
Contents

VOLUME 1

Colour plates appear between pages 220 and 221, and pages 364 and 365

List of Contributors	xxix
Editor's Preface to the Third Edition	xxxv
Glossary of Terms	xxxvii
Abbreviations and Acronyms	xlix

Part 1 GENOMES 1

1 Chromosome Maps 3

T.P. Speed and H. Zhao

1.1	Introduction	3
1.2	Genetic Maps	5
1.2.1	Mendel's Two Laws	5
1.2.2	Basic Principles in Genetic Mapping	7
1.2.3	Meiosis, Chromatid Interference, Chiasma Interference, and Crossover Interference	8
1.2.4	Genetic Map Functions	9
1.2.5	Genetic Mapping for Three Markers	9
1.2.6	Genetic Mapping for Multiple Markers	11
1.2.7	Tetrads	14
1.2.8	Half-tetrads	17
1.2.9	Other Types of Data	17
1.2.10	Current State of Genetic Maps	17
1.2.11	Programs for Genetic Mapping	18
1.3	Physical Maps	20
1.3.1	Polytene Chromosomes	21
1.3.2	Cytogenetic Maps	21
1.3.3	Restriction Maps	21
1.3.4	Restriction Mapping via Optical Mapping	22
1.3.5	Ordered Clone Maps	22

1.3.6	Contig Mapping Using Restriction Fragments	24
1.3.7	Sequence-tagged Site Maps	26
1.4	Radiation Hybrid Mapping	28
1.4.1	Haploid Data	29
1.4.2	Diploid Data	29
1.5	Other Physical Mapping Approaches	31
1.6	Gene Maps	32
	Acknowledgments	32
	References	32
2	Statistical Significance in Biological Sequence Comparison	40
	<i>W.R. Pearson and T.C. Wood</i>	
2.1	Introduction	40
2.2	Statistical Significance and Biological Significance	41
2.2.1	'Molecular' Homology	42
2.2.2	Examples of Similarity in Proteins	42
2.2.3	Inferences from Protein Homology	43
2.3	Estimating Statistical Significance for Local Similarity Searches	44
2.3.1	Measuring Sequence Similarity	44
2.3.2	Statistical Significance of Local Similarity Scores	47
2.3.3	Evaluating Statistical Estimates	58
2.4	Summary: Exploiting Statistical Estimates	62
	Acknowledgments	63
	References	63
3	Bayesian Methods in Biological Sequence Analysis	67
	<i>Jun S. Liu and T. Logvinenko</i>	
3.1	Introduction	67
3.2	Overview of the Bayesian Methodology	68
3.2.1	The Procedure	68
3.2.2	Model Building and Prior	69
3.2.3	Model Selection and Bayes Evidence	69
3.2.4	Bayesian Computation	70
3.3	Hidden Markov Model: A General Introduction	71
3.4	Pairwise Alignment of Biological Sequences	73
3.4.1	Bayesian Pairwise Alignment	73
3.5	Multiple Sequence Alignment	76
3.5.1	The Rationale of Using HMM for Sequence Alignment	77
3.5.2	Bayesian Estimation of HMM Parameters	79
3.5.3	PROBE and Beyond: Motif-based MSA Methods	82
3.5.4	Bayesian Progressive Alignment	83
3.6	Finding Recurring Patterns in Biological Sequences	85
3.6.1	Block-motif Model with iid Background	85
3.6.2	Block-motif Model with a Markovian Background	86
3.6.3	Block-motif Model with Inhomogeneous Background	86

3.6.4	Extension to Multiple Motifs	87
3.6.5	HMM for <i>cis</i> Regulatory Module Discovery	88
3.7	Joint Analysis of Sequence Motifs and Expression Microarrays	89
3.8	Summary	90
	Acknowledgments	91
Appendix A:	Markov Chain Monte Carlo Methods	91
	References	93
4	Statistical Approaches in Eukaryotic Gene Prediction	97
	<i>V. Solovyev</i>	
4.1	Structural Organization and Expression of Eukaryotic Genes	97
4.2	Methods of Functional Signal Recognition	100
4.2.1	Position-specific Measures	102
4.2.2	Content-specific Measures	104
4.2.3	Frame-specific Measures	104
4.2.4	Performance Measures	104
4.3	Linear Discriminant Analysis	105
4.4	Prediction of Donor and Acceptor Splice Junctions	106
4.4.1	Splice-sites Characteristics	106
4.4.2	Donor Splice-site Characteristics	110
4.4.3	Acceptor Splice-site Recognition	112
4.5	Identification of Promoter Regions in Human DNA	113
4.6	Recognition of PolyA Sites	121
4.7	Characteristics for Recognition of 3'-Processing Sites	123
4.8	Identification of Multiple Genes in Genomic Sequences	124
4.9	Discriminative and Probabilistic Approaches for Multiple Gene Prediction	124
4.9.1	HMM-based Multiple Gene Prediction	125
4.9.2	Pattern-based Multiple Gene-prediction Approach	127
4.10	Internal Exon Recognition	128
4.11	Recognition of Flanking Exons	129
4.11.1	5'-terminal Exon-coding Region Recognition	129
4.11.2	3'-exon-coding Region Recognition	129
4.12	Performance of Gene Identification Programs	131
4.13	Using Protein Similarity Information to Improve Gene Prediction	132
4.13.1	Components of Fgenesh++ Gene-prediction Pipeline	133
4.14	Genome Annotation Assessment Project (EGASP)	135
4.15	Annotation of Sequences from Genome Sequencing Projects	136
4.15.1	Finding Pseudogenes	137
4.15.2	Selecting Potential Pseudogenes	137
4.15.3	Selecting a Reliable Part of Alignment	140
4.16	Characteristics and Computational Identification of miRNA genes	141
4.17	Prediction of microRNA Targets	145
4.18	Internet Resources for Gene Finding and Functional Site Prediction	147
	Acknowledgments	150
	References	150

5	Comparative Genomics	160
	<i>J. Dicks and G. Savva</i>	
5.1	Introduction	160
5.2	Homology	163
5.3	Genomic Mutation	164
5.4	Comparative Maps	166
5.5	Gene Order and Content	170
5.5.1	Gene Order	170
5.5.2	Fragile Breakage versus Random Breakage Models	177
5.5.3	Gene Content	179
5.5.4	Comparison of Gene Order and Gene Content Methods	183
5.6	Whole Genome Sequences	184
5.6.1	Whole Genome Alignment	185
5.6.2	Finding Conserved Blocks	187
5.6.3	Dating Duplicated Genes and Blocks	190
5.7	Conclusions and Future Research	192
	Acknowledgments	193
	References	193
Part 2	BEYOND THE GENOME	201
6	Analysis of Microarray Gene Expression Data	203
	<i>W. Huber, A. von Heydebreck and M. Vingron</i>	
6.1	Introduction	203
6.2	Data Visualization and Quality Control	205
6.2.1	Image Quantification	205
6.2.2	Dynamic Range and Spatial Effects	206
6.2.3	Scatterplot	206
6.2.4	Batch Effects	209
6.3	Error Models, Calibration and Measures of Differential Expression	210
6.3.1	Multiplicative Calibration and Noise	211
6.3.2	Limitations	212
6.3.3	Multiplicative and Additive Calibration and Noise	214
6.4	Identification of Differentially Expressed Genes	216
6.4.1	Regularized t -Statistics	218
6.4.2	Multiple Testing	219
6.5	Pattern Discovery	221
6.5.1	Projection Methods	222
6.5.2	Cluster Algorithms	223
6.5.3	Local Pattern Discovery Methods	225
6.6	Conclusions	225
	Acknowledgments	226
	References	226
7	Statistical Inference for Microarray Studies	231
	<i>S.B. Pounds, C. Cheng and A. Onar</i>	
7.1	Introduction	231

7.2	Initial Data Processing	235
7.2.1	Normalization	236
7.2.2	Filtering	238
7.3	Testing the Association of Phenotype with Expression	240
7.3.1	Two-group and k -group Comparisons	240
7.3.2	Association with a Quantitative Phenotype	242
7.3.3	Association with a Time-to-event Endpoint	242
7.3.4	Computing p Values	244
7.4	Multiple Testing	245
7.4.1	Family-wise Error Rate	245
7.4.2	The False Discovery Rate	245
7.4.3	Significance Criteria for Multiple Hypothesis Tests	248
7.4.4	Significance Analysis of Microarrays	249
7.4.5	Selecting an MTA Method for a Specific Application	250
7.5	Annotation Analysis	253
7.6	Validation Analysis	254
7.7	Study Design and Sample Size	256
7.8	Discussion	259
	Related Chapters	260
	References	260
8	Bayesian Methods for Microarray Data	267
	<i>A. Lewin and S. Richardson</i>	
8.1	Introduction	267
8.2	Extracting Signal From Observed Intensities	269
8.2.1	Spotted cDNA Arrays	272
8.2.2	Oligonucleotide Arrays	273
8.3	Differential Expression	275
8.3.1	Normalization	275
8.3.2	Gene Variability	276
8.3.3	Expression Levels	277
8.3.4	Classifying Genes as Differentially Expressed	280
8.3.5	Multiclass Data	283
8.4	Clustering Gene Expression Profiles	283
8.4.1	Unordered Samples	284
8.4.2	Ordered Samples	287
8.5	Multivariate Gene Selection Models	288
8.5.1	Variable Selection Approach	289
8.5.2	Bayesian Shrinkage with Sparsity Priors	290
	Acknowledgments	291
	Related Chapters	291
	References	291
9	Inferring Causal Associations between Genes and Disease via the Mapping of Expression Quantitative Trait Loci	296
	<i>S.K. Sieberts and E.E. Schadt</i>	
9.1	Introduction	297

9.2	An Overview of Transcription as a Complex Process	299
9.3	Human Versus Experimental Models	301
9.4	Heritability of Expression Traits	302
9.5	Joint eQTL Mapping	303
9.6	Multilocus Models AND FDR	305
9.7	eQTL and Clinical Trait Linkage Mapping to Infer Causal Associations	307
9.7.1	A Simple Model for Inferring Causal Relationships	309
9.7.2	Distinguishing Proximal eQTL Effects from Distal	312
9.8	Using eQTL Data to Reconstruct Coexpression Networks	313
9.8.1	More Formally Assessing eQTL Overlaps in Reconstructing Coexpression Networks	315
9.8.2	Identifying Modules of Highly Interconnected Genes in Coexpression Networks	316
9.9	Using eQTL Data to Reconstruct Probabilistic Networks	317
9.10	Conclusions	321
9.11	Software	322
	References	323
10	Protein Structure Prediction	327
	<i>D.P. Klose and W.R. Taylor</i>	
10.1	History	327
10.2	Basic Structural Biology	328
10.2.1	The Hydrophobic Core	329
10.2.2	Secondary Structure	329
10.3	Protein Structure Prediction	329
10.3.1	Homology Modelling	330
10.3.2	Threading	330
10.3.3	True Threading	331
10.3.4	3D/1D Alignment	331
10.3.5	New Fold (NF) Prediction (<i>Ab initio</i> and <i>De novo</i> Approaches)	331
10.4	Model Evaluation	332
10.4.1	Decision Trees	332
10.4.2	Genetic Algorithms	334
10.4.3	k and Fuzzy k-nearest Neighbour	335
10.4.4	Bayesian Approaches	336
10.4.5	Artificial Neural Networks (ANNs)	338
10.4.6	Support Vector Machines	339
10.5	Conclusions	342
	References	344
11	Statistical Techniques in Metabolic Profiling	347
	<i>M. De Iorio, T.M.D. Ebbels and D.A. Stephens</i>	
11.1	Introduction	347
11.1.1	Spectroscopic Techniques	348
11.1.2	Data Pre-processing	350
11.1.3	Example Data	351

11.2	Principal Components Analysis and Regression	352
11.2.1	Principal Components Analysis	352
11.2.2	Principal Components Regression	354
11.3	Partial Least Squares and Related Methods	355
11.3.1	Partial Least Squares	355
11.3.2	PLS and Discrimination	357
11.3.3	Orthogonal Projections to Latent Structure	358
11.4	Clustering Procedures	360
11.4.1	Partitioning Methods	361
11.4.2	Hierarchical Clustering	361
11.4.3	Model-based Hierarchical Clustering	362
11.4.4	Choosing the Number of Clusters	363
11.4.5	Displaying and Interpreting Clustering Results	363
11.5	Neural Networks, Kernel Methods and Related Approaches	364
11.5.1	Mathematical Formulation	364
11.5.2	Kernel Density Estimates, PNNs and CLOUDS	366
11.6	Evolutionary Algorithms	367
11.7	Conclusions	369
	Acknowledgments	370
	References	370

Part 3 EVOLUTIONARY GENETICS 375

12 Adaptive Molecular Evolution 377

Z. Yang

12.1	Introduction	377
12.2	Markov Model of Codon Substitution	379
12.3	Estimation of Synonymous (d_S) and Nonsynonymous (d_N) Substitution Rates Between Two Sequences	381
12.3.1	Counting Methods	381
12.3.2	Maximum Likelihood Estimation	382
12.3.3	A Numerical Example and Comparison of Methods	386
12.4	Likelihood Calculation on a Phylogeny	388
12.5	Detecting Adaptive Evolution Along Lineages	390
12.5.1	Likelihood Calculation under Models of Variable ω Ratios among Lineages	390
12.5.2	Adaptive Evolution in the Primate Lysozyme	391
12.5.3	Comparison with Methods Based on Reconstructed Ancestral Sequences	392
12.6	Inferring Amino Acid Sites Under Positive Selection	394
12.6.1	Likelihood Ratio Test under Models of Variable ω Ratios among Sites	394
12.6.2	Methods That Test One Site at a Time	396
12.6.3	Positive Selection in the HIV-1 <i>vif</i> Genes	397
12.7	Testing Positive Selection Affecting Particular Sites and Lineages	399
12.7.1	Branch-site Test of Positive Selection	399

12.7.2	Similar Models	400
12.8	Limitations of Current Methods	401
12.9	Computer Software	402
	Acknowledgments	402
	References	402
13	Genome Evolution	407
	<i>J.F.Y. Brookfield</i>	
13.1	Introduction	407
13.2	The Structure and Function of Genomes	409
13.2.1	Genome Sequencing Projects	409
13.2.2	The Origins and Functions of Introns	411
13.3	The Organisation of Genomes	414
13.3.1	The Relative Positions of Genes: Are They Adaptive?	414
13.3.2	Functional Linkage Among Prokaryotes	415
13.3.3	Gene Clusters	415
13.3.4	Integration of Genetic Functions	416
13.3.5	Gene Duplications as Individual Genes or Whole Genome Duplications?	417
13.3.6	Apparent Genetic Redundancy	421
13.4	Population Genetics and the Genome	422
13.4.1	The Impact of Chromosomal Position on Population Genetic Variability	422
13.4.2	Codon Usage Bias	423
13.4.3	Effective Population Size	424
13.5	Mobile DNAs	425
13.5.1	Repetitive Sequences	425
13.5.2	Selfish Transposable Elements and Sex	426
13.5.3	Copy Number Control	426
13.5.4	Phylogenies of Transposable Elements	426
13.5.5	Functional Variation between Element Copies	430
13.6	Conclusions	430
	References	431
14	Probabilistic Models for the Study of Protein Evolution	439
	<i>J.L. Thorne and N. Goldman</i>	
14.1	Introduction	439
14.2	Empirically Derived Models of Amino Acid Replacement	440
14.2.1	The Dayhoff and Eck Model	440
14.2.2	Descendants of the Dayhoff Model	442
14.3	Amino Acid Composition	442
14.4	Heterogeneity of Replacement Rates Among Sites	443
14.5	Protein Structural Environments	444
14.6	Variation of Preferred Residues Among Sites	446
14.7	Models with a Physicochemical Basis	447
14.8	Codon-Based Models	448

14.9	Dependence Among Positions: Simulation	449
14.10	Dependence Among Positions: Inference	450
14.11	Conclusions	453
	Acknowledgments	454
	References	454
15	Application of the Likelihood Function in Phylogenetic Analysis	460
	<i>J.P. Huelsenbeck and J.P. Bollback</i>	
15.1	Introduction	460
15.2	History	462
	15.2.1 A Brief History of Maximum Likelihood in Phylogenetics	462
	15.2.2 A Brief History of Bayesian Inference in Phylogenetics	463
15.3	Likelihood Function	463
15.4	Developing an Intuition of Likelihood	469
15.5	Method of Maximum Likelihood	471
15.6	Bayesian Inference	474
15.7	Markov Chain Monte Carlo	476
15.8	Assessing Uncertainty of Phylogenies	480
15.9	Hypothesis Testing and Model Choice	481
15.10	Comparative Analysis	482
15.11	Conclusions	484
	References	484
16	Phylogenetics: Parsimony, Networks, and Distance Methods	489
	<i>D. Penny, M.D. Hendy and B.R. Holland</i>	
16.1	Introduction	489
16.2	DATA	490
	16.2.1 Character State Matrix	491
	16.2.2 Genetic Distances (Including Generalised Distances)	491
	16.2.3 Splits (Bipartitions)	496
	16.2.4 Sampling Error	498
16.3	Theoretical Background	499
	16.3.1 Terminology for Graphs and Trees	499
	16.3.2 Computational Complexity, Numbers of Trees	501
	16.3.3 Three Parts of an Evolutionary Model	504
	16.3.4 Stochastic Mechanisms of Evolution	507
16.4	Methods for Inferring Evolutionary Trees	509
	16.4.1 Five Desirable Properties for a Method	510
	16.4.2 Optimality Criteria	512
16.5	Phylogenetic Networks	520
	16.5.1 Reconstructing Reticulate Evolutionary Histories	520
	16.5.2 Displaying Conflicting Phylogenetic Signals	521
16.6	Search Strategies	524
	16.6.1 Complete or Exact Searches	524
	16.6.2 Heuristic Searches I—Limited (Local) Searches	526
	16.6.3 Heuristic Searches II—Hill-climbing and Related Methods	527

16.6.4	Quartets and Supertrees	528
16.7	Overview and Conclusions	529
	References	530
17	Evolutionary Quantitative Genetics	533
	<i>B. Walsh</i>	
17.1	Introduction	533
17.1.1	Resemblances, Variances, and Breeding Values	534
17.1.2	Single Trait Parent–Offspring Regressions	535
17.1.3	Multiple Trait Parent–Offspring Regressions	536
17.2	Selection Response Under the Infinitesimal Model	537
17.2.1	The Infinitesimal Model	537
17.2.2	Changes in Variances	538
17.2.3	The Roles of Drift and Mutation under the Infinitesimal Model	541
17.3	Fitness	542
17.3.1	Individual Fitness	543
17.3.2	Episodes of Selection	544
17.3.3	The Robertson–Price Identity	545
17.3.4	The Opportunity for Selection	545
17.3.5	Some Caveats in Using the Opportunity for Selection	547
17.4	Fitness Surfaces	548
17.4.1	Individual and Mean Fitness Surfaces	548
17.4.2	Measures of Selection on the Mean	549
17.4.3	Measures of Selection on the Variance	550
17.4.4	Gradients and the Geometry of Fitness Surfaces	551
17.4.5	Estimating the Individual Fitness Surface	552
17.4.6	Linear and Quadratic Approximations of $W(z)$	553
17.5	Measuring Multivariate Selection	555
17.5.1	Changes in the Mean Vector: The Directional Selection Differential	555
17.5.2	The Directional Selection Gradient	556
17.5.3	Directional Gradients, Fitness Surface Geometry, and Selection Response	557
17.5.4	Changes in the Covariance Matrix: The Quadratic Selection Differential	558
17.5.5	The Quadratic Selection Gradient	559
17.5.6	Quadratic Gradients, Fitness Surface Geometry, and Selection Response	561
17.5.7	Estimation, Hypothesis Testing, and Confidence Intervals	561
17.5.8	Geometric Interpretation of the Quadratic Fitness Regression	563
17.5.9	Unmeasured Characters and Other Biological Caveats	565
17.6	Multiple Trait Selection	566
17.6.1	Short-Term Changes in Means: The Multivariate Breeders' Equation	566
17.6.2	The Effects of Genetic Correlations: Direct and Correlated Responses	566

17.6.3	Evolutionary Constraints Imposed by Genetic Correlations	567
17.6.4	Inferring the Nature of Previous Selection	567
17.6.5	Changes in G under the Infinitesimal Model	568
17.6.6	Effects of Drift and Mutation	570
17.7	Phenotypic Evolution Models	570
17.7.1	Selection versus Drift in the Fossil Record	571
17.7.2	Stabilizing Selection	573
17.8	Theorems of Natural Selection: Fundamental and Otherwise	575
17.8.1	The Classical Interpretation of Fishers' Fundamental Theorem	576
17.8.2	What did Fisher Really Mean?	578
17.8.3	Heritabilities of Characters Correlated with Fitness	580
17.8.4	Robertson's Secondary Theorem of Natural Selection	580
17.9	Final Remarks	582
	Acknowledgments	582
	References	582

Part 4 ANIMAL AND PLANT BREEDING 587

18 Quantitative Trait Loci in Inbred Lines 589

R.C. Jansen

18.1	Introduction	589
18.1.1	Mendelian Factors and Quantitative Traits	589
18.1.2	The Genetics of Inbred Lines	590
18.1.3	Phenotype, Genotype and Environment	591
18.2	Segregation Analysis	592
18.2.1	Visualisation of Quantitative Variation in a Histogram	592
18.2.2	Plotting Mixture Distributions on Top of the Histogram	594
18.2.3	Fitting Mixture Distributions	595
18.2.4	Wanted: QTLs!	596
18.3	Dissecting Quantitative Variation With the Aid of Molecular Markers	597
18.3.1	Molecular Markers	597
18.3.2	Mixture Models	598
18.3.3	Alternative Regression Mapping	602
18.3.4	Highly Incomplete Marker Data	603
18.3.5	ANOVA and Regression Tests	603
18.3.6	Maximum Likelihood Tests	604
18.3.7	Analysis-of-deviance Tests	605
18.3.8	How Many Parameters Can We Fit Safely?	606
18.4	Qtl Detection Strategies	607
18.4.1	Model Selection and Genome Scan	607
18.4.2	Single-marker Analysis and Interval Mapping	608
18.4.3	Composite Interval Mapping	610
18.4.4	Multiple-QTL Mapping	611
18.4.5	Uncritical use of Model Selection Procedures	614
18.4.6	Final Comments	615
18.5	Bibliographic Notes	615

18.5.1	Statistical Approaches	615
18.5.2	Learning More about Important Genetic Parameters	616
18.5.3	QTL Analysis in Inbred Lines on a Large Scale	617
	Acknowledgments	618
	References	618
19	Mapping Quantitative Trait Loci in Outbred Pedigrees	623
	<i>I. Hoeschele</i>	
19.1	Introduction	623
19.2	Linkage Mapping via Least Squares or Maximum Likelihood and Fixed Effects Models	625
19.2.1	Least Squares	625
19.2.2	Maximum Likelihood	628
19.3	Linkage Mapping via Residual Maximum Likelihood and Random Effects Models	629
19.3.1	Identity-by-descent Probabilities of Alleles	629
19.3.2	Mixed Linear Model with Random QTL Allelic Effects	633
19.3.3	Mixed Linear Model with Random QTL Genotypic Effects	634
19.3.4	Relationship with Other Likelihood Methods	636
19.4	Linkage Mapping via Bayesian Methodology	638
19.4.1	General	638
19.4.2	Bayesian Mapping of a Monogenic Trait	639
19.4.3	Bayesian QTL Mapping	641
19.5	Deterministic Haplotyping In Complex Pedigrees	650
19.6	Genotype Sampling In Complex Pedigrees	653
19.7	Fine Mapping of Quantitative Trait Loci	665
19.7.1	Fine Mapping Using Current Recombinations	665
19.7.2	Fine Mapping Using Historical Recombinations	665
19.8	Concluding Remarks	668
	Acknowledgments	669
	References	669
20	Inferences from Mixed Models in Quantitative Genetics	678
	<i>D. Gianola</i>	
20.1	Introduction	678
20.2	Landmarks	680
20.2.1	Statistical Genetic Models	680
20.2.2	Best Linear Unbiased Prediction (BLUP)	681
20.2.3	Variance and Covariance Component Estimation	684
20.2.4	BLUP and Unknown Dispersion Parameters	687
20.2.5	Bayesian Procedures	688
20.2.6	Nonlinear, Generalized Linear Models, and Longitudinal Responses	690
20.2.7	Effects of Selection on Inferences	695
20.2.8	Massive Molecular Data: Semiparametric Methods	697
20.2.9	Computing Software	705

20.3	Future Developments	705
20.3.1	Model Development and Criticism	706
20.3.2	Model Dimensionality	706
20.3.3	Robustification of Inference	707
20.3.4	Inference Under Selection	708
20.3.5	Mixture Models	708
	Acknowledgments	709
	References	710
21	Marker-assisted Selection and Introgression	718
	<i>L. Moreau, F. Hospital and J. Whittaker</i>	
21.1	Introduction	718
21.2	Marker-assisted Selection: Inbred Line Crosses	719
21.2.1	Lande and Thompson's Formula	719
21.2.2	Efficiency of Marker-assisted Selection	722
21.2.3	Refinements	724
21.3	Marker-assisted selection: outbred populations	730
21.3.1	MAS via BLUP	731
21.3.2	Comments	732
21.3.3	Within-family MAS	734
21.4	Marker-assisted Introgression	736
21.4.1	Inbred Line Crosses	736
21.4.2	Outbred Populations	739
21.5	Marker-assisted Gene Pyramiding	740
21.6	Discussion	742
	Acknowledgments	745
	References	745

Reference Author Index	I
-------------------------------	----------

Subject Index	LXIII
----------------------	--------------

VOLUME 2

List of Contributors	xxix
Editor's Preface to the Third Edition	xxxv
Glossary of Terms	xxxvii
Abbreviations and Acronyms	xlix

Part 5 POPULATION GENETICS	753
-----------------------------------	------------

22 Mathematical Models in Population Genetics	755
<i>C. Neuhauser</i>	
22.1 A Brief History of The Role of Selection	755

22.2	Mutation, Random Genetic Drift, and Selection	756
22.2.1	Mutation	757
22.2.2	Random Genetic Drift	757
22.2.3	Selection	759
22.2.4	The Wright-Fisher Model	759
22.3	The Diffusion Approximation	760
22.3.1	Fixation	763
22.3.2	The Kolmogorov Forward Equation	764
22.3.3	Random Genetic Drift Versus Mutation and Selection	764
22.4	The Infinite Allele Model	765
22.4.1	The Infinite Allele Model with Mutation	766
22.4.2	Ewens's Sampling Formula	767
22.4.3	The Infinite Allele Model with Selection and Mutation	767
22.5	Other Models of Mutation and Selection	768
22.5.1	The Infinitely Many Sites Model	768
22.5.2	Frequency-dependent Selection	768
22.5.3	Overlapping Generations	769
22.6	Coalescent Theory	769
22.6.1	The Neutral Coalescent	769
22.6.2	The Ancestral Selection Graph	771
22.6.3	Varying Population Size	774
22.7	Detecting Selection	775
	Acknowledgments	777
	References	777
23	Inference, Simulation and Enumeration of Genealogies	781
	<i>C. Cannings and A. Thomas</i>	
23.1	Genealogies as Graphs	781
23.2	Relationships	782
23.2.1	The Algebra of Pairwise Relationships	782
23.2.2	Measures of Genetic Relationship	785
23.2.3	Identity States for Two Individuals	786
23.2.4	More Than Two Individuals	787
23.2.5	Example: Two Siblings Given Parental States	789
23.3	The Identity Process Along a Chromosome	790
23.3.1	Theory of Junctions	790
23.3.2	Random Walks	791
23.3.3	Other Methods	791
23.4	State Space Enumeration	792
23.4.1	Applying the <i>Peeling</i> Method	792
23.4.2	Recursions	793
23.4.3	More Complex Linear Systems	795
23.4.4	A Non-linear System	796
23.5	Marriage Node Graphs	796
23.5.1	Drawing Marriage Node Graphs	796
23.5.2	Zero-loop Pedigrees	798

23.6	Moral Graphs	801
23.6.1	Significance for Computation	801
23.6.2	Derivation from Marriage Node Graphs	802
23.6.3	Four Colourability and Triangulation	804
	References	805
24	Graphical Models in Genetics	808
	<i>S.L. Lauritzen and N.S. Sheehan</i>	
24.1	Introduction	808
24.2	Bayesian Networks and Other Graphical Models	809
24.2.1	Graph Terminology	809
24.2.2	Conditional Independence	809
24.2.3	Elements of Bayesian Networks	810
24.2.4	Object-oriented Specification of Bayesian Networks	810
24.3	Representation of Pedigree Information	811
24.3.1	Graphs for Pedigrees	811
24.3.2	Pedigrees and Bayesian Networks	812
24.4	Peeling and Related Algorithms	816
24.4.1	Compilation	817
24.4.2	Propagation	821
24.4.3	Random and Other Propagation Schemes	823
24.4.4	Computational Shortcuts	824
24.5	Pedigree Analysis and Beyond	824
24.5.1	Single-point Linkage Analysis	824
24.5.2	QTL Mapping	825
24.5.3	Pedigree Uncertainty	827
24.5.4	Forensic Applications	829
24.5.5	Bayesian Approaches	832
24.6	Causal Inference	832
24.6.1	Causal Concepts	833
24.6.2	Mendelian Randomisation	833
24.7	Other Applications	836
24.7.1	Graph Learning for Genome-wide Associations	836
24.7.2	Gene Networks	838
	References	838
25	Coalescent Theory	843
	<i>M. Nordborg</i>	
25.1	Introduction	843
25.2	The coalescent	844
25.2.1	The Fundamental Insights	844
25.2.2	The Coalescent Approximation	847
25.3	Generalizing the Coalescent	850
25.3.1	Robustness and Scaling	850
25.3.2	Variable Population Size	851
25.3.3	Population Structure on Different Time Scales	853

25.4	Geographical Structure	854
25.4.1	The Structured Coalescent	855
25.4.2	The Strong-migration Limit	856
25.5	Segregation	857
25.5.1	Hermaphrodites	857
25.5.2	Males and Females	859
25.6	Recombination	859
25.6.1	The Ancestral Recombination Graph	860
25.6.2	Properties and Effects of Recombination	863
25.7	Selection	865
25.7.1	Balancing Selection	865
25.7.2	Selective Sweeps	868
25.7.3	Background Selection	868
25.8	Neutral Mutations	869
25.9	Conclusion	870
25.9.1	The Coalescent and 'Classical' Population Genetics	870
25.9.2	The Coalescent and Phylogenetics	870
25.9.3	Prospects	872
	Acknowledgments	872
	References	872
26	Inference Under the Coalescent	878
	<i>M. Stephens</i>	
26.1	Introduction	878
26.1.1	Likelihood-based Inference	879
26.2	The Likelihood and the Coalescent	883
26.3	Importance Sampling	885
26.3.1	Likelihood Surfaces	887
26.3.2	Ancestral Inference	888
26.3.3	Application and Assessing Reliability	888
26.4	Markov Chain Monte Carlo	889
26.4.1	Introduction	889
26.4.2	Choosing a Good Proposal Distribution	891
26.4.3	Likelihood Surfaces	891
26.4.4	Ancestral Inference	893
26.4.5	Example Proposal Distributions	894
26.4.6	Application and Assessing Reliability	897
26.4.7	Extensions to More Complex Demographic and Genetic Models	899
26.5	Other Approaches	900
26.5.1	Rejection Sampling and Approximate Bayesian Computation	900
26.5.2	Composite Likelihood Methods	902
26.5.3	Product of Approximate Conditionals (PAC) Models	903
26.6	Software and Web Resources	903
26.6.1	Population Genetic Simulations	904
26.6.2	Inference Methods	904
	Acknowledgments	905
	References	906

27	Linkage Disequilibrium, Recombination and Selection	909
	<i>G. McVean</i>	
27.1	What Is Linkage Disequilibrium?	909
27.2	Measuring Linkage Disequilibrium	911
27.2.1	Single-number Summaries of LD	913
27.2.2	The Spatial Distribution of LD	914
27.2.3	Various Extensions of Two-locus LD Measures	918
27.2.4	The Relationship between r^2 and Power in Association Studies	919
27.3	Modelling LD and Genealogical History	922
27.3.1	A Historical Perspective	922
27.3.2	Coalescent Modelling	924
27.3.3	Relating Genealogical History to LD	930
27.4	Inference	932
27.4.1	Formulating the Hypotheses	932
27.4.2	Parameter Estimation	933
27.4.3	Hypothesis Testing	937
27.5	Prospects	938
	Acknowledgments	939
	Related Chapters	940
	References	940
28	Inferences from Spatial Population Genetics	945
	<i>F. Rousset</i>	
28.1	Introduction	945
28.2	Neutral Models of Geographical Variation	946
28.2.1	Assumptions and Parameters	946
28.3	Methods of Inference	948
28.3.1	F -statistics	948
28.3.2	Likelihood Computations	952
28.4	Inference Under the Different Models	955
28.4.1	Migration-Matrix Models	955
28.4.2	Island Model	955
28.4.3	Isolation by Distance	956
28.4.4	Likelihood Inferences	959
28.5	Separation of Timescales	960
28.6	Other Methods	962
28.6.1	Assignment and Clustering	962
28.6.2	Inferences from Clines	964
28.7	Integrating Statistical Techniques into the Analysis of Biological Processes	965
	Acknowledgments	966
	Related Chapters	966
	References	967
	Appendix A: Analysis of Variance and Probabilities of Identity	972
	Appendix B: Likelihood Analysis of the Island Model	977

29 Analysis of Population Subdivision	980
<i>L. Excoffier</i>	
29.1 Introduction	980
29.1.1 Effects of Population Subdivision	980
29.2 The Fixation Index F	982
29.3 Wright's F Statistics in Hierarchic Subdivisions	983
29.3.1 Multiple Alleles	985
29.3.2 Sample Estimation of F Statistics	986
29.3.3 G Statistics	987
29.4 Analysis of Genetic Subdivision Under an Analysis of Variance Framework	988
29.4.1 The Model	989
29.4.2 Estimation Procedure	991
29.4.3 Dealing with Mutation and Migration using Identity Coefficients	996
29.5 Relationship Between Different Definitions of Fixation Indexes	997
29.6 F Statistics and Coalescence Times	999
29.7 Analysis of Molecular Data: The Amova Framework	1001
29.7.1 Haplotypic Diversity	1001
29.7.2 Genotypic Data	1004
29.7.3 Multiallelic Molecular Data	1004
29.7.4 Dominant Data	1007
29.7.5 Relation of AMOVA with other Approaches	1008
29.8 Significance Testing	1009
29.8.1 Resampling Techniques	1009
29.9 Related and Remaining Problems	1011
29.9.1 Testing Departure from Hardy-Weinberg Equilibrium	1011
29.9.2 Detecting Loci under Selection	1012
29.9.3 What is the Underlying Genetic Structure of Populations?	1012
Acknowledgments	1013
References	1013
30 Conservation Genetics	1021
<i>M.A. Beaumont</i>	
30.1 Introduction	1021
30.2 Estimating Effective Population Size	1022
30.2.1 Estimating N_e Using Two Samples from the Same Population: The Temporal Method	1023
30.2.2 Estimating N_e from Two Derived Populations	1025
30.2.3 Estimating N_e Using One Sample	1030
30.2.4 Inferring Past Changes in Population Size: Population Bottlenecks	1033
30.2.5 Approximate Bayesian Computation	1040
30.3 Admixture	1041
30.4 Genotypic Modelling	1046
30.4.1 Assignment Testing	1046
30.4.2 Genetic Mixture Modelling and Clustering	1048
30.4.3 Hybridisation and the Use of Partially Linked Markers	1051

30.4.4	Inferring Current Migration Rates	1052
30.4.5	Spatial Modelling	1053
30.5	Relatedness and Pedigree Estimation	1054
	Acknowledgments	1057
	Related Chapters	1057
	References	1058
31	Human Genetic Diversity and its History	1067
	<i>G. Barbujani and L. Chikhi</i>	
31.1	Introduction	1068
31.2	Human Genetic Diversity: Historical Inferences	1069
31.2.1	Some Data on Fossil Evidence	1069
31.2.2	Models of Modern Human Origins	1070
31.2.3	Methods for Inferring Past Demography	1071
31.2.4	Reconstructing Past Human Migration and Demography	1075
31.3	Human Genetic Diversity: Geographical Structure	1081
31.3.1	Catalogues of Humankind	1081
31.3.2	Methods for Describing Population Structure	1084
31.3.3	Identifying the Main Human Groups	1087
31.3.4	Continuous versus Discontinuous Models of Human Variation	1091
31.4	Final Remarks	1092
	Acknowledgments	1096
	References	1096
Part 6	GENETIC EPIDEMIOLOGY	1109
32	Epidemiology and Genetic Epidemiology	1111
	<i>P.R. Burton, J.M. Bowden and M.D. Tobin</i>	
32.1	Introduction	1111
32.2	Descriptive Epidemiology	1112
32.2.1	Incidence and Prevalence	1114
32.2.2	Modelling Correlated Responses	1115
32.3	Descriptive Genetic Epidemiology	1117
32.3.1	Is There Evidence of Phenotypic Aggregation within Families?	1117
32.3.2	Is the Pattern of Correlation Consistent with a Possible Effect of Genes?	1118
32.3.3	Segregation Analysis	1126
32.3.4	Ascertainment	1127
32.4	Studies Investigating Specific Aetiological Determinants	1130
32.5	The Future	1131
	Acknowledgments	1132
	References	1132
33	Linkage Analysis	1141
	<i>E.A. Thompson</i>	
33.1	Introduction	1141

33.2	The Early Years	1142
33.3	The Development of Human Genetic Linkage Analysis	1144
33.4	The Pedigree Years; Segregation and Linkage Analysis	1146
33.5	Likelihood and Location Score Computation	1149
33.6	Monte Carlo Multipoint Linkage Likelihoods	1151
33.7	Linkage Analysis of Complex Traits	1155
33.8	Map Estimation, Map Uncertainty, and The Meiosis Model	1158
33.9	The Future	1162
	Acknowledgments	1163
	References	1163
34	Non-parametric Linkage	1168
	<i>P. Holmans</i>	
34.1	Introduction	1168
34.2	Pros and Cons of Model-free Methods	1169
34.3	Model-free Methods for Dichotomous Traits	1171
34.3.1	Affected Sib-pair Methods	1171
34.3.2	Parameter Estimation and Power Calculation Using Affected Sib Pairs	1172
34.3.3	Typing Unaffected Relatives in Sib-pair Analyses	1173
34.3.4	Application of Sib-pair Methods to Multiplex Sibships	1174
34.3.5	Methods for Analysing Larger Pedigrees	1175
34.3.6	Extensions to Multiple Marker Loci	1176
34.3.7	Multipoint Analysis with Tightly Linked Markers	1177
34.3.8	Inclusion of Covariates	1177
34.3.9	Multiple Disease Loci	1179
34.3.10	Significance Levels for Genome Scans	1180
34.3.11	Meta-analysis of Genome Scans	1180
34.4	Model-free Methods for Analysing Quantitative Traits	1181
34.5	Conclusions	1182
	Related Chapters	1182
	References	1183
35	Population Admixture and Stratification in Genetic Epidemiology	1190
	<i>P.M. McKeigue</i>	
35.1	Background	1191
35.2	Admixture Mapping	1192
35.2.1	Basic Principles	1192
35.2.2	Statistical Power and Sample Size	1194
35.2.3	Distinguishing between Genetic and Environmental Explanations for Ethnic Variation in Disease Risk	1196
35.3	Statistical Models	1198
35.3.1	Modelling Admixture	1198
35.3.2	Modelling Stratification	1199
35.3.3	Modelling Allele Frequencies	1201

35.3.4	Fitting the Statistical Model	1202
35.3.5	Model Comparison	1203
35.3.6	Assembling and Evaluating Panels of Ancestry-informative Marker Loci	1204
35.4	Testing For Linkage With Locus Ancestry	1205
35.4.1	Modelling Population Stratification	1207
35.5	Conclusions	1212
	References	1213
36	Population Association	1216
	<i>D. Clayton</i>	
36.1	Introduction	1216
36.2	Measures of Association	1217
36.3	Case-Control Studies	1219
36.4	Tests For Association	1221
36.5	Logistic Regression And Log-Linear Models	1225
36.6	Stratification And Matching	1227
36.7	Unmeasured Confounding	1230
36.8	Multiple Alleles	1232
36.9	Multiple Loci	1234
36.10	Discussion	1236
	Acknowledgments	1236
	References	1236
37	Whole Genome Association	1238
	<i>A.P. Morris and L.R. Cardon</i>	
37.1	Introduction	1238
37.1.1	Linkage Disequilibrium and Tagging	1239
37.1.2	Current WGA Studies	1240
37.2	Genotype Quality Control	1242
37.3	Single-Locus Analysis	1245
37.3.1	Logistic Regression Modelling Framework	1247
37.3.2	Interpretation of Results and Correction for Multiple Testing	1249
37.4	Population Structure	1250
37.5	Multi-Locus Analysis	1251
37.5.1	Haplotype-based Analyses	1252
37.5.2	Haplotype Clustering Techniques	1253
37.6	Epistasis	1254
37.7	Replication	1256
37.8	Prospects for Whole-Genome Association Studies	1257
	References	1258
38	Family-based Association	1264
	<i>F. Dudbridge</i>	
38.1	Introduction	1264
38.2	Transmission/Disequilibrium Test	1266

38.3	Logistic Regression Models	1268
38.4	Haplotype Analysis	1271
38.5	General Pedigree Structures	1273
38.6	Quantitative Traits	1276
38.7	Association in the Presence of Linkage	1278
38.8	Conclusions	1281
	References	1282
39	Cancer Genetics	1286
	<i>M.D. Teare</i>	
39.1	Introduction	1286
39.2	Armitage-Doll Models of Carcinogenesis	1287
39.2.1	The Multistage Model	1287
39.2.2	The Two-stage Model	1289
	Electronic Resources	1298
	References	1298
40	Epigenetics	1301
	<i>K.D. Siegmund and S. Lin</i>	
40.1	A Brief Introduction	1301
40.2	Technologies for CGI Methylation Interrogation	1303
40.2.1	MethyLight	1304
40.2.2	Methylation Microarrays	1304
40.3	Modeling Human Cell Populations	1305
40.3.1	Background	1305
40.3.2	Methylation Patterns	1305
40.3.3	Modeling Human Colon Crypts	1306
40.3.4	Summary	1307
40.4	Mixture Modeling	1307
40.4.1	Cluster Analysis	1308
40.4.2	Modeling Exposures for Latent Disease Subtypes	1310
40.4.3	Differential Methylation with Single-slide Data	1311
40.5	Recapitulation of Tumor Progression Pathways	1313
40.5.1	Background	1313
40.5.2	Heritable Clustering	1313
40.5.3	Further Comments	1315
40.6	Future Challenges	1316
	Acknowledgments	1316
	References	1317
Part 7	SOCIAL AND ETHICAL ASPECTS	1323
41	Ethics Issues in Statistical Genetics	1325
	<i>R.E. Ashcroft</i>	
41.1	Introduction: Scope of This Chapter	1325
41.1.1	What is Ethics?	1326

41.1.2	Models for Analysing the Ethics of Population Genetic Research	1327
41.2	A Case Study in Ethical Regulation of Population Genetics Research: UK Biobank's Ethics and Governance Framework	1329
41.2.1	The Scientific and Clinical Value of the Research	1330
41.2.2	Recruitment of Participants	1332
41.2.3	Consent	1334
41.2.4	Confidentiality and Security	1339
41.3	Stewardship	1339
41.3.1	Benefit Sharing	1340
41.3.2	Community Involvement	1341
41.4	Wider Social Issues	1341
41.4.1	Geneticisation	1341
41.4.2	Race, Ethnicity and Genetics	1342
41.5	Conclusions	1343
	Acknowledgments	1343
	References	1343
42	Insurance	1346
	<i>A.S. Macdonald</i>	
42.1	Principles of Insurance	1346
42.1.1	Long-term Insurance Pricing	1346
42.1.2	Life Insurance Underwriting	1348
42.1.3	Familial and Genetic Risk Factors	1349
42.1.4	Adverse Selection	1349
42.1.5	Family Medical Histories	1350
42.1.6	Legislation and Regulation	1351
42.1.7	Quantitative Questions	1352
42.2	Actuarial Modelling	1352
42.2.1	Actuarial Models for Life and Health Insurance	1352
42.2.2	Parameterising Actuarial Models	1354
42.2.3	Market Models and Missing Information	1355
42.2.4	Modelling Strategies	1357
42.2.5	Statistical Issues	1358
42.2.6	Economics Issues	1359
42.3	Examples and Conclusions	1359
42.3.1	Single-gene Disorders	1359
42.3.2	Multifactorial Disorders	1361
	References	1365
43	Forensics	1368
	<i>B.S. Weir</i>	
43.1	Introduction	1368
43.2	Principles of Interpretation	1369
43.3	Profile Probabilities	1371
43.3.1	Allelic Independence	1371
43.3.2	Allele Frequencies	1373

43.3.3	Joint Profile Probabilities	1375
43.4	Parentage Issues	1377
43.5	Identification of Remains	1379
43.6	Mixtures	1379
43.7	Sampling Issues	1383
43.7.1	Allele Probabilities	1383
43.7.2	Coancestry	1384
43.8	Other Forensic Issues	1385
43.8.1	Common Fallacies	1385
43.8.2	Relevant Population	1385
43.8.3	Database Searches	1386
43.8.4	Uniqueness of Profiles	1386
43.8.5	Assigning Individuals to Phenotypes, Populations or Families	1388
43.8.6	Hierarchy of Propositions	1389
43.9	Conclusions	1390
	References	1390

Reference Author Index

I

Subject Index

LXIII