Documentation of the RMAGEML package

Steffen Durinck‡, Joke Allemeersch‡†, Vincent J Carey¶, Yves Moreau‡, Bart De Moor‡

March 18, 2006

and ¶Channing Laboratory, Brigham and Women’s Hospital, 75 Francis Street, Boston 02115, USA

Contents

1 Introduction

MAGE-ML or Microarray Gene Expression Markup Language is a language designed to describe and exchange information about microarray experiments. MAGE-ML is based on XML and can describe microarray designs, microarray experiment setups, gene expression data, and data analysis results.
This package provides the link between MAGE-ML files and BioConductor. It gives the possibility to read in MAGE-ML files that describe cDNA microarray experiments. The functions convert the MAGE-ML files into the customary BioConductor objects (i.e., marrayLayout, marrayInfo and marrayRaw objects or limma RGLIst objects).

Here we give a short introduction to the Microarray and GeneExpression Object Model (MAGE-OM) and how we implemented the extraction of information necessary to make BioConductor objects. For a full description of MAGE-OM, we refer to the Gene Expression Specification: http://www.omg.org/cgi-bin/doc?formal/03-02-03.
In MAGE-ML these translate into packages with the same name. The packages needed for

*Steffen.Durinck@esat.kuleuven.ac.be
†Joke.Allemeersch@esat.kuleuven.ac.be
building BioConductor objects are BioAssayData, BioAssay, BioMaterial, BioSequence, ArrayDesign, and DesignElement. The DesignElement package contains a mapping of Features, which are the actual features present on the array, to Reporters, the reporter a feature represents. The DesignElement package also provides a mapping from Reporters to their corresponding BioSequence references. These BioSequence objects are characterized by their name and database entries in the BioSequence package. The ArrayDesign package contains information on the layout of the array. From this package, we can derive the position of each Feature on the array in terms of Zone (block or grid) and row and column within each Zone. The BioAssayData package describes the feature references that were assayed and the measured and derived QuantitationTypes. The BioAssay package describes the different steps in the microarray experiment. The last package used to make BioConductor objects is the BioMaterial package and describes how a sample is treated to obtain, for example, labeled samples used for hybridization.

2 Prerequisites

The RMAGEML package depends on SJava(>= 0.68) and a Java VM, e.g. j2resdk1.4.0. Other dependencies are as the Java-MAGEstk API and Java Xerces included in the package itself.

3 Getting started

Installing the package. The package can be installed as a normal R package: download the RMAGEML_2.0.4.tar.gz package and under Unix use the command

R CMD INSTALL RMAGEML_2.1.0.tar.gz.

The equivalent command for Windows is

Rcmd INSTALL RMAGEML_2.1.0.zip.

The package automatically loads the Biobase and marrayInput packages from BioConductor and the SJava libraries, so these should be installed as well.

Starting R. Before starting R one should be aware that the RMAGEML package uses SJava and that SJava requires to set the LD_LIBRARY_PATH environment variable before starting R. Without setting this variable the package won’t work.

Loading the package. You can load the package into R by typing
4 Import to marray packages

4.1 One step import and creation of an marrayRaw object from MAGE-ML files

In the marray packages of BioConductor the design of an array experiment is typically described by an `marrayLayout` and `marrayInfo` object. The function `importMAGEML` parses all MAGE-ML files present in the directory, which is given as a parameter to the function. From these files it creates an `marrayLayout` object, containing the Layout of one type of microarrays, and an `marrayInfo` object containing the gene names and database entries of the features spotted on the array. The name of the database to which the entries refer, is given in the 'notes' slot of the Gnames object. Next the function will extract the raw data values and output a complete `marrayRaw` object as a result.

The function can be tested on the MEXP-14 dataset. This example is available from ArrayExpress at [http://www.ebi.ac.uk/arrayexpress/](http://www.ebi.ac.uk/arrayexpress/). If one knows which `DesignElement Dimension`, `QuantitationType Dimension` and `Quantitation Types` are required, the import function can be used as:

```r
> datadir <- system.file("MAGEMLdata", package = "RMAGEML")
> raw <- importMAGEML(directory = datadir, package = "marray",
+ name.Rf = "QT:F635 Mean", name.Rb = "QT:B635 Median", name.Gf = "QT:F532 Mean",
+ name.Gb = "QT:B532 Median")
```

- Java Virtual Machine is running -
- parsing MAGEML files
- making Layout and Gnames objects
- Reading am2730miame.txt
- Reading am2731miame.txt
- Reading am2732m.txt
An object of class "marrayRaw"

@maRf

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>5841</td>
<td>2030</td>
<td>2968</td>
<td>45</td>
<td>1828</td>
<td>1975</td>
<td>2077</td>
<td>1775</td>
<td>2202</td>
<td>841</td>
</tr>
<tr>
<td>[2,]</td>
<td>2002</td>
<td>1312</td>
<td>421</td>
<td>96</td>
<td>399</td>
<td>557</td>
<td>295</td>
<td>748</td>
<td>465</td>
<td>83</td>
</tr>
<tr>
<td>[3,]</td>
<td>2254</td>
<td>2057</td>
<td>1097</td>
<td>649</td>
<td>917</td>
<td>755</td>
<td>1276</td>
<td>985</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>[4,]</td>
<td>2212</td>
<td>1492</td>
<td>782</td>
<td>767</td>
<td>709</td>
<td>1144</td>
<td>985</td>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5,]</td>
<td>73</td>
<td>76</td>
<td>42</td>
<td>45</td>
<td>49</td>
<td>49</td>
<td>46</td>
<td>47</td>
<td>43</td>
<td></td>
</tr>
</tbody>
</table>

955 more rows ...

@maGf

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>852</td>
<td>750</td>
<td>1587</td>
<td>135</td>
<td>1625</td>
<td>1183</td>
<td>1598</td>
<td>1108</td>
<td>807</td>
<td>1746</td>
</tr>
<tr>
<td>[2,]</td>
<td>652</td>
<td>529</td>
<td>404</td>
<td>162</td>
<td>397</td>
<td>515</td>
<td>285</td>
<td>390</td>
<td>291</td>
<td>190</td>
</tr>
<tr>
<td>[3,]</td>
<td>576</td>
<td>615</td>
<td>634</td>
<td>386</td>
<td>734</td>
<td>820</td>
<td>696</td>
<td>573</td>
<td>457</td>
<td>572</td>
</tr>
<tr>
<td>[4,]</td>
<td>781</td>
<td>589</td>
<td>733</td>
<td>366</td>
<td>758</td>
<td>848</td>
<td>667</td>
<td>597</td>
<td>559</td>
<td>716</td>
</tr>
<tr>
<td>[5,]</td>
<td>157</td>
<td>143</td>
<td>111</td>
<td>124</td>
<td>130</td>
<td>148</td>
<td>135</td>
<td>137</td>
<td>146</td>
<td>131</td>
</tr>
</tbody>
</table>

955 more rows ...

@maRb

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>42</td>
<td>42</td>
<td>36</td>
<td>39</td>
<td>41</td>
<td>47</td>
<td>40</td>
<td>45</td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td>[2,]</td>
<td>41</td>
<td>41</td>
<td>35</td>
<td>39</td>
<td>40</td>
<td>45</td>
<td>39</td>
<td>43</td>
<td>42</td>
<td>41</td>
</tr>
<tr>
<td>[3,]</td>
<td>41</td>
<td>42</td>
<td>34</td>
<td>41</td>
<td>40</td>
<td>44</td>
<td>40</td>
<td>43</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>[4,]</td>
<td>41</td>
<td>40</td>
<td>34</td>
<td>41</td>
<td>40</td>
<td>43</td>
<td>39</td>
<td>43</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>[5,]</td>
<td>41</td>
<td>39</td>
<td>34</td>
<td>40</td>
<td>40</td>
<td>42</td>
<td>39</td>
<td>43</td>
<td>41</td>
<td>41</td>
</tr>
</tbody>
</table>

955 more rows ...

@maGb

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>150</td>
<td>130</td>
<td>87</td>
<td>120</td>
<td>104</td>
<td>135</td>
<td>117</td>
<td>137</td>
<td>168</td>
<td>127</td>
</tr>
<tr>
<td>[2,]</td>
<td>147</td>
<td>128</td>
<td>88</td>
<td>111</td>
<td>105</td>
<td>136</td>
<td>116</td>
<td>131</td>
<td>137</td>
<td>121</td>
</tr>
<tr>
<td>[3,]</td>
<td>140</td>
<td>124</td>
<td>85</td>
<td>106</td>
<td>105</td>
<td>133</td>
<td>116</td>
<td>128</td>
<td>135</td>
<td>121</td>
</tr>
<tr>
<td>[4,]</td>
<td>138</td>
<td>122</td>
<td>88</td>
<td>108</td>
<td>106</td>
<td>133</td>
<td>116</td>
<td>128</td>
<td>134</td>
<td>120</td>
</tr>
<tr>
<td>[5,]</td>
<td>138</td>
<td>122</td>
<td>87</td>
<td>106</td>
<td>104</td>
<td>133</td>
<td>114</td>
<td>128</td>
<td>133</td>
<td>118</td>
</tr>
</tbody>
</table>
955 more rows ...

@maW
<0 x 0 matrix>

@maLayout
An object of class "marrayLayout"
@maNgr
[1] 4

@maNgc
[1] 4

@maNsr
[1] 10

@maNsc
[1] 6

@maNspots
[1] 960

@maSub
[1] TRUE

@maPlate
factor(0)
Levels:

@maControls
factor(0)
Levels:

@maNotes
[1] ""

@maGnames
An object of class "marrayInfo"
@maLabels
[1] "none" "none" "none" "none" "none"
955 more elements ...

@maInfo
An object of class "marrayInfo"

@maLabels
[1] "am2730miame.txt"  "am2731miame.txt"  "am2732m.txt"  "am2736m.txt"
[5] "am2737m.txt"  "tm1826m.txt"  "tm1827m.txt"  "tm1829m.txt"
[9] "tm1830m.txt"  "tm1831m.txt"

@maInfo

<table>
<thead>
<tr>
<th>Cy3</th>
<th>Cy5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AM-Pool AM2730-I</td>
</tr>
<tr>
<td>2</td>
<td>AM-Pool AM2731-I</td>
</tr>
<tr>
<td>3</td>
<td>AM-Pool AM2732-I</td>
</tr>
<tr>
<td>4</td>
<td>AM-Pool AM2736-I</td>
</tr>
<tr>
<td>5</td>
<td>AM-Pool AM2737-I</td>
</tr>
<tr>
<td>6</td>
<td>AM-Pool TM1826-I</td>
</tr>
<tr>
<td>7</td>
<td>AM-Pool TM1827-I</td>
</tr>
<tr>
<td>8</td>
<td>AM-Pool TM1829-I</td>
</tr>
<tr>
<td>9</td>
<td>AM-Pool TM1830-I</td>
</tr>
<tr>
<td>10</td>
<td>AM-Pool TM1831-I</td>
</tr>
</tbody>
</table>

@maNotes
[1] "Description of the targets"

@maNotes
character(0)

If however you do not know which DesignElement Dimension, QuantitationType Dimension and Quantitation Types to use, you can call the function as follows:

```r
> datadir <- system.file("MAGEMLdata", package = "RMAGEML")
> if (interactive()) {
  +   raw <- importMAGEML(directory = datadir, package = "marray")
+ }
```

This will generate a few selection panels which allow selection of the appropriate DesignElement Dimension, QuantitationType Dimension and Quantitation Types.
4.2 Creation of a Gnames marrayInfo object

If one just wants to make an marrayInfo object containing the gene names and database identifiers of the spotted features the function getGnames can be used.

```r
> data <- system.file("MAGEMLdata", package = "RMAGEML")
> mageom <- importMAGEOM(directory = data)
parsing MAGEML files

> getGnames(mageom, arrayID = "A-MEXP-14", DED = "DED:707", package = "marray")
An object of class "marrayInfo"
@maLabels
[1] "none" "none" "none" "none" "none"
955 more elements ...

@maInfo
[1] aj508733 VO0618 aj291984 aj306233 aj310439
142 Levels: af025843 af034412 af135499 aj132353 aj291832 aj291833 ... y17187
955 more rows ...

@maNotes
[1] "Identifiers refer to database: DB:embl"

Again leaving out the ‘DED’ parameter will cause selection panels to pop up displaying the available DesignElement Dimensions.

4.3 Creation of an marrayLayout object

In the marray packages the information on the array layout is stored in an marrayLayout object which can be created by the getArrayLayout function.

```r
> data <- system.file("MAGEMLdata", package = "RMAGEML")
> mageom <- importMAGEOM(directory = data)
parsing MAGEML files

> getArrayLayout(mageom, arrayID = "A-MEXP-14", DED = "DED:707")
An object of class "marrayLayout"
@maNgr
[1] 4

@maNgc
[1] 4
4.4 Make an marrayRaw object

The function makeMarrayRaw takes a Gnames and Layout object and parameters corresponding to the DesignElement Dimension, QuantitationType Dimension and Quantitation Types to create an marrayRaw object.

```r
> data <- system.file("MAGEMLdata", package = "RMAGEML")
> mageom <- importMAGEOM(directory = data)
```

Parsing MAGEML files

```r
> gnames <- getGnames(mageom, arrayID = "A-MEXP-14", DED = "DED:707",
+   package = "marray")
> layout <- getArrayLayout(mageom, arrayID = "A-MEXP-14", DED = "DED:707")
> raw <- makeMarrayRaw(mageOM = mageom, layout = layout, gnames = gnames,
+   directory = data, arrayID = "A-MEXP-14", DED = "DED:707",
+   QTD = "QTD:707", name.Rf = "QT:F635 Mean", name.Rb = "QT:B635 Median",
+   name.Gf = "QT:F532 Mean", name.Gb = "QT:B532 Median")
```

Reading am2730miame.txt
Reading am2731miame.txt
Reading am2732m.txt
5 Import to limma package

5.1 One step import and creation of a limma RGList object from MAGE-ML files

In the limma package of BioConductor the raw data is stored in an RGList object. The function `importMAGEML` parses all MAGE-ML files present in the directory which is given as a parameter to the function. From these files it creates the RGList object, containing the layout, gene names and database entries of the features spotted on the array and the foreground and background intensities for the green and red channels.

The function can be tested on the MEXP-14 dataset. This example is available from Array-Express at [http://www.ebi.ac.uk/arrayexpress/](http://www.ebi.ac.uk/arrayexpress/).

For import to limma the same function as MAGEML import to marray packages can be used, just adapt the name of the package into limma as follows:

```r
> datadir <- system.file("MAGEMLdata", package = "RMAGEML")
```

```r
> print(raw)
An object of class "RGList"
```

$R$
<table>
<thead>
<tr>
<th></th>
<th>5841</th>
<th>2030</th>
<th>2968</th>
<th>45</th>
<th>1828</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,</td>
<td>2002</td>
<td>1312</td>
<td>421</td>
<td>96</td>
<td>399</td>
</tr>
<tr>
<td>3,</td>
<td>2254</td>
<td>2057</td>
<td>1097</td>
<td>1163</td>
<td>649</td>
</tr>
<tr>
<td>4,</td>
<td>2212</td>
<td>1492</td>
<td>782</td>
<td>767</td>
<td>709</td>
</tr>
<tr>
<td>5,</td>
<td>73</td>
<td>76</td>
<td>42</td>
<td>54</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>1312</th>
<th>421</th>
<th>96</th>
<th>399</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,</td>
<td>557</td>
<td>295</td>
<td>748</td>
<td>465</td>
<td>83</td>
</tr>
<tr>
<td>3,</td>
<td>917</td>
<td>755</td>
<td>1276</td>
<td>985</td>
<td>335</td>
</tr>
<tr>
<td>4,</td>
<td>1114</td>
<td>620</td>
<td>1004</td>
<td>860</td>
<td>488</td>
</tr>
<tr>
<td>5,</td>
<td>49</td>
<td>49</td>
<td>46</td>
<td>47</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1975</th>
<th>2077</th>
<th>1775</th>
<th>2202</th>
<th>841</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,</td>
<td>557</td>
<td>295</td>
<td>748</td>
<td>465</td>
<td>83</td>
</tr>
<tr>
<td>3,</td>
<td>917</td>
<td>755</td>
<td>1276</td>
<td>985</td>
<td>335</td>
</tr>
<tr>
<td>4,</td>
<td>1114</td>
<td>620</td>
<td>1004</td>
<td>860</td>
<td>488</td>
</tr>
<tr>
<td>5,</td>
<td>49</td>
<td>49</td>
<td>46</td>
<td>47</td>
<td>43</td>
</tr>
</tbody>
</table>

955 more rows ...

$G$

<table>
<thead>
<tr>
<th></th>
<th>852</th>
<th>750</th>
<th>1587</th>
<th>135</th>
<th>1625</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,</td>
<td>652</td>
<td>529</td>
<td>404</td>
<td>162</td>
<td>397</td>
</tr>
<tr>
<td>3,</td>
<td>576</td>
<td>615</td>
<td>634</td>
<td>386</td>
<td>734</td>
</tr>
<tr>
<td>4,</td>
<td>781</td>
<td>589</td>
<td>733</td>
<td>366</td>
<td>758</td>
</tr>
<tr>
<td>5,</td>
<td>157</td>
<td>143</td>
<td>111</td>
<td>124</td>
<td>130</td>
</tr>
</tbody>
</table>

955 more rows ...

$Rb$

<table>
<thead>
<tr>
<th></th>
<th>42</th>
<th>42</th>
<th>36</th>
<th>39</th>
<th>41</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,</td>
<td>41</td>
<td>41</td>
<td>35</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>3,</td>
<td>41</td>
<td>42</td>
<td>34</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>4,</td>
<td>41</td>
<td>40</td>
<td>34</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>5,</td>
<td>41</td>
<td>39</td>
<td>34</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

955 more rows ...
Similarly if one only specifies the ‘directory’ and the ‘package’, selection panels will pop up to select the DesignElement Dimension, QuantitationType Dimension and Quantitation Types.

5.2 Creating the genes dataframe of an RGList object

In limma the gene names, gene identifiers and layout information is stored in a dataframe which can be created by the getArrayLayoutLimma function.

```r
> data <- system.file("MAGEMLdata", package = "RMAGEML")
> mageom <- importMAGEOM(directory = data)
parsing MAGEML files

> genes <- getArrayLayoutLimma(mageom, arrayID = "A-MEXP-14", DED = "DED:707")
> print(genes[1:10, ])

<table>
<thead>
<tr>
<th>Block</th>
<th>Row</th>
<th>Column</th>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>aj508733</td>
<td>none</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>VO0618</td>
<td>none</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>aj291984</td>
<td>none</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>aj306233</td>
<td>none</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>aj310439</td>
<td>none</td>
</tr>
</tbody>
</table>
```

955 more rows ...
5.3 Make an RGList object

The function makeRG takes a genes dataframe (containing the layout, gene identifiers and gene names), and parameters corresponding to DesignElement Dimension, QuantitationType Dimension and Quantitation Types to create a limma RGList object.

```r
> data <- system.file("MAGEMLdata", package = "RMAGEML")
> mageom <- importMAGEOM(directory = data)

> genes <- getArrayLayoutLimma(mageom, arrayID = "A-MEXP-14", DED = "DED:707")
> raw <- makeRG(mageom = mageom, genes = genes, directory = data,
+ name.Rf = "QT:F635 Mean", name.Rb = "QT:B635 Median", name.Gf = "QT:F532 Mean",
+ name.Gb = "QT:B532 Median")

Reading am2730miame.txt
Reading am2731miame.txt
Reading am2732m.txt
Reading am2736m.txt
Reading am2737m.txt
Reading tm1826m.txt
Reading tm1827m.txt
Reading tm1829m.txt
Reading tm1830m.txt
Reading tm1831m.txt

>raw <- importMAGEML(directory = "/home/steffen/data/MEXP-14", name.Gf = "QT:F635 Mean",
>norm<-maNorm(raw)
>mageom <- importMAGEOM(directory = "/home/steffen/data/E-MEXP-14")
>outputDirectory <- "/home/steffen/XMLout"
>magemlFile <- "RMAGEMLtest2.xml"
>rawDataFiles <- raw@maTargets@maLabels
>externalDataFiles <- c("deriv_test1.txt","deriv_test2.txt","deriv_test3.txt","deriv_test7.txt",
>test8.txt","deriv_test9.txt","deriv_test10.txt")
>protocolID <- "P-maNorm-test"
>protocol <- "This is a test protocol! Applied maNorm to the raw signal intensities"
```
```r
> qtID <- c("esat.kuleuven.ac.be:maNorm")
> qtName <- c("applied marrayNorm")
> qtDescription <- c("some description")
> qtScale <- c("linear")
> qtDataType <- c("scalar")
> qtDimID <- "esat.kuleuven.ac.be:QTD-test"
> date <- "testdate"
> tfmID <- "TFM-testID"
> addNormToMAGEML(mageOM = mageom, norm = norm, outputDirectory = outputDirectory, externalDataFiles = externalDataFiles, protocolID = protocolID, protocol = protocol, date=date, qtID = qtID, qtName = qtName, qtDescription = qtDescription, qtScale = qtScale, qtDataType = qtDataType, qtDimID = qtDimID, transformationID = tfmID, arrayID="A-MEXP-14", DED = "none", BADIDs = BADIDs, derivedBioAssayIDs = derivedBioAssayIDs, derivedBioAssayDataIDs = derivedBioAssayDataIDs, rawDataFiles=rawDataFiles)
> writeMAGEML(mageOM = mageom, directory = outputDirectory, file = magemlFile)
```