Dexter User’s Guide

# Introduction

Dexter is an open-source Java program that supports exploration, analysis, visualization, and comparison of time-series gene expression data sets. The software is freely available for non-commercial use under revision 3 of the GNU General Public License.

# Installation and Execution

The Dexter archive file should be downloaded into an empty directory and its contents should be extracted. Extraction creates an executable jar file called Dexter.jar and 4 directories: src, bin, lib, and data.

Execution consists of two phases: a wizard phase and an analysis phase. During the wizard phase, time-series expression data seta and supplementary data files are imported, and a “reference schedule” is designed. Results of the wizard phase are stored in session files with the extension “.dex”. During the analysis phase, you analyze and explore data stored in session files.

To execute, cd into the extraction directory and type “java –jar Dexter.jar”. Dexter displays an initial window with 3 buttons: “Wizard”, “Analysis”, and “Quit”. Clicking “Wizard” starts the wizard phase, at the end of which imported data is saved in a session file and you have the option of continuing to the analysis phase. Clicking “Analysis” in the initial window bypasses the wizard phase and presents a file browser for selecting a previously created session file, which will be immediately analyzed.

# Wizard Phase

The wizard phase supports importing expression and supplementary data, and the creation of a “reference schedule” (Figure 1).



Figure - Wizard phase block diagram

The wizard phase proceeds in 8 stages: Import studies, Select studies, Duration, Dark-light phases, Align schedules, Orthologs, Coregulation, and Wrap. A roadmap strip at the top of the wizard screen (Figure 2) shows the current stage and navigates to the next stage on completion of the current stage. The roadmap also allows navigation back to previous stages.



Figure - Wizard roadmap. The current stage is highlighted.

## Wizard Stage: Import Studies



Figure - Import studies phase, unpopulated.

In the Import studies stage (Figure 3) begin by clicking the “Open spreadsheet…” button to browse for a data set spreadsheet file, which must be in csv (comma-separated values) or tsv (tab-separated values) format. The screen becomes populated with a representation of the first few rows of the spreadsheet (Figure 4).



Figure - Import studies phase, populated.

The “Organism” menu (circled in red) is used to specify the organism associated with the study; specification of an organism is mandatory.

Below the “Organism” menu is the representation of the spreadsheet. Select one of the radio buttons (circled in blue) to specify the division between column titles and data cells. Toggle buttons (circled in green) are used to select which columns of the spreadsheet are to be used by Dexter. Selected columns are represented at the bottom of the screen by menus (circled in purple). Use the menus to specify the role of each column. A role is one of Gene I.D., Gene name, Annotation, KEGG Pathway, and Timepoint.

Additional information may be supplied via supplemental spreadsheets. For example, some data sets only contain gene i.d.s and timepoints; you might want to provide a supplemental spreadsheet to provide annotations and/or KEGG pathways. A supplemental spreadsheet must contain a column with i.d.s whose values match the i.d.s in the original data set; however, rows in the supplemental spreadsheet need not be in the same i.d. order as rows in the original spreadsheet, and supplemental values need not be provided for all rows in the original spreadsheet. As with the original spreadsheet, use the wizard to specify roles for the columns of the supplemental spreadsheet.

When all data has been imported for a study, click the “Apply” button. You can then import more studies or move on to the next stage of the wizard.

## Wizard Stage: Select Studies

The second wizard stage presents a checklist of all studies that have ever been imported. Check the ones you want to analyze.

## Wizard Stage: Reference Schedule Duration

Since there is no guarantee that the studies you have chosen all followed the same experimental design with regard to experiment duration, timing of dark/light phases, and selection of timepoints, the next 3 wizard stages guide you through the creation of a “reference schedule”, an idealized experiment design to which imported actual experiment results can be mapped.

In the Duration phase (Figure 5), set the duration in hours of the reference schedule by dragging the pin left or right. Duration is rounded to the nearest hour.



Figure - Reference Schedule Duration stage.

## Wizard Stage: Reference Schedule Dark/Light Periods

In the phase the reference schedule duration is fixed. Use the menu (“4 phases” in Figure 6) to specify the number of dark and light periods. Use the radio buttons to specify whether the initial period should be dark or light. The drag the pins to set the durations (in hours) of the dark and light periods.

When this phase is completed, the image on the screen is a background on which Dexter will paint all expression signature graphs.



Figure - Reference Schedule Dark/Light Periods phase.

## Wizard Stage: Align Schedules

In the next phase, horizontal (time) positions are assigned to each timepoint column that was imported during the first wizard phase.



Figure - Align Schedules Wizard Stage, initial appearance.

On entry to this wizard phase (Figure 7), the reference schedule appears at the top of the screen; each imported study is represented by a horizontal band at the bottom of the screen. Each study should be aligned to the reference by clicking the corresponding “Align” button. Clicking “Align” in a band adds alignment tools to the screen (Figure 8).



Figure - Align Schedules Wizard Stage, about to align a study.

The light blue band contains vertical lines representing the timepoint columns for the study being aligned. Each timepoint column must be mapped to a position in the reference schedule. To map a timepoint column to the reference, click near the timepoint and drag a line to the desired position in the image of the reference schedule. Figure 9 shows the result of mapping the first and last columns.



Figure - Align Schedules Wizard Stage, with first and last timepoints mapped to reference schedule.

At this point the remaining unmapped timepoints need to be mapped to the reference schedule. This can be accomplished by dragging each timepoint individually, or by clicking the “Interpolate” button to cause all unmapped timepoints to be mapped to the reference by linear interpolation (Figure 10).



Figure - Align Schedules Wizard Stage, with all timepoints mapped to reference schedule.

When a study has been aligned, a green check mark appears at the right of its band. When all studies have been aligned, Dexter has the information it needs to display all time series data on the reference schedule.

## Wizard Stage: Orthologs

In this stage, the wizard provides a browser for loading one or more ortholog files. Two file formats are supported:

* One orthologous group per line. Each line is a comma-separated list of gene i.d.s.
* Tabular BLAST results. Queries should be imported genes, with gene i.d. as the first element of each fasta defline. The search should be performed against a database of the sequences of all imported genes.

## Wizard Stage: Coregulation

In the Coregulation stage, a list of operon predictions can be loaded for each organism associated with an imported study. The files should be in the format supported by <http://www.microbesonline.org/operons/OperonList.html>, and at present Dexter has only been tested with files from that source. To download an operon list file, click on the link for the desired organism, then click on “A tab-delimited version of this table” near the top of the page. After a file has been loaded into Dexter, you can inspect the predictions by clicking the “Inspect” button for the associated organism. A dialog will display a list of all predicted operons. Since the imported study might not include all genes of an organism, it is possible that expression data is not available for all genes of an operon. In the dialog, genes that are represented in the data set are marked with a green check mark; genes that are not represented are marked with a red “X” (Figure 11).



Figure - Operon prediction inspector.

## Wizard Stage: Wrap

In the Wrap stage, the wizard terminates by saving all work in a session file (“.dex” filename extension). You then have the option of exiting Dexter or proceding to analysis of the imported data.

# Analysis

Dexter’s main screen, which organizes expression data for all imported experiments, is the starting point for exploration. In the course of exploration, individual genes can be collected into “experiments” for further analysis. The following two sections describe the functionality available in the main screen and in experiments.

## Main Screen

Dexter’s main analysis screen (Figure 12) initially contains one vertical scrolling column for each inported study, plus an extra column for experiments (ad-hoc collections of expression profiles).



Figure - Main Screen

The vertical columns contain thumbnail graphs that organize the gene expression graphs of each data set. Genes in thumbnail graphs may be grouped in 3 ways, under control of the “Group by” menu (circled in red):

* KEGG pathway. If pathways have been imported during the Wizard phase, each thumbnail represents a different pathway.
* Predicted operon. If predicted operons have been imported during the Wizard phase, each thumbnail represents a different operon.
* Order of appearance. Each thumbnail contains 20 genes, in order of appearance in the imported data set spreadsheet.

Within each vertical column, thumbnail graphs can be ordered alphabetically by name, by differential expression, or by population, under control of the “Order by” menu (circled in blue).

An individual thumbnail graph can be selected by clicking on it with the mouse, and expanded by SHIFT-clicking (Figure 13). Rolling the mouse over a gene name or i.d. in the legend highlights the gene’s signature in the graph. Clicking on a gene name or i.d. in the legend selects or deselects the gene; selected genes can be copied to an experiment.



Figure - Expanded Graph

The “Restrict” button in the main screen (Figure 12, circled in purple) displays a dialog (Figure 14) that allows specification of ranges for mean expression, range of expression, and change of expression. Ranges are edited by dragging the pins vertically. Genes whose expression falls outside any of the three ranges will not be displayed.



Figure - Restriction Dialog

The “Cluster” button in the main screen (Figure 12, circled in green) builds and displays a neighbor-joining tree by expression similarity of all expression profiles in all selected thumbnail graphs. The distance metric for the neighbor-joining algorithm can be Euclidean or Pearson Correlation Coefficient, selected by the “Metric” menu (Figure 12, circled in yellow). Figure 15 shows a typical tree.



Figure - Neighbor-joining tree of selected thumbnail graphs.

Rolling over a node in the tree displays a popup thumbnail graph of the subtree represented by the node. Node can be selected by clicking. When the tree dialog is dismissed, selected subtrees appear in a “Trees” column to the right of the “Experiments”. The “Save tree…” button allows the tree to be exported to a file in Newick format for display, exploration, or analysis by any compatible third-party application.

The “Export…” button (Figure 12, circled in brown) writes all data associated with all selected thumbnail graphs to a tsv (tab-separated) file. Optionally, data can be exported to the display for immediate visual inspection.

The “+” button at the top of the “Experiments” column (Figure 12, circled in orange) creates a new experiment containing all genes in all selected thumbnail graphs. Clicking “+” when no graphs are selected creates a new empty experiment. Shift-clicking on an experiment thumbnail graph brings up a dialog containing an expanded view of the experiment. The following section describes the capabilities of the expanded experiment view.

## Experiments

Figure 16 shows an expanded experiment dialog.



Figure - Expanded experiment dialog.

Initially genes are colored arbitrarily by a scheme that minimizes repetition of colors. Other color schemes, selectable via the “Color by” menu (circled in red) include coloring by data set, by organism (similar to coloring by data set, for situations where an organism might be represented in more than one data set), by KEGG pathway, or by order of addition to the experiment).

The experiment is assigned a default name that can be changed by clicking “Edit name…” (circled in green). The experiment name appears in the dialog banner and above the experiment thumbnail graph in the main display.

The “Zero mean” checkbox (circled in brown) normalizes every gene’s expression to a constant mean.

Genes can be added to an experiment by 5 criteria (buttons circled in blue):

To add by expression similarity, ensure that exactly one gene (the “query” gene) is selected in the legend and click “Expression Similarity…”. A dialog appears (Figure 17) displaying the 100 genes whose expression profile is most similar to the query. Genes are color-coded by data set. The pin at the top-right corner of the graph can be dragged to restrict the set of genes to the *n* genes most similar to the query, where 1 <= *n* <= 100. After restriction, clicking “Apply” causes all selected genes to be added to the experiment.



Figure - Add genes to an experiment by expression similarity.

To add genes by specifying a gene name or i.d., click “Gene Name…” to display a dialog (Figure 18). The main panel of the dialog is initially blank. Type a search string into the textfield (wildcard asterisks are supported) and type Enter or click “Search”. The main panel will become populated with one checkbox for every gene whose name or i.d. matches the search string. Select the desired genes. After the dialog is dismissed, selected genes will be added to the experiment.



Figure - Add genes to an experiment by name or i.d.

If KEGG pathways have been imported, you can add genes belonging to a specified pathway by clicking the “Pathway…” button and proceeding as with gene names or i.d.s.

If lists of orthologous genes have been imported, you can add genes that are orthologous to selected genes by clicking the “Orthology…” button. A dialog presents all genes orthologous to selections; desired genes in this dialog can be selected, and will be added to the experiment when the dialog is dismissed.

If operon predictions have been imported, you can add genes by operon membership. Ensure that exactly one gene (the “query” gene) is selected in the experiment and click the “Operon…” button. A dialog presents all genes that are predicted to be in the same operon as the query gene and are represented in the relevant data set. Desired genes in this dialog can be selected, and will be added to the experiment when the dialog is dismissed.

The “Proximity…” button (Figure 16, circled in orange) displays a proximity map for each organism represented in the experiment. Figure 19 shows proximity maps for two organisms, with 9 and 2 genes respectively in the originating experiment. Each proximity map displays its genes in order of appearance on the organism’s chromosome. Vertical lines display the number of intervening genes (a blue line indicates that genes are consecutive, a magenta line indicates that there is 1 intervening gene, etc.) If there are more than 3 intervening genes, the vertical line is black and text shows the number of intervening genes. (Note that consecutive genes may still be distant in terms of nucleotides.) If operon predictions have been imported, the extent of any predicted operons to which the genes belong is displayed as a dark green line at the right of the proximity map.



Figure - Proximity dialog.

# Afterword

After result sets and supplemental data are imported via the Wizard, Dexter’s main screen functions as an index, providing an overview of small thumbnail graphs. Exploration of thumbnail graphs is expected to draw attention to individual genes of interest, which can be added to experiments. The experiment display provides functionality for finding and adding additional genes.

Please send questions, comments, and bug reports to Phil Heller at pheller“at”soe.ucsc.edu.