UMR - BGPI
Biologie et Génétique
des Interactions Plante-Parasite

Méthodes ABC

Approximate Bayesian Computation

V. Ravigné, B. Barrès, D. Tharreau

INRA

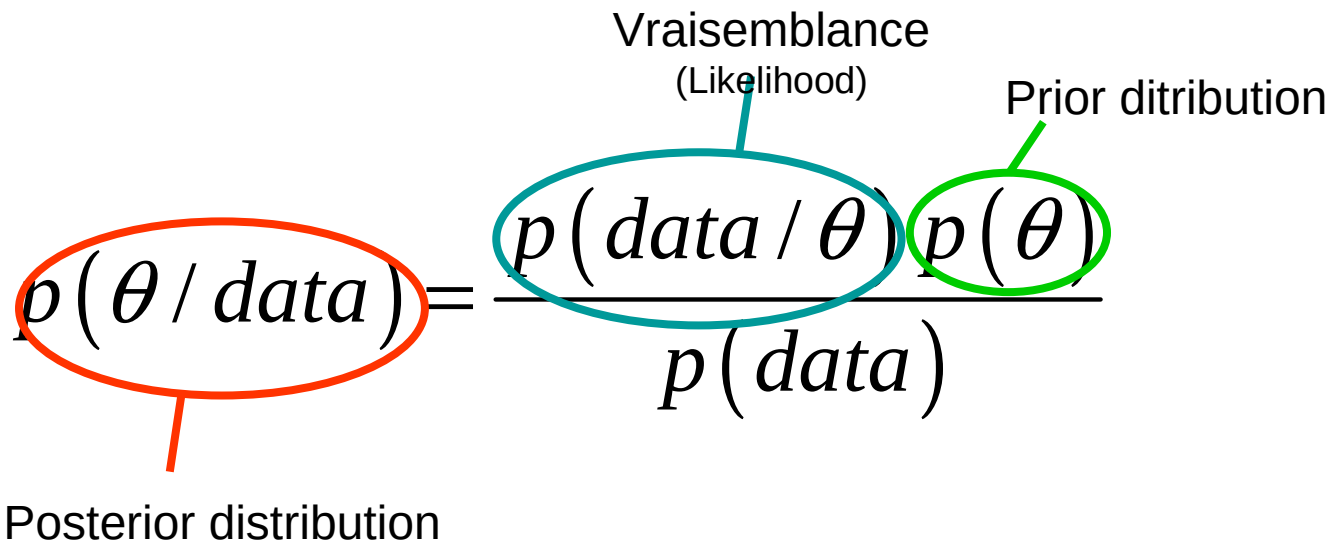
CIRAD

Objectifs et principe

Observations (données génétiques, démographiques, historiques etc…)

- Comparer des scénarios démographiques et/ou évolutifs
- Estimer au sein de ces scénarios des paramètres de la dynamique actuelle ou passée des populations (Ne, taux de recombinaison, taux de migration, date d'introduction, taux d'admixture, temps de divergence …)

Théorème de Bayes θ paramètre à estimer



Principe

$$p(\theta / data) = \frac{p(data / \theta)\, p(\theta)}{p(data)}$$

Utilisation ABC basique du théorème de Bayes

Dans la cas où la vraisemblance ne peut être écrite
(scénarios démographiques et/ ou évolutifs trop complexes)

1-Tirer une valeur du paramètre dans le prior

Ex : taux de mutation $\theta = 10^{-6}$

2-Simuler un jeu de données avec cette valeur de paramètre

Ex : générer un arbre par le modèle coalescent et distribuer des mutations le long de ses branches à une fréquence de θ

3-Calculer une statistique *S'* qui résume l'info du jeu de données simulé

Ex : le nombre de sites qui ségrègent

4-Comparer S' simulé et S réel

Si $\|s' - s\| \leq \delta$ on stocke la valeur de paramètre sinon on la rejette

5-Retour à 1 (beaucoup de fois)

L'ensemble des valeurs stockées permet de tracer la posterior distribution

Pourquoi « approximate » ?

Utilisation ABC basique du théorème de Bayes

Dans la cas où la vraisemblance ne peut être écrite
(scénarios démographiques et/ ou évolutifs trop complexes)

1-Tirer une valeur du paramètre dans le prior

2-Simuler un jeu

Par opposition au calcul exact de vraisemblance et recherche du max de vraisemblance par MCMC

distribuer des mutations le long de ses branches à une fréquence de θ

3-Calculer une statistique S' qui résume l'info du jeu de données simulé

Ex : le nombre de sites qui ségrègent

4-Comparer S' simulé et S ré

Si $\|s' - s\| \leq \delta$ on

Par opposition au méthodes rejection-sampling

on la rejette

5-Retour à 1 (beaucoup de fois)

L'ensemble des valeurs stockées permet de tracer la posterior distribution

Améliorations

- Choix de la (des) statistique(s) décrivant le jeu de données

On suppose que $p(\theta / data) \approx p(\theta / S)$

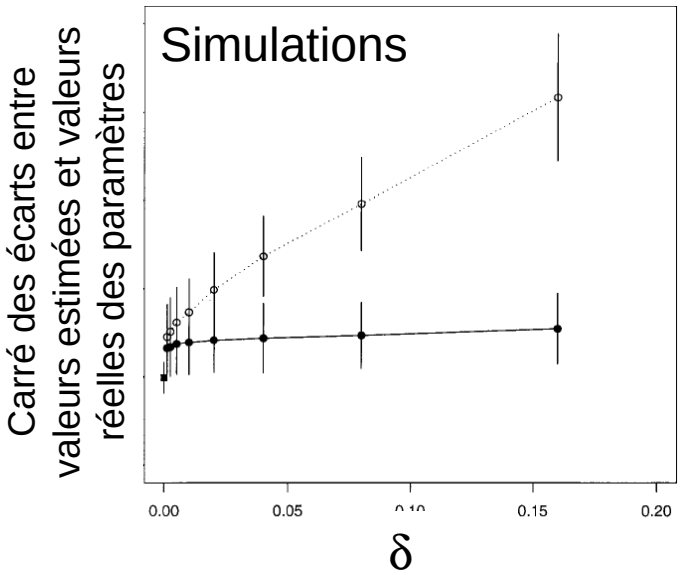
Pour le moment, pas de théorie disponible pour guider le choix de *S*

- L'estimation dépend largement du seuil de rejet δ

•Méthode par régression linéaire locale de *Beaumont et al. 2002 Genetics*

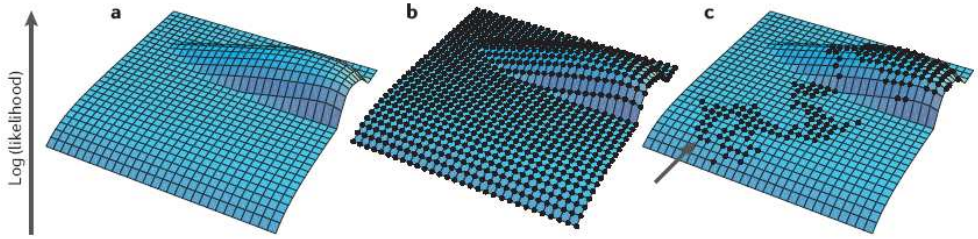
•Régression entre les valeurs testées du paramètre θ_i et $||s_i - s||$

• Améliore la méthode en pondérant les valeurs testées du paramètre θ_i par $||s_i - s||$ et en diminuant l'impact de l'écart entre s_i et s sur les estimations



- Exploration de l'espace des paramètres

Couplage possible avec des techniques de MCMC



Excoffier & heckel 2006 Nature Reviews | Genetics

Marjoram et al. 2003 PNAS
Sisson et al. 2007 PNAS

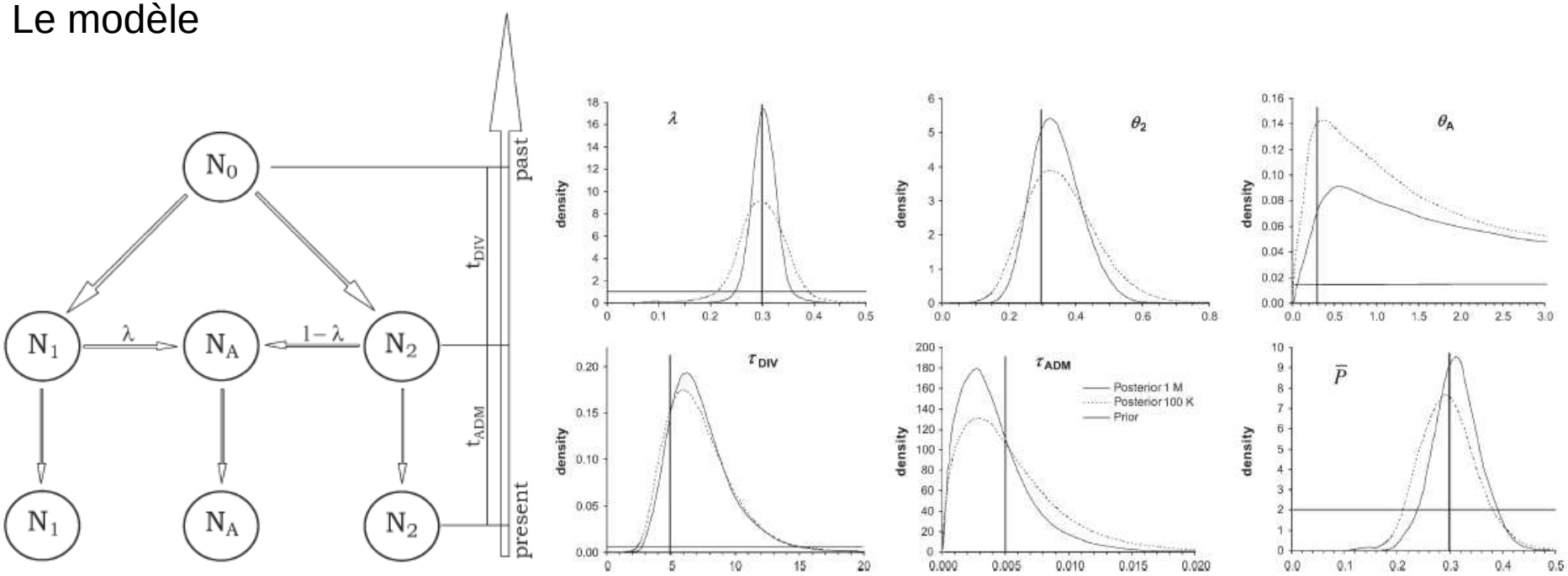
Bayesian Analysis of an Admixture Model With Mutations and Arbitrarily Linked Markers

Laurent Excoffier,*,†,1 Arnaud Estoup* and Jean-Marie Cornuet*

*Institut National de la Recherche Agronomique, Centre de Biologie et de Gestion des Populations (CBGP), Campus International de Baillarguet, 34988 Montferrier-sur-Lez Cedex, France and †Computational and Molecular Population Genetics Lab (CMPG), Zoological Institute, University of Bern, 3012 Bern, Switzerland

Genetics **169:** 1727–1738 (March 2005)

Le modèle



Application : *Apis mellifera* 3 pops (33, 19, 49 ind.) 8 microsats

Comparaison de modèles

Comparer différents scénarios démographiques ou évolutifs

Bayes factor $$B(M_1, M_2) = \frac{p(M_1 / data)\, p(M_2)}{p(M_2 / data)\, p(M_1)}$$

Nombre d'occurrences du modèle 2 parmi les simulations non rejetées

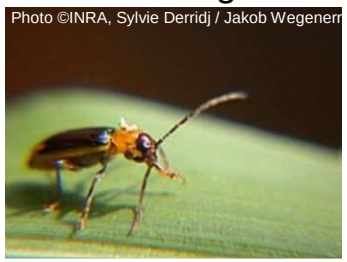
Fréquence des simulations avec le modèle 1

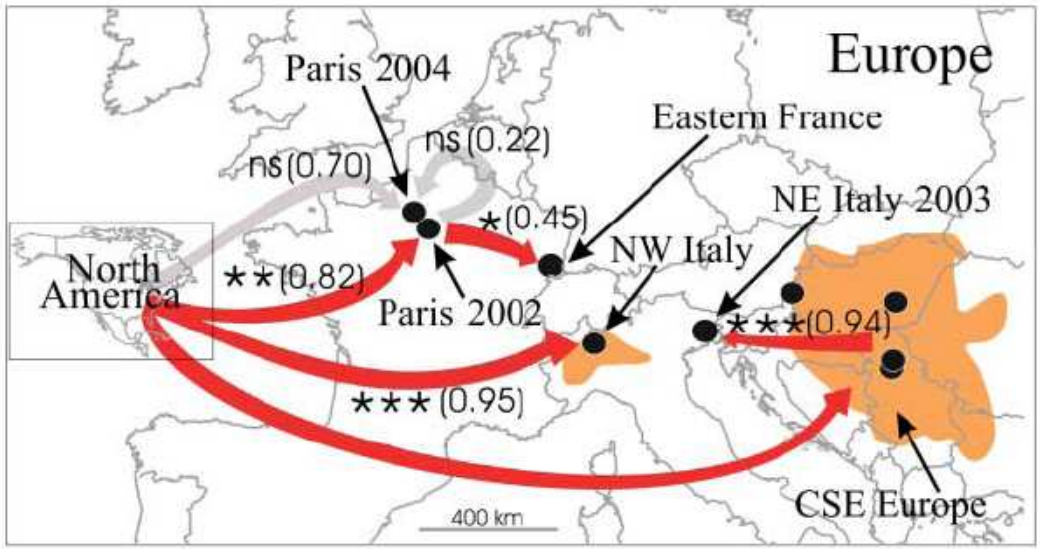
Multiple Transatlantic Introductions of the Western Corn Rootworm

Nicholas Miller,[1] Arnaud Estoup,[2] Stefan Toepfer,[3]
Denis Bourguet,[2] Laurent Lapchin,[1] Sylvie Derridj,[4]
Kyung Seok Kim,[5] Philippe Reynaud,[6]

11 NOVEMBER 2005 VOL 310 SCIENCE www.sciencemag.org

Diabrotica virgifera virgifera





Utilisations de l'ABC

Inférences de paramètres

Vitesse de propagation

Taux de mutation

Taux de recombinaison

Intensité d'un bottleneck

Taille efficace : OneSamp Tallmon et al. 2008

Taux d'admixture : Excoffier et al. 2005

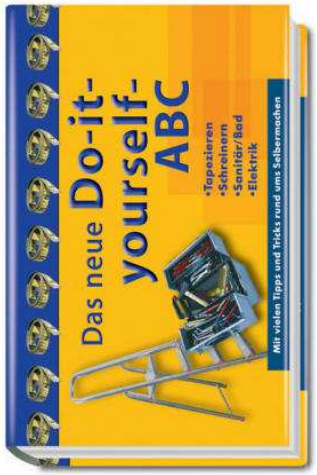
Dates (ancêtre commun, dernier bottleneck…)

Tests de scénarios démographiques et/ou évolutifs

Lieu d'introduction : Estoup et al. 2001

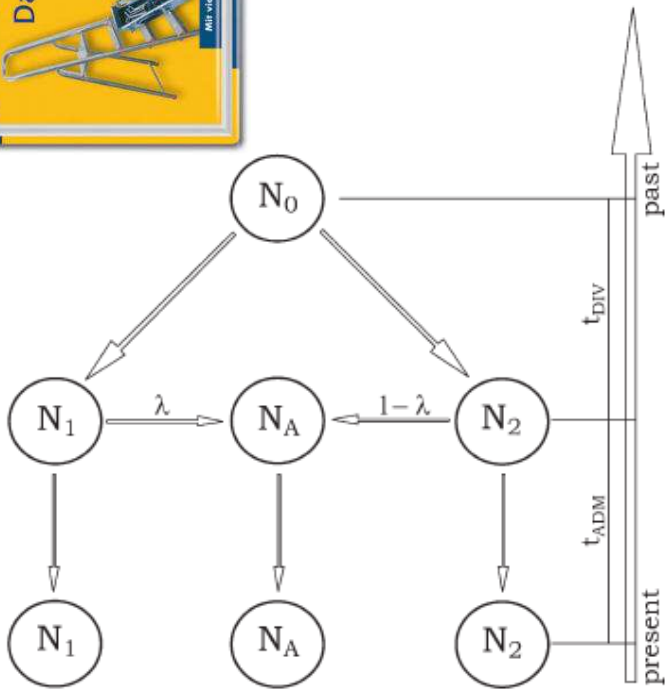
Itinéraires d'introduction : Miller et al. 2001, Pascual et al. 2007

DoItYourselfABC



Un programme convivial (« clic-clic ») pour réaliser par ABC des inférences de paramètres démographiques, historiques et génétiques à partir de données génétiques et comparer des scénarios

Construit à partir de abc-sim et abcEst (Cornuet et al. 2006 Actes du BRG)



Limites actuelles

- échelle de temps moyenne (phylogéographie, histoire avant émergence)
- données de marqueurs microsatellites
- espèces diploïdes ★
- reproduction sexuée panmictique ★
- divergence sans migration et/ou admixtures

Spécifications

- données réelles au format Genepop
- nombre de populations
- histoire présumée
- paramètres connus (exemple : dates)
- statistiques à utiliser

Sorties : bayes factors des scénarios et posterior distributions des paramètres

A MULTILOCUS PERSPECTIVE ON COLONIZATION ACCOMPANIED BY SELECTION AND GENE FLOW

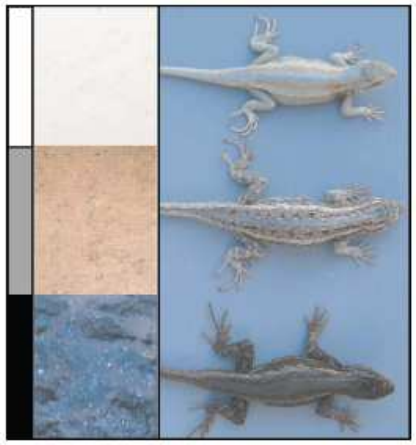
Erica Bree Rosenblum,[1,2,3] Michael J. Hickerson,[1] and Craig Moritz[1]

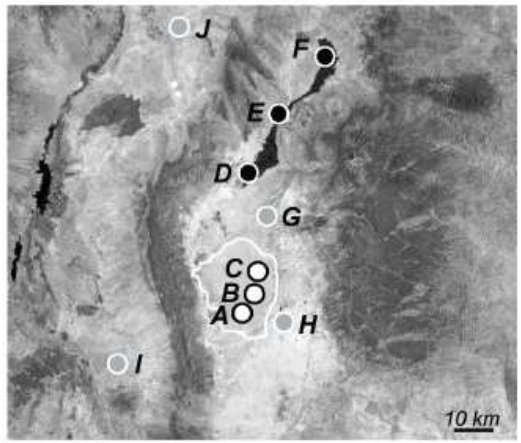
[1]*Museum of Vertebrate Zoology, University of California, Berkeley, 3101 Valley Life Sciences California 94720*

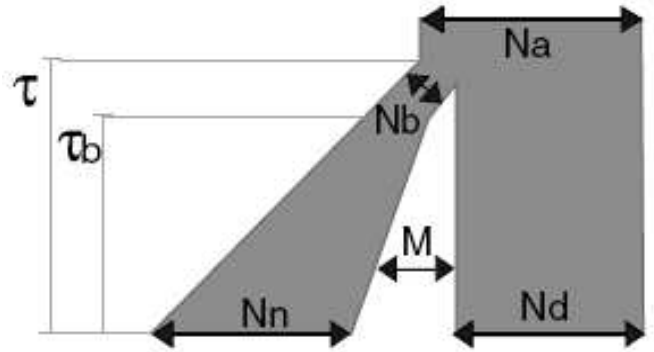
[2]*E-mail: rosenblum@berkeley.edu*

Evolution 61-12: 2971–2985

Sceloporus undulatus







Populations Sampled	mtDNA No. Inds	mtDNA No. Var. Sites	mtDNA π	mtDNA θ	nucDNA No. Inds	nucDNA No. Var Sites	nucDNA π	nucDNA θ
All Combined	89	64	0.0215	0.0156	91	191	0.0065	0.0074
Dark Soil Combined	36	57	0.0241	0.0169	37	148	0.0063	0.0068
Non-Tularosa Basin Dark Soil	17	25	0.0149	0.0091	17	105	0.0058	0.0047
Tularosa Basin Dark Soil	19	46	0.0238	0.0162	20	112	0.0064	0.0065
White Sand	29	26	0.0060	0.0082	29	119	0.0056	0.0059
Black lava	24	10	0.0026	0.0033	25	115	0.0060	0.0054