

Contents

Part I Genetic Linkage Map

1	Map Functions	3
1.1	Physical Map and Genetic Map	3
1.2	Derivation of Map Functions	5
1.3	Haldane Map Function	8
1.4	Kosambi Map Function	8
2	Recombination Fraction	11
2.1	Mating Designs	11
2.2	Maximum Likelihood Estimation of Recombination Fraction	13
2.3	Standard Error and Significance Test	15
2.4	Fisher's Scoring Algorithm for Estimating r	17
2.5	EM Algorithm for Estimating r	21
3	Genetic Map Construction	23
3.1	Criteria of Optimality	24
3.2	Search Algorithms	25
3.2.1	Exhaustive Search	25
3.2.2	Heuristic Search	26
3.2.3	Simulated Annealing	28
3.2.4	Branch and Bound	29
3.3	Bootstrap Confidence of a Map	33
4	Multipoint Analysis of Mendelian Loci	35
4.1	Joint Distribution of Multiple-Locus Genotype	35
4.1.1	BC Design	36
4.1.2	F ₂ Design	37
4.1.3	Four-Way Cross Design	40
4.2	Incomplete Genotype Information	40
4.2.1	Partially Informative Genotype	40

4.2.2	BC and F_2 Are Special Cases of FW	42
4.2.3	Dominance and Missing Markers.....	43
4.3	Conditional Probability of a Missing Marker Genotype	44
4.4	Joint Estimation of Recombination Fractions.....	46
4.5	Multipoint Analysis for m Markers	47
4.6	Map Construction with Unknown Recombination Fractions	49

Part II Analysis of Quantitative Traits

5	Basic Concepts of Quantitative Genetics.....	53
5.1	Gene Frequency and Genotype Frequency.....	53
5.2	Genetic Effects and Genetic Variance	55
5.3	Average Effect of Allelic Substitution.....	56
5.4	Genetic Variance Components.....	57
5.5	Heritability	58
5.6	An F_2 Family Is in Hardy–Weinberg Equilibrium	59
6	Major Gene Detection.....	61
6.1	Estimation of Major Gene Effect	61
6.1.1	BC Design	61
6.1.2	F_2 Design	63
6.2	Hypothesis Tests.....	64
6.2.1	BC Design	64
6.2.2	F_2 Design	65
6.3	Scale of the Genotype Indicator Variable	67
6.4	Statistical Power	71
6.4.1	Type I Error and Statistical Power	72
6.4.2	Wald-Test Statistic	72
6.4.3	Size of a Major Gene	74
6.4.4	Relationship Between W -test and Z -test	75
6.4.5	Extension to Dominance Effect	76
7	Segregation Analysis	79
7.1	Gaussian Mixture Distribution.....	79
7.2	EM Algorithm	81
7.2.1	Closed Form Solution	81
7.2.2	EM Steps	82
7.2.3	Derivation of the EM Algorithm	83
7.2.4	Proof of the EM Algorithm	85
7.3	Hypothesis Tests.....	87
7.4	Variances of Estimated Parameters	88
7.5	Estimation of the Mixing Proportions	93
8	Genome Scanning for Quantitative Trait Loci	95
8.1	The Mouse Data	96
8.2	Genome Scanning	96

8.3	Missing Genotypes	98
8.4	Test Statistics	99
8.5	Bonferroni Correction	102
8.6	Permutation Test	103
8.7	Piepho's Approximate Critical Value	106
8.8	Theoretical Consideration	107
9	Interval Mapping	109
9.1	Least-Squares Method	110
9.2	Weighted Least Squares	113
9.3	Fisher Scoring	115
9.4	Maximum Likelihood Method	120
9.4.1	EM Algorithm	121
9.4.2	Variance–Covariance Matrix of $\hat{\theta}$	122
9.4.3	Hypothesis Test	126
9.5	Remarks on the Four Methods of Interval Mapping	126
10	Interval Mapping for Ordinal Traits	131
10.1	Generalized Linear Model	132
10.2	ML Under Homogeneous Variance	134
10.3	ML Under Heterogeneous Variance	136
10.4	ML Under Mixture Distribution	137
10.5	ML via the EM Algorithm	139
10.6	Logistic Analysis	146
10.7	Example	147
11	Mapping Segregation Distortion Loci	151
11.1	Probabilistic Model	152
11.1.1	The EM Algorithm	153
11.1.2	Hypothesis Test	155
11.1.3	Variance Matrix of the Estimated Parameters	156
11.1.4	Selection Coefficient and Dominance	159
11.2	Liability Model	159
11.2.1	EM Algorithm	161
11.2.2	Variance Matrix of Estimated Parameters	163
11.2.3	Hypothesis Test	164
11.3	Mapping QTL Under Segregation Distortion	164
11.3.1	Joint Likelihood Function	164
11.3.2	EM Algorithm	165
11.3.3	Variance–Covariance Matrix of Estimated Parameters ...	166
11.3.4	Hypothesis Tests	168
11.3.5	Example	169
12	QTL Mapping in Other Populations	171
12.1	Recombinant Inbred Lines	171
12.2	Double Haploids	175

12.3	Four-Way Crosses	175
12.4	Full-Sib Family	181
12.5	F_2 Population Derived from Outbreds	182
12.6	Example.....	183
13	Random Model Approach to QTL Mapping	187
13.1	Identity by Descent.....	189
13.2	Random Effect Genetic Model	191
13.3	Sib-Pair Regression	192
13.4	Maximum Likelihood Estimation	193
13.4.1	EM Algorithm	194
13.4.2	EM Algorithm Under Singular Value Decomposition	196
13.4.3	Multiple Siblings	197
13.5	Estimating the IBD Value for a Marker	199
13.6	Multipoint Method for Estimating the IBD Value	201
13.7	Genome Scanning and Hypothesis Tests	203
13.8	Multiple QTL Model.....	204
13.9	Complex Pedigree Analysis	207
14	Mapping QTL for Multiple Traits	209
14.1	Multivariate Model	210
14.2	EM Algorithm for Parameter Estimation	211
14.3	Hypothesis Tests.....	213
14.4	Variance Matrix of Estimated Parameters.....	214
14.5	Derivation of the EM Algorithm	215
14.6	Example.....	218
15	Bayesian Multiple QTL Mapping.....	223
15.1	Bayesian Regression Analysis	223
15.2	Markov Chain Monte Carlo.....	228
15.3	Mapping Multiple QTL	234
15.3.1	Multiple QTL Model	234
15.3.2	Prior, Likelihood, and Posterior.....	235
15.3.3	Summary of the MCMC Process	243
15.3.4	Post-MCMC Analysis	243
15.4	Alternative Methods of Bayesian Mapping	245
15.4.1	Reversible Jump MCMC	245
15.4.2	Stochastic Search Variable Selection	250
15.4.3	Lasso and Bayesian Lasso.....	252
15.5	Example: Arabidopsis Data	255
16	Empirical Bayesian QTL Mapping	257
16.1	Classical Mixed Model.....	257
16.1.1	Simultaneous Updating for Matrix G	259
16.1.2	Coordinate Descent Method	262
16.1.3	Block Coordinate Descent Method	264
16.1.4	Bayesian Estimates of QTL Effects	267

16.2	Hierarchical Mixed Model	268
16.2.1	Inverse Chi-Square Prior	268
16.2.2	Exponential Prior	272
16.2.3	Dealing with Sparse Models	275
16.3	Infinitesimal Model for Whole Genome Sequence Data	275
16.3.1	Data Trimming	276
16.3.2	Concept of Continuous Genome	276
16.4	Example: Simulated Data	278

Part III Microarray Data Analysis

17	Microarray Differential Expression Analysis	283
17.1	Data Preparation	283
17.1.1	Data Transformation	284
17.1.2	Data Normalization	285
17.2	F -test and T -test	288
17.3	Type I Error and False Discovery Rate	289
17.4	Selection of Differentially Expressed Genes	291
17.4.1	Permutation Test	291
17.4.2	Selecting Genes by Controlling FDR	293
17.4.3	Problems of the Previous Methods	295
17.4.4	Regularized T-test	296
17.5	General Linear Model	296
17.5.1	Fixed Model Approach	298
17.5.2	Random Model Approach	300
18	Hierarchical Clustering of Microarray Data	303
18.1	Distance Matrix	303
18.2	UPGMA	304
18.3	Neighbor Joining	306
18.3.1	Principle of Neighbor Joining	307
18.3.2	Computational Algorithm	312
18.4	Other Methods	317
18.5	Bootstrap Confidence	318
19	Model-Based Clustering of Microarray Data	321
19.1	Cluster Analysis with the K-means Method	321
19.2	Cluster Analysis Under Gaussian Mixture	323
19.2.1	Multivariate Gaussian Distribution	324
19.2.2	Mixture Distribution	325
19.2.3	The EM Algorithm	326
19.2.4	Supervised Cluster Analysis	327
19.2.5	Semisupervised Cluster Analysis	329
19.3	Inferring the Number of Clusters	329
19.4	Microarray Experiments with Replications	330

20	Gene-Specific Analysis of Variances	335
20.1	General Linear Model	335
20.2	The SEM Algorithm	338
20.3	Hypothesis Testing	341
21	Factor Analysis of Microarray Data	343
21.1	Background of Factor Analysis	343
21.1.1	Linear Model of Latent Factors	343
21.1.2	EM Algorithm	344
21.1.3	Number of Factors	348
21.2	Cluster Analysis	348
21.3	Differential Expression Analysis	350
21.4	MCMC Algorithm	351
22	Classification of Tissue Samples Using Microarrays	355
22.1	Logistic Regression	355
22.2	Penalized Logistic Regression	358
22.3	The Coordinate Descent Algorithm	359
22.4	Cross Validation	361
22.5	Prediction of Disease Outcome	361
22.6	Multiple-Category Classification	362
23	Time-Course Microarray Data Analysis	365
23.1	Gene Expression Profiles	365
23.2	Orthogonal Polynomial	366
23.3	B-spline	371
23.4	Mixed Effect Model	374
23.5	Mixture Mixed Model	375
23.6	EM Algorithm	376
23.7	Best Linear Unbiased Prediction	378
23.8	SEM Algorithm	379
23.8.1	Monte Carlo Sampling	380
23.8.2	SEM Steps	380
24	Quantitative Trait-Associated Microarray Data Analysis	383
24.1	Linear Association	384
24.1.1	Linear Model	384
24.1.2	Cluster Analysis	384
24.1.3	Three-Cluster Analysis	385
24.1.4	Differential Expression Analysis	387
24.2	Polynomial and B-spline	391
24.3	Multiple-Trait Association	394
25	Mapping Expression Quantitative Trait Loci	395
25.1	Individual Marker Analysis	396
25.1.1	SEM Algorithm	397
25.1.2	MCMC Algorithm	399

25.2	Joint Analysis of All Markers	402
25.2.1	Multiple eQTL Model	402
25.2.2	SEM Algorithm	403
25.2.3	MCMC Algorithm	405
25.2.4	Hierarchical Evolutionary Stochastic Search.....	408
References		413
Index		423



<http://www.springer.com/978-0-387-70806-5>

Principles of Statistical Genomics

Xu, S.

2013, XVI, 428 p., Hardcover

ISBN: 978-0-387-70806-5