

Table des matières

Avant-propos	1
Gilles DIDIER et Stéphane GUINDON	
Chapitre 1. Arbres : combinatoire et modèles	7
Gilles DIDIER et Stéphane GUINDON	
1.1. Introduction	7
1.2. Définitions préliminaires	8
1.3. Compter les arbres	10
1.3.1. Arbres non enracinés entièrement étiquetés	10
1.3.2. Arbres binaires à feuilles étiquetées	12
1.3.3. Arbres binaires à feuilles étiquetées et à nœuds internes ordonnés	13
1.3.4. Nombre d'ordres de nœuds internes d'un arbre donné	14
1.3.5. Arbres binaires orientés	15
1.4. Probabilités d'arbres résultant de processus de branchement	15
1.5. Processus de naissance-mort	18
1.5.1. Densité de probabilité d'un arbre de naissance-mort	21
1.6. Le coalescent	24
1.6.1. Liens avec les modèles « classiques » en génétique des populations	25
1.6.2. Modèle de Moran	26
1.6.3. Modèle de Wright-Fisher	27
1.6.4. Modèle générique	28
1.6.5. Densité de probabilité d'un arbre généré par le coalescent	29
1.7. Conclusion	30
1.8. Bibliographie	32

Chapitre 2. Modèles d'évolution pour les séquences et les caractères discrets	33
Étienne PARDOUX	
2.1. Introduction	33
2.2. Processus de Markov en temps continu à valeurs dans un ensemble discret	34
2.2.1. Processus de Poisson	34
2.2.2. Processus de Markov en temps continu à valeurs dans un ensemble fini	35
2.3. Modèles d'évolution des séquences d'ADN	38
2.3.1. Modèle de Jukes-Cantor	39
2.3.2. Modèle de Kimura	39
2.3.3. Modèle de Felsenstein	40
2.3.4. Modèle HKY	40
2.3.5. Modèle réversible général (REV)	41
2.4. Modèles d'évolution des taux le long de la séquence	41
2.4.1. Taux i.i.d. le long de la séquence	42
2.4.2. Modèle de Markov caché	42
2.5. Modèles d'évolution de caractères discrets	43
2.6. Bibliographie	44
Chapitre 3. Modèles d'évolution de caractères continus	47
Paul BASTIDE, Mahendra MARIADASSOU et Stéphane ROBIN	
3.1. Motivations	47
3.1.1. Méthodes comparatives	48
3.1.2. Études des phénomènes évolutifs	49
3.2. Mouvement brownien	50
3.2.1. Description	51
3.2.2. Régression phylogénétique et transformations statistiques	52
3.2.3. Algorithmes récursifs pour l'inférence	55
3.3. Analyse multivariée	56
3.3.1. Description	56
3.3.2. Contrastes phylogénétiques	57
3.3.3. ACP phylogénétique	58
3.4. Modèles gaussiens	59
3.4.1. Quelques limites du mouvement brownien	59
3.4.2. Processus d'Ornstein-Uhlenbeck	60
3.4.3. Interprétations biologiques et mises en garde	64
3.4.4. Autres processus gaussiens	66
3.4.5. Évolution hétérogène	69
3.4.6. Modèles d'observation	72
3.4.7. Sélection de modèle	74

3.5. Extensions et généralisations	75
3.5.1. Modèles non gaussiens	75
3.5.2. Interactions entre l'arbre et le trait	76
3.5.3. Interactions entre espèces	76
3.5.4. Trait de grande dimension	77
3.6. Références utiles	78
3.7. Remerciements	79
3.8. Bibliographie	79

Chapitre 4. Évolution corrélée : modèles et méthodes 87

Guillaume ACHAZ et Julien Y. DUTHEIL

4.1. Introduction	87
4.2. Évolution corrélée entre traits	90
4.2.1. Non-indépendance des espèces	90
4.2.2. Contrastes phylogénétiquement indépendants	93
4.2.3. Extension du modèle linéaire pour prendre en compte la phylogénie	95
4.2.4. Corrélation entre traits discrets	99
4.2.5. Exemples de traits corrélés	101
4.2.6. Modélisation conjointe des traits et des séquences	102
4.3. Évolution corrélée au sein des génomes	102
4.3.1. Au sein des gènes, entre les nucléotides	103
4.3.2. Au sein des protéines, entre les acides aminés	105
4.3.3. Au sein des génomes, entre les gènes	110
4.4. La génétique vue comme évolution corrélée	112
4.4.1. Au sein des individus	112
4.4.2. Dans les pédigrées	114
4.4.3. Dans la population	116
4.5. Conclusion	119
4.6. Bibliographie	121

Chapitre 5. Un siècle de réarrangements génomiques 127

Anne BERGERON et Krister M. SWENSON

5.1. Introduction	127
5.2. Ordre linéaire des gènes et opérations de réarrangement	128
5.2.1. Représentations et définitions de base	129
5.2.2. Opérations DCJ et graphe des cassures	131
5.3. Dénombrement de scénarios	135
5.3.1. Scénarios possibles pour un cycle équilibré de longueur $2m$	135
5.3.2. (Nombreuses) décompositions en cycles du graphe des cassures	136

5.4. Données de contact et scénarios pondérés	139
5.4.1. Modélisation des données de contact	139
5.4.2. Arbres planaires et algorithme d'exploration	141
5.5. Conclusion	146
5.6. Bibliographie	148

Chapitre 6. Inférence phylogénétique : méthodes basées sur les distances

151

Fabio PARDI

6.1. Introduction	151
6.2. Fondements mathématiques	153
6.3. Estimation des distances	157
6.3.1. Distances estimées à partir de séquences alignées	157
6.3.2. Approches alternatives pour estimer les distances	159
6.4. L'inférence d'arbre	161
6.4.1. Ajustement des longueurs de branches par les moindres carrés	162
6.4.2. Des moindres carrés au minimum d'évolution	164
6.4.3. NJ et autres algorithmes agglomératifs	166
6.4.4. Au-delà des distances	169
6.5. Conclusion	170
6.6. Bibliographie	171

Chapitre 7. Calculs d'inférence dans les arbres phylogénétiques

177

Laurent GUÉGUEN

7.1. Inférences et modélisation	177
7.1.1. Inférences	177
7.1.2. Parcimonie et vraisemblance	178
7.2. Programmation dynamique	181
7.2.1. Sur les branches	182
7.2.2. Sur les nœuds	183
7.2.3. Sur l'arbre	183
7.2.4. À la racine	184
7.2.5. Formules de récursivité	184
7.2.6. Réduction de la complexité	186
7.2.7. Gestion de la racine	187
7.3. Maximum de parcimonie	188
7.3.1. Inférence ancestrale	191
7.4. Vraisemblance	191
7.4.1. Gestion de la racine	194
7.4.2. Calcul aux nœuds	194

7.4.3. Maximisation, dérivation	196
7.4.4. Inférence ancestrale	199
7.5. Bibliographie	202

Chapitre 8. Le paradigme bayésien en phylogénie moléculaire 203

Nicolas RODRIGUE

8.1. Introduction	203
8.2. Principes généraux de l'approche bayésienne en phylogénie	204
8.2.1. Échantillonnage Monte Carlo par chaîne de Markov	207
8.2.2. Résumé de la distribution postérieure et de l'échantillonnage	210
8.3. Démarginalisation de la fonction de vraisemblance	210
8.3.1. Expansion de paramètres	211
8.3.2. Augmentation de données	212
8.4. Sélection bayésienne de modèles de substitution	214
8.4.1. Comparaison relative de modèles <i>via</i> le facteur de Bayes	214
8.4.2. Évaluation absolue de modèles <i>via</i> la simulation postérieure prédictive	217
8.5. Impacts et directions à venir	218
8.6. Bibliographie	219

Chapitre 9. Phylogénétique : quelles mesures de support pour les branches d'un arbre ? 223

Olivier GASCUEL et Frédéric LEMOINE

9.1. Introduction	223
9.2. Supports locaux : aLRT paramétrique et non paramétrique	225
9.2.1. Test de branche nulle et ses limites	225
9.2.2. Test local aLRT, version paramétrique	227
9.2.3. Test local aLRT, version SH-like non paramétrique	228
9.2.4. Comparaison sur un exemple des supports aLRT et bootstrap	229
9.3. Bootstrap phylogénétique	231
9.3.1. Bootstrap statistique	231
9.3.2. Bootstrap de Felsenstein	232
9.3.3. Bootstrap de transfert	233
9.3.4. Comparaison sur un exemple des supports bootstrap	237
9.4. Supports bayésiens	239
9.4.1. Principe, utilisation de chaînes de Markov Monte Carlo	239
9.4.2. Support bayésien local	241
9.4.3. Comparaison des supports bayésiens sur un exemple	242
9.5. Discussion	242
9.6. Bibliographie	245

Chapitre 10. Fossiles et phylogénie 247

Michel LAURIN

- 10.1. Inférences sur la topologie 247
 - 10.1.1. Premières approches 247
 - 10.1.2. Caractères utilisables en paléontologie 249
 - 10.1.3. Première approche quantitative : la phénétique 251
 - 10.1.4. La stratophénétique 251
 - 10.1.5. La cladistique 252
 - 10.1.6. Approches par modèles : vraisemblance, approches bayésiennes 253
 - 10.1.7. Fossiles et données moléculaires 255
- 10.2. Datation de l'arbre du vivant 255
 - 10.2.1. Premières approches qualitatives 256
 - 10.2.2. Premières approches statistiques 256
 - 10.2.3. Datation moléculaire 257
 - 10.2.4. Datation par les bouts de branches 259
 - 10.2.5. Datation par modèles de naissance et de mort 260
- 10.3. Conclusion 262
- 10.4. Bibliographie 263

Chapitre 11. Phylodynamique 269

Samuel ALIZON

- 11.1. Réconcilier écologie, évolution et mathématiques 269
- 11.2. Données et processeurs 270
 - 11.2.1. Séquencage de nouvelle génération 271
 - 11.2.2. PCR et capture 271
- 11.3. Phylogénies d'infections 272
 - 11.3.1. Lien avec les chaînes de transmission 272
 - 11.3.2. Datation et vitesse d'évolution 273
 - 11.3.3. Applications biologiques de la calibration temporelle 274
- 11.4. Phylodynamique 275
 - 11.4.1. Un champ en quête de définition 275
 - 11.4.2. Au plus près de l'épidémiologie 276
 - 11.4.3. Le coalescent 277
 - 11.4.4. Modèles de naissance et de mort 279
 - 11.4.5. Limites des approches avec vraisemblance 280
 - 11.4.6. Phylodynamique ABC 280
- 11.5. Phylogéographie d'infections 281
- 11.6. Traits d'histoire de vie des infections et des virus 282

11.7. Perspectives et défis	284
11.8. Bibliographie	285

**Chapitre 12. Inférence de processus démographiques
chez les populations humaines 293**

Frédéric AUSTERLITZ

12.1. Introduction	293
12.2. Inférences démographiques sur des données de génétique des populations	296
12.2.1. Reconstitution de l'histoire des Pygmées d'Afrique centrale . . .	296
12.2.2. Inférence de l'histoire des populations en Asie centrale	297
12.2.3. Impact du mode de vie sur la dynamique de croissance des populations	299
12.3. Inférence des expansions humaines à partir des données de séquences nouvelle génération	300
12.4. Reconstruction de la dynamique démographique à partir des données génétiques et culturelles	303
12.4.1. Analyse simultanée des diversités génétiques et linguistiques . .	303
12.4.2. Détection de la transmission intergénérationnelle du succès reproducteur	305
12.5. Conclusion	307
12.6. Bibliographie	307

Liste des auteurs 313

Index 315