Statistical Methods for QTL Mapping
Aims and scope:
This series aims to capture new developments and summarize what is known over the entire spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical, and computational methods into biology by publishing a broad range of textbooks, reference works, and handbooks. The titles included in the series are meant to appeal to students, researchers, and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

Series Editors

N. F. Britton
Department of Mathematical Sciences
University of Bath

Xihong Lin
Department of Biostatistics
Harvard University

Hershel M. Safer
School of Computer Science
Tel Aviv University

Maria Victoria Schneider
European Bioinformatics Institute

Mona Singh
Department of Computer Science
Princeton University

Anna Tramontano
Department of Biochemical Sciences
University of Rome La Sapienza

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
3 Park Square, Milton Park
Abingdon, Oxfordshire OX14 4RN
UK
Published Titles

Algorithms in Bioinformatics: A Practical Introduction
Wing-Kin Sung

Bioinformatics: A Practical Approach
Shui Qing Ye

Biological Computation
Ehud Lamm and Ron Unger

Biological Sequence Analysis Using the SeqAn C++ Library
Andreas Gogol-Döring and Knut Reinert

Cancer Modelling and Simulation
Luigi Preziosi

Cancer Systems Biology
Edwin Wang

Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling
Arnaud Chauvière, Luigi Preziosi, and Claude Verdier

Clustering in Bioinformatics and Drug Discovery
John D. MacCuish and Norah E. MacCuish

Combinatorial Pattern Matching Algorithms in Computational Biology Using Perl and R
Gabriel Valiente

Computational Biology: A Statistical Mechanics Perspective
Ralf Blossey

Computational Hydrodynamics of Capsules and Biological Cells
C. Pozrikidis

Computational Neuroscience: A Comprehensive Approach
Jianfeng Feng

Computational Systems Biology of Cancer
Emmanuel Barillot, Laurence Calzone, Philippe Hupé, Jean-Philippe Vert, and Andrei Zinovyev

Data Analysis Tools for DNA Microarrays
Sorin Draghici

Differential Equations and Mathematical Biology, Second Edition
D.S. Jones, M.J. Plank, and B.D. Sleeman

Dynamics of Biological Systems
Michael Small

Engineering Genetic Circuits
Chris J. Myers

Exactly Solvable Models of Biological Invasion
Sergei V. Petrovskii and Bai-Lian Li

Game-Theoretical Models in Biology
Mark Broom and Jan Rychtář

Gene Expression Studies Using Affymetrix Microarrays
Hinrich Göhlmann and Willem Talloen

Genome Annotation
Jung Soh, Paul M.K. Gordon, and Christoph W. Sensen

Glycome Informatics: Methods and Applications
Kiyoko F. Aoki-Kinoshita

Handbook of Hidden Markov Models in Bioinformatics
Martin Gollery

Introduction to Bioinformatics
Anna Tramontano

Introduction to Bio-Ontologies
Peter N. Robinson and Sebastian Bauer

Introduction to Computational Proteomics
Golan Yona

Introduction to Proteins: Structure, Function, and Motion
Amit Kessel and Nir Ben-Tal

An Introduction to Systems Biology: Design Principles of Biological Circuits
Uri Alon

Kinetic Modelling in Systems Biology
Oleg Demin and Igor Goryanin

Knowledge Discovery in Proteomics
Igor Jurisica and Dennis Wigle
Published Titles (continued)

Meta-analysis and Combining Information in Genetics and Genomics
Rudy Guerra and Darlene R. Goldstein

Jules J. Berman

Modeling and Simulation of Capsules and Biological Cells
C. Pozrikidis

Niche Modeling: Predictions from Statistical Distributions
David Stockwell

Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems
Qiang Cui and Iveta Bahar

Optimal Control Applied to Biological Models
Suzanne Lenhart and John T. Workman

Pattern Discovery in Bioinformatics: Theory & Algorithms
Laxmi Parida

Python for Bioinformatics
Sebastian Bassi

Quantitative Biology: From Molecular to Cellular Systems
Sebastian Bassi

Spatial Ecology
Stephen Cantrell, Chris Cosner, and Shigui Ruan

Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulation
Horst Malchow, Sergei V. Petrovskii, and Ezio Venturino

Statistical Methods for QTL Mapping
Zehua Chen

Statistics and Data Analysis for Microarrays Using R and Bioconductor, Second Edition
Sorin Drăghici

Stochastic Modelling for Systems Biology, Second Edition
Darren J. Wilkinson

Structural Bioinformatics: An Algorithmic Approach
Forbes J. Burkowski

The Ten Most Wanted Solutions in Protein Bioinformatics
Anna Tramontano
Statistical
Methods for QTL Mapping

Zehua Chen
National University of Singapore
1 Biological Background

1.1 A Brief Sketch of the History of Genetics and Mendel’s Pea Plant Experiments .............................. 1
1.2 Basic Elements of Genetics ................................. 3
  1.2.1 Cells, Chromosomes and DNA Sequences .............. 3
  1.2.2 Genes .................................................. 5
  1.2.3 Genetic Traits and Phenotypes ........................ 5
  1.2.4 Gene Effects and Genetic Modes ....................... 6
1.3 Cell Division, Crossover and Recombination ............... 7
  1.3.1 Mitosis .................................................. 8
  1.3.2 Meiosis .................................................. 9
  1.3.3 Recombination ......................................... 11
1.4 Gene Expression ............................................. 12
  1.4.1 Transcription ........................................... 12
  1.4.2 Translation .............................................. 13
  1.4.3 Gene Regulation ....................................... 14
1.5 Genetic Maps and Mapping Functions ....................... 15
  1.5.1 Cytogenetic Maps ..................................... 15
  1.5.2 Physical Maps .......................................... 15
  1.5.3 Genetic Linkage Maps .................................. 15
  1.5.4 Crossover Interference ................................ 17
  1.5.5 A General Form of Mapping Functions ............... 18
  1.5.6 Special Mapping Functions ............................ 20
1.6 Experimental Crosses ...................................... 21
  1.6.1 The Mechanism of Selfing and Sibling Mating ........ 22
  1.6.2 The F1 Generation ..................................... 24
  1.6.3 Backcross .............................................. 25
  1.6.4 Intercross ............................................. 26
  1.6.5 Recombinant Inbred Lines ............................ 26
## Contents

1.6.6 Other Experimental Crosses .......................................................... 30

2 Selected Topics in Statistics ................................................................. 31

2.1 Population and Distribution .............................................................. 31
   2.1.1 Natural Population and Statistical Population .......................... 31
   2.1.2 Distributions ............................................................................ 32
   2.1.3 Special Distributions ............................................................... 34

2.2 Random Variable, Samples, Statistics and Related Distribution ........... 39
   2.2.1 Random Sample and Statistics ............................................... 39
   2.2.2 $t$, $\chi^2$ and $F$ Distributions .............................................. 41
   2.2.3 Distributions of Sample Mean and Sample Variance ............... 42

2.3 Estimation ......................................................................................... 42
   2.3.1 Maximum Likelihood Estimation .............................................. 42
   2.3.2 Method of Moments ................................................................. 44
   2.3.3 Unbiasedness, Mean Square Error and Consistency .............. 45
   2.3.4 Interval Estimation .................................................................. 46
     - 2.3.4.1 Asymptotic Confidence Intervals ................................. 46
     - 2.3.4.2 Bootstrap Confidence Intervals .................................... 47

2.4 Hypothesis Testing ........................................................................... 48
   2.4.1 Neyman-Pearson Framework ................................................... 49
   2.4.2 $t$-Test and $F$-Test ............................................................... 50
   2.4.3 Pearson’s $\chi^2$-Test ............................................................... 52
   2.4.4 Likelihood Ratio Test ............................................................... 54

2.5 Linear and Generalized Linear Models ............................................. 55
   2.5.1 Linear Models ........................................................................ 55
   2.5.2 Generalized Linear Models ..................................................... 57

2.6 Mixture Models and EM Algorithm .................................................. 59

2.7 Bayesian Analysis ............................................................................ 61
   2.7.1 General Framework of Bayesian Analysis ............................... 62
   2.7.2 Empirical and Hierarchical Bayesian Analysis ..................... 63
   2.7.3 Markov Chain Monte Carlo Simulation .................................. 64
   2.7.4 Bayes Factor .......................................................................... 65

2.8 Samplers for Markov Chain Monte Carlo Simulation ...................... 66
   2.8.1 The Gibbs Sampler ................................................................. 66
   2.8.2 Metropolis-Hastings Algorithm .............................................. 67
   2.8.3 Reversible Jump MCMC ........................................................... 68
   2.8.4 Extracting Information from the MCMC Sequence ............. 70

2.9 An Overview on Feature Selection with Small-n-Large-p Models ....... 71

2.10 Model Selection Criteria for Small-n-Large-p Models ...................... 73

3 Quantitative Genetics and General Issues on QTL Mapping ............... 77

3.1 Distributional Features of Genetic Quantitative Traits .................... 77
   3.1.1 Features of Genetic and Non-genetic Quantitative Traits ....... 78
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.2</td>
<td>Features of Genetic Traits Affected by Major and Minor QTL</td>
<td>78</td>
</tr>
<tr>
<td>3.2</td>
<td>Genetic Values and Variances</td>
<td>80</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Genotype and Allele Frequencies</td>
<td>81</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Genetic Values</td>
<td>82</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Genetic Variance and Heritability</td>
<td>85</td>
</tr>
<tr>
<td>3.3</td>
<td>Statistical Models with Known QTL Genotypes</td>
<td>86</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Models with Single QTL</td>
<td>86</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Models with Multiple QTL</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Conditional Probabilities of Putative QTL Genotypes Given Markers</td>
<td>88</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Conditional Probabilities of Putative QTL Genotypes in Backcross and Intercross Populations</td>
<td>88</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Conditional Probabilities of Putative QTL Genotypes in RILs</td>
<td>93</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Comparison of Conditional Probabilities under Different Assumptions on Crossover Interference</td>
<td>96</td>
</tr>
<tr>
<td>3.5</td>
<td>QTL Mapping Data</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>One-dimensional Mapping Approaches</td>
<td>101</td>
</tr>
<tr>
<td>4.1</td>
<td>QTL Mapping by Single-marker Analysis</td>
<td>101</td>
</tr>
<tr>
<td>4.2</td>
<td>Single Interval Mapping</td>
<td>104</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Interval Mapping by Normal Mixture Models</td>
<td>104</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Interval Mapping by Haley-Knott Regression</td>
<td>109</td>
</tr>
<tr>
<td>4.2.3</td>
<td>A Remark</td>
<td>111</td>
</tr>
<tr>
<td>4.3</td>
<td>Composite Interval Mapping</td>
<td>112</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Composite Interval Mapping with Normal Mixture Models</td>
<td>112</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Composite Interval Mapping with Haley-Knott Regression</td>
<td>114</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Choice of Adjusting Markers</td>
<td>114</td>
</tr>
<tr>
<td>4.4</td>
<td>Determination of Threshold Values</td>
<td>115</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Theoretical Approximation</td>
<td>116</td>
</tr>
<tr>
<td>4.4.2</td>
<td>A Simulation Approach</td>
<td>117</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Permutation Test</td>
<td>118</td>
</tr>
<tr>
<td>4.5</td>
<td>Determination of Sample Sizes</td>
<td>120</td>
</tr>
<tr>
<td>4.5.1</td>
<td>The Case with Known QTL Genotypes</td>
<td>120</td>
</tr>
<tr>
<td>4.5.2</td>
<td>The Case of Interval Mapping</td>
<td>123</td>
</tr>
<tr>
<td>4.6</td>
<td>Selective Genotyping</td>
<td>125</td>
</tr>
<tr>
<td>5</td>
<td>Multiple Interval Mapping</td>
<td>131</td>
</tr>
<tr>
<td>5.1</td>
<td>Gaussian Mixture Models for QTL Mapping</td>
<td>132</td>
</tr>
<tr>
<td>5.2</td>
<td>An EM algorithm for Gaussian Mixture Models</td>
<td>134</td>
</tr>
</tbody>
</table>
## Contents

5.3 Mixture Generalized Linear Models ........................................... 137
5.4 General EM Algorithms for Mixture Generalized Linear Models  139
  5.4.1 EM Algorithm for Backcross Design ................................. 140
  5.4.2 Adaptation for Intercross Design ..................................... 144
  5.4.3 A Special Issue on Binary Traits ...................................... 145
5.5 Multiple Interval Mapping Procedures ..................................... 146
5.6 Example: The Analysis of Radiata Pine Data ............................ 149
5.7 Multiple Interval Mapping with Polytomous Traits ...................... 149
  5.7.1 Mixture Models for Ordinal Traits ................................. 150
  5.7.2 EM Algorithm for Mixture Polytomous Models ............... 155
  5.7.3 Multiple Interval Mapping with Nominal Traits .............. 157
5.8 Multiple Interval Mapping of Traits with Spike Distributions .... 158
  5.8.1 The Mixture Models for Traits with Spike Distributions .... 159
  5.8.2 EM Algorithm for Traits with Spike Distributions .......... 160
  5.8.3 The Mapping Strategy .................................................. 163
  5.8.4 Example: The Analysis of Mouse *Listeria monocytogenes* Infection Data ................................................. 163
  5.8.5 Single-QTL Methods and Comparisons ................................. 164

6 QTL Mapping with Dense Markers .............................................. 167
  6.1 Feature Selection Methods for QTL Mapping ......................... 167
    6.1.1 Preliminary Screening ............................................. 168
    6.1.2 QTL Mapping by Penalized Likelihood Methods ............ 170
    6.1.3 Example: Analysis of CGEMS Prostate Cancer Data .... 171
    6.1.4 QTL Mapping by Sequential Methods ........................... 173
      6.1.4.1 Sequential Methods for Gaussian Traits ............. 174
      6.1.4.2 Sequential Methods for Non-Gaussian Traits .......... 176
      6.1.4.3 Example: Analysis of F<sub>2</sub> Rat Data ........ 178
  6.2 QTL Mapping with Consideration of QTL Epistatic Effects ....... 179
    6.2.1 Special Natures of Interaction Models for QTL Mapping ... 179
    6.2.2 The Mapping Procedure with Interaction Models ........ 181
    6.2.3 Example: Mapping QTL for Locomotor Activation and Anxiety with F<sub>2</sub> Mouse Model ................................. 183

7 Bayesian Approach to QTL Mapping .......................................... 185
  7.1 The Framework of the Bayesian Approach to QTL Mapping ........ 185
  7.2 Bayesian QTL Mapping with Inbred Line Cross Data ................. 187
    7.2.1 Bayesian Models without Epistatic Effects .................... 188
      7.2.1.1 The Number of QTLs Treated as Fixed ................... 189
      7.2.1.2 The Number of QTLs Treated as a Random Variable ........ 192
      7.2.1.3 Example: Mapping Loci Controlling Flowering Time in *Brassica napus* ........................................... 195
### Contents

7.2.2 Bayesian Models with Epistatic Effects .......................... 199
7.2.3 Example: Mapping QTL Affecting Heading in Two-Row Barley ......................................................... 205
7.2.4 Missing Marker Genotypes and Other Issues ................. 206
7.3 Bayesian QTL Mapping for Outbred Offsprings ................. 209
    7.3.1 The Prior on Marker Haplotypes and QTL Genotypes .... 209
    7.3.2 The MCMC Procedures for Mapping with Outbred Offsprings ....................................................... 213
7.4 Bayesian QTL Mapping for Ordinal Polytomous Traits .... 214

8 Multi-trait QTL Mapping and eQTL Mapping .................. 219
    8.1 Multi-trait QTL Mapping I: Single-QTL Model Approaches 220
        8.1.1 One-trait-at-a-time Method .............................. 220
        8.1.2 Meta-trait Method ....................................... 221
        8.1.3 Multivariate Composite Interval Mapping ............ 222
    8.2 Multi-trait QTL Mapping II: Multi-QTL Model Approaches 227
        8.2.1 Sparse Partial Least Squares Regression .............. 228
        8.2.2 Multivariate Sequential Procedures ................. 230
        8.2.3 Bayesian Approaches .................................. 237
    8.3 eQTL Mapping Approaches ................................ 244
        8.3.1 Specific Features of eQTL Mapping .................... 245
        8.3.2 Approaches Based on Single-QTL Models .......... 246
        8.3.3 Multi-QTL Model Methods ................................ 251
        8.3.4 Clustering of Transcripts ............................. 252
        8.3.5 General eQTL Mapping Strategies ................... 260
        8.3.6 Case Study: eQTL Mapping with GEO Mouse Data GSE3330 ....................................................... 261

Bibliography .................................................................... 267

Index ............................................................................. 285