They seldom made use of the spatial location of the items on the Cartesian plot nor the actual colour of the boxes. Further testing need to be done to determine how best to utilise these aspects of the display.

## 6 Acknowledgements

The authors wish to acknowledge the assistance of the New Zealand Foundation for Research, Science and Technology (C02X0203)

## 7 References

Ahlberg, C. and Shneiderman, B. (1994). *Visual Information Seeking*: *Tight Coupling of Dynamic Query Filters with Starfield Displays.* Proceedings of ACM CHI '94 Conference.

Kampke, T., Kieninger, M. and Mecklenburg, M. (2001). *Efficient primer design algorithms.* Bioinformatics **17**(3): 214-225.

Klug, W. S. and Cummings, M. R. (2000). *Concepts of Genetics*, 6th ed. New Jersey, Prentice-Hall.

McCarthy, J. and Bergel, P. (2003). *Search Without Keywords*. CHI Workshop: Best Practices and Future Visions for Search User Interfaces, CHI 2003, Fort Lauderdale, Florida.

Ousterhout, K. K. (1994). *Tcl and the Tk Toolkit*. Massachusetts, Addison-Wesley Publishing Company.

Rozen, S. and Skaletsky, H. J.,1996, 1997, 1998, *Primer3 Release 0.9*, Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html

Rozen, S. and Skaletsky, H. J. (2000). *Primer3 on the WWW for general users and for biologist programmers.* Bioinformatics Methods and Protocols: Methods in Molecular Biology. M. S. Krawetz S. New Jersey, Humana Press**:** pp.365-386.

Schroeder, W., Martin, K. and Lorensen, W. (1998). *The Visualisation Toolkit: an object oriented approach to 3D graphics.* New Jersey, Prentice-Hall.

Spence, R. (1999). *A framework for navigation*. International Journal of Human-Computer Studies **51**(5): 919-945.

Spence, R. (2001). *Information Visualization.* New York, ACM Press.

Spence, R. and Tweedie, L. (1998). *The Attribute Explorer: information synthesis via exploration.* Interacting with Computers **11**: 137–146.

Ware, C. (1999). *Information Visualization: design for perception.* San Francisco, Morgan Kaufmann Publishers.