

Comportement collectif de plusieurs fronts d'avance en génétique des populations : agrégation et accélération induite

Benoît Sarels

Laboratoire Jacques-Louis Lions & Laboratoire de biologie intégrative des modèles marins
Station Biologique de Roscoff, Sorbonne Université



LBI2M



CNRS • SORBONNE UNIVERSITÉ
Station Biologique
de Roscoff



Colloque d'ouverture, 50 ans du LJLL
7 Mars 2019

Une anecdote ET un conseil touristique



Georges Teissier (1900–1972)

Georges Teissier et les affaires qui nous intéressent

<http://www.histcnrs.fr/histoire-genetique/teissier.html>

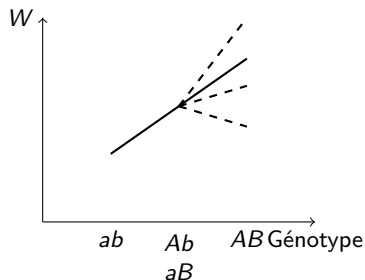
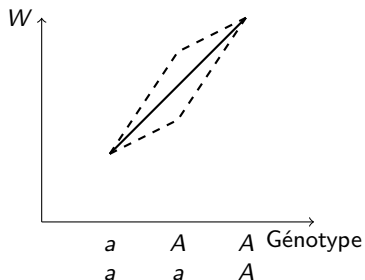
- ▶ Mathématicien de formation
- ▶ Directeur de la station biologique de Roscoff : 1945–1971
- ▶ Directeur du CNRS : 1945–1950
- ▶ «[...] Teissier n'a de cesse de vouloir imposer cette approche mathématique à l'étude biologique pour établir une "biologie mathématique", c'est-à-dire une "biologie rationnelle", comme il l'explique lui-même lors d'une conférence faite à l'École des Hautes Études de Belgique le 17 mars 1933.»
- ▶ «Dans la conclusion de son article "La statistique en Biologie", on peut en fait trouver l'expression de toute sa démarche : "le rôle que jouera la statistique en biologie s'accroîtra sans doute possible dans l'avenir [...] elle pourra même contribuer à poser de nouveaux problèmes et à étendre ainsi dans de vastes proportions le domaine actuel des sciences de la vie. Mais il faut, pour cela, que cela ne soit pas le privilège de quelques rares spécialistes plus versés dans le maniement des formules que dans celui des êtres vivants. Le statisticien et le biométricien, si loin que puissent les conduire leurs calculs, si abstraites que puissent être leurs conclusions, ne devront jamais oublier que sous les algorithmes se cache une réalité vivante et qu'un progrès véritable ne sera possible qu'autant qu'ils sauront être et rester des biologistes".»

Évolution des systèmes génétiques



Denis Roze (EBEA, Station Biologique Roscoff, CNRS), Nicolas Bierne (ISEM, Univ. Montpellier, CNRS), Alexis Simon (ISEM, Univ. Montpellier, CNRS)

- ▶ Fronts d'avance en génétique des populations
- ▶ Étude des phénomènes de spéciation
- ▶ Interactions non-linéaires : dominance, épistasie



Un locus, deux allèles

- ▶ Population dans un habitat linéaire
- ▶ Individus diploïdes
- ▶ Cycle de vie
 1. Fusion au hasard des gamètes en tout point → naissance de nouveaux individus
 2. Production de gamètes (selon la fitness)
 3. Migration des gamètes
- ▶ $p(x, t)$ fréquence de l'allèle A
 $1 - p(x, t)$ fréquence de l'allèle a
- ▶ W_{aa}, W_{Aa}, W_{AA} fitness des individus

$$W_{aa} = 1$$

$$W_{Aa} = 1 + s_A - S_A$$

$$W_{AA} = 1 + 2s_A$$

avec $0 \leq s_A < S_A$ (sous-dominance)

Équation de réaction-diffusion bistable

- ▶ Supposant $S_A \ll 1$, on obtient

$$\partial_t p = \left(s_A + S_A(2p - 1) \right) p(1 - p) + \frac{\sigma^2}{2} \partial_{xx}^2 p$$

ou

$$\partial_t p = 2S_A p(p - \alpha)(1 - p) + \frac{\sigma^2}{2} \partial_{xx}^2 p \quad (\star_p)$$

où $\alpha = \frac{S_A - s_A}{2S_A} > 0$

- ▶ Deux locus, deux allèles (mêmes hypothèses)
 $q(x, t)$ fréquence de l'allèle B
 $1 - q(x, t)$ fréquence de l'allèle b
- ▶ W_{bb}, W_{Bb}, W_{BB} fitness des individus

$$W_{bb} = 1$$

$$W_{Bb} = 1 + s_B - S_B$$

$$W_{BB} = 1 + 2s_B$$

avec $s_B \geq 0$ et $S_B > s_B$

- ▶ $W_{x,y} = W_x W_y$

Deux locus, deux allèles

► Possibilité de recombinaison pendant la méiose (taux de recombinaison $0 < r \leq \frac{1}{2}$)

► Individu $\begin{array}{c|c} A & a \\ B & b \end{array}$ produit:

1. $\frac{A}{B}$ en proportion $\frac{1-r}{2}$

2. $\frac{a}{b}$ en proportion $\frac{1-r}{2}$

3. $\frac{A}{b}$ en proportion $\frac{r}{2}$

4. $\frac{a}{B}$ en proportion $\frac{r}{2}$

► y_1, y_2, y_3, y_4 fréquences des gamètes AB, Ab, aB, ab (resp.)

► $z_{i,j}$ fréquence des individus diploïdes formés par fusion des gamètes i et j :

$$z_{i,j} = \begin{cases} 2y_i y_j & \text{si } i \neq j \\ y_i^2 & \text{sinon} \end{cases}$$

On obtient

$$\begin{cases} y_1 = z_{1,1} + \frac{1}{2}(z_{1,2} + z_{1,3} + (1-r)z_{1,4} + rz_{2,3}) \\ y_2 = z_{2,2} + \frac{1}{2}(z_{2,1} + z_{2,4} + (1-r)z_{2,3} + rz_{1,4}) \\ y_3 = z_{3,3} + \frac{1}{2}(z_{3,1} + z_{3,4} + (1-r)z_{3,2} + rz_{1,4}) \\ y_4 = z_{4,4} + \frac{1}{2}(z_{4,2} + z_{4,3} + (1-r)z_{4,1} + rz_{2,3}) \end{cases}$$

Système d'évolution

- Supposant $S_A, S_B, r \ll 1$, on obtient

$$\begin{cases} \frac{dy_1}{dt} = f_1(y) := y_1 P_1(y) + r(y_2 y_3 - y_1 y_4) \\ \frac{dy_2}{dt} = f_2(y) := y_2 P_2(y) + r(y_1 y_4 - y_2 y_3) \\ \frac{dy_3}{dt} = f_3(y) := y_3 P_3(y) + r(y_1 y_4 - y_2 y_3) \\ \frac{dy_4}{dt} = f_4(y) := y_4 P_4(y) + r(y_2 y_3 - y_1 y_4) \end{cases}$$

- Le changement de variables

$$p = y_1 + y_2 \quad \text{fréquence de l'allèle A}$$

$$q = y_1 + y_3 \quad \text{fréquence de l'allèle B}$$

$$D = y_1 y_4 - y_2 y_3 \quad \text{déséquilibre de liaison}$$

donne

$$\begin{cases} \frac{dp}{dt} = 2S_A p(p - \alpha)(1 - p) + 2S_B (q - \beta)D \\ \frac{dq}{dt} = 2S_B q(q - \beta)(1 - q) + 2S_A (p - \alpha)D \\ \frac{dD}{dt} = -D \left(r + 4S_A (p - \alpha) \left(p - \frac{1}{2} \right) + 4S_B (q - \beta) \left(q - \frac{1}{2} \right) \right) \end{cases}$$

avec

$$\alpha = \frac{S_A - s_A}{2S_A}$$

$$\beta = \frac{S_B - s_B}{2S_B}$$

Système de réaction-diffusion

Avec la migration

$$\partial_t y_i = f_i(y) + \partial_{xx}^2 y_i$$

on obtient

$$\begin{cases} \partial_t p = f_1(y) + \partial_{xx}^2 y_1 + f_2(y) + \partial_{xx}^2 y_2 \\ \partial_t q = f_1(y) + \partial_{xx}^2 y_1 + f_3(y) + \partial_{xx}^2 y_3 \\ \partial_t D = y_1 f_4(y) + y_1 \partial_{xx}^2 y_4 + y_4 f_1(y) + y_4 \partial_{xx}^2 y_1 - y_2 f_3(y) - y_2 \partial_{xx}^2 y_3 - y_3 f_2(y) - y_3 \partial_{xx}^2 y_2 \end{cases}$$

Puis

$$\begin{aligned} y_1 \partial_{xx}^2 y_4 + y_4 \partial_{xx}^2 y_1 - y_2 \partial_{xx}^2 y_3 - y_3 \partial_{xx}^2 y_2 &= \partial_{xx}^2 D - (2\partial_x y_1 \partial_x y_4 - 2\partial_x y_2 \partial_x y_3) \\ &= \partial_{xx}^2 D + 2\partial_x p \partial_x q \end{aligned}$$

$$\begin{cases} \partial_t p = 2S_A p(p - \alpha)(1 - p) + 2S_B (q - \beta)D + \partial_{xx}^2 p \\ \partial_t q = 2S_B q(q - \beta)(1 - q) + 2S_A (p - \alpha)D + \partial_{xx}^2 q \\ \partial_t D = -D \left(r + 4S_A (p - \alpha) \left(p - \frac{1}{2} \right) + 4S_B (q - \beta) \left(q - \frac{1}{2} \right) \right) + 2\partial_x p \partial_x q + \partial_{xx}^2 D \\ p(\cdot, 0) = p_0(x) \quad q(\cdot, 0) = q_0(x) \quad D(\cdot, 0) = 0 \end{cases}$$

($\star_{p,q,D}$)

Rappel sur l'équation de réaction-diffusion bistable

$$\partial_t p = 2S_A p(p - \alpha)(1 - p) + \partial_{xx}^2 p \quad (\star_p)$$

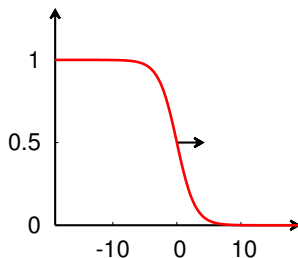
- ▶ Front progressif

$$p(x, t) = \tilde{p}(z) \quad \text{avec} \quad z = \sqrt{S_A}(x - ct)$$

- ▶ Vitesse $c = \sqrt{S_A}(1 - 2\alpha) = \frac{s_A}{\sqrt{S_A}}$

- ▶ Profil

$$\tilde{p}(z) = \frac{1}{1 + \exp(z)}$$



- ▶ Propagation de l'allèle avantageux A
- ▶ Vocabulaire : «cline de fréquence allélique»

Simulation #1 : fronts progressifs ?

Même direction, même vitesse

$$s_A = 0.08,$$

$$S_A = 0.1,$$

$$c_p = \frac{s_A}{\sqrt{S_A}} = 0.25$$

$$s_B = 0.08,$$

$$S_B = 0.1,$$

$$c_q = \frac{s_B}{\sqrt{S_B}} = 0.25$$

$$r = 0.01$$

Simulation #2 : attention à la condition initiale !

Même direction, même vitesse

$$s_A = 0.08,$$

$$s_B = 0.08,$$

$$r = 0.01$$

$$S_A = 0.1,$$

$$S_B = 0.1,$$

$$c_p = \frac{s_A}{\sqrt{S_A}} = 0.25$$

$$c_q = \frac{s_B}{\sqrt{S_B}} = 0.25$$

Simulation #3

Même direction, écart de vitesse

$$s_A = 0.08,$$

$$s_B = 0,$$

$$r = 0.01$$

$$S_A = 0.1,$$

$$S_B = 0.1,$$

$$c_p = \frac{s_A}{\sqrt{S_A}} = 0.25$$

$$c_q = \frac{s_B}{\sqrt{S_B}} = 0$$

Simulation #4

Même direction, (plus petit) écart de vitesse

$$s_A = 0.04,$$

$$s_B = 0,$$

$$r = 0.01$$

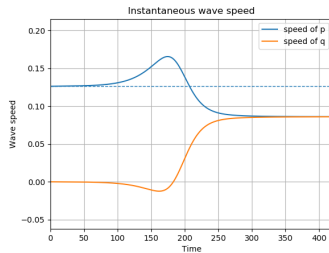
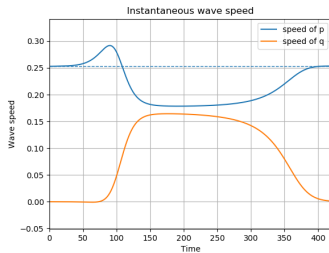
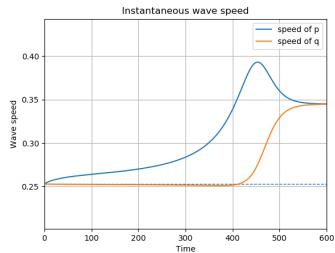
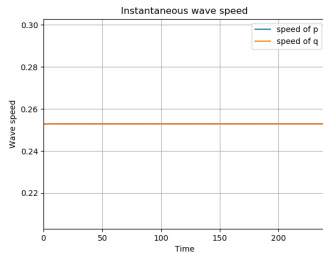
$$S_A = 0.1,$$

$$S_B = 0.1,$$

$$c_p = \frac{s_A}{\sqrt{S_A}} = 0.13$$

$$c_q = \frac{s_B}{\sqrt{S_B}} = 0$$

Simulations : une vue condensée



Comment sont calculées les vitesses ?

- ▶ À partir des solutions $(p(x_n, t))_{0 \leq n \leq N}$ et $(q(x_n, t))_{0 \leq n \leq N}$
- ▶ Approcher (à chaque pas de temps) les solutions numériques par un front de référence, *i.e.* trouver $x_p(t)$ qui minimise

$$\left\| p(x, t) - \frac{1}{1 + \exp(\sqrt{S_A}(x - x_p(t)))} \right\|_{\ell^2}^2 = \sum_{n=0}^N \left(p(x_n, t) - \frac{1}{1 + \exp(\sqrt{S_A}(x_n - x_p(t)))} \right)^2$$

et $x_q(t)$ qui minimise

$$\left\| q(x, t) - \frac{1}{1 + \exp(\sqrt{S_B}(x - x_q(t)))} \right\|_{\ell^2}^2 = \sum_{n=0}^N \left(q(x_n, t) - \frac{1}{1 + \exp(\sqrt{S_B}(x_n - x_q(t)))} \right)^2$$

- ▶ S'assurer que les erreurs restent petites
 $\varepsilon(x_p(t)) < 10^{-5}$ et $\varepsilon(x_q(t)) < 10^{-5}$
- ▶ Vitesses définies par $c_p(t) = x_p'(t)$ et $c_q(t) = x_q'(t)$

Fronts progressifs généralisés en réaction–diffusion

- ▶ Hétérogénéité dans le domaine : environnement, milieu...

$$\partial_t p = d(x)\Delta p + f(x, p)$$

- ▶ Front de transition (BERESTYCKI, HAMEL, MATANO 2009):

1. solution globale en temps
2. qui relie deux états d'équilibre
3. de largeur bornée uniformément

- ▶ Front progressif généralisé :

1. solution globale en temps
2. proche d'un front de référence \tilde{p}

$$p(x, t) \approx \tilde{p}\left(\frac{x - x_0(t)}{w}\right)$$

$$p(x, t) \approx \tilde{p}\left(\frac{x - x_0(t)}{w(t)}\right)$$

- ▶ Notions de position (et largeur) instantanée(s), donc de **vitesse** instantanée

Fronts progressifs généralisés pour notre problème

- ▶ Hypothèse sur p

$$p(x, t) \approx \tilde{p} \left(\sqrt{S_A}(x - x_p(t)) \right)$$

- ▶ Hypothèse sur q

$$q(x, t) \approx \tilde{q} \left(\sqrt{S_B}(x - x_q(t)) \right)$$

- ▶ 2 coordonnées collectives : x_p et x_q

- ▶ Ici,

$$\tilde{p}(z) = \tilde{q}(z) = \frac{1}{1 + \exp(z)} =: k(z)$$

Cas $S_A, S_B \ll r$: faible déséquilibre de liaison

- ▶ On s'attend à ce que D reste proche de

$$D \approx \frac{2}{r} \partial_x p \partial_x q$$

- ▶ Supposons

$$p(x, t) = k(z_p) = k\left(\sqrt{S_A}(x - x_p(t))\right) \quad q(x, t) = k(z_q) = k\left(\sqrt{S_B}(x - x_q(t))\right)$$

et calculons

$$\begin{aligned} \partial_t p &= -\sqrt{S_A} x'_p(t) k'(z_p) & \partial_x p &= \sqrt{S_A} k'(z_p) & \partial_{xx}^2 p &= S_A k''(z_p) \\ \partial_t q &= -\sqrt{S_B} x'_q(t) k'(z_q) & \partial_x q &= \sqrt{S_B} k'(z_q) & \partial_{xx}^2 q &= S_B k''(z_q) \end{aligned}$$

- ▶ En remplaçant dans les deux premières équations de $(\star_{p,q,D})$

$$\begin{cases} -\sqrt{S_A} x'_p(t) k'(z_p) = 2S_A R_\alpha(k(z_p)) + \frac{2S_B \sqrt{S_A S_B}}{r} \left(k(z_p) - \beta \right) k'(z_p) k'(z_q) + S_A k''(z_p) \\ -\sqrt{S_B} x'_q(t) k'(z_q) = 2S_B R_\beta(k(z_q)) + \frac{2S_A \sqrt{S_A S_B}}{r} \left(k(z_q) - \alpha \right) k'(z_q) k'(z_p) + S_B k''(z_q) \end{cases}$$

Cas $S_A, S_B \ll r$: faible déséquilibre de liaison (cont.)

- ▶ En intégrant sur le domaine, tenant compte des relations

$$\int_{-\infty}^{+\infty} k'(z) dz = 1 \qquad \int_{-\infty}^{+\infty} k''(z) dz = 0$$

$$\int_{-\infty}^{+\infty} R_\alpha(k'(z)) dz = \frac{1 - 2\alpha}{2} = \frac{s_A}{S_A} \qquad \int_{-\infty}^{+\infty} R_\beta(k'(z)) dz = \frac{1 - 2\beta}{2} = \frac{s_B}{S_B}$$

- ▶ On obtient les équations d'évolution suivantes

$$\left\{ \begin{array}{l} x'_p(t) = \frac{s_A}{\sqrt{S_A}} + \frac{4S_B}{r} \int_{-\infty}^{+\infty} k(z)k'(z)k' \left(\sqrt{S_A}(x_q - x_p) + z\sqrt{\frac{S_A}{S_B}} \right) dz \\ \qquad \qquad \qquad - \frac{2(S_B - s_B)}{r} \int_{-\infty}^{+\infty} k'(z)k' \left(\sqrt{S_A}(x_q - x_p) + z\sqrt{\frac{S_A}{S_B}} \right) dz \\ x'_q(t) = \frac{s_B}{\sqrt{S_B}} + \frac{4S_A}{r} \int_{-\infty}^{+\infty} k(z)k'(z)k' \left(\sqrt{S_B}(x_p - x_q) + z\sqrt{\frac{S_B}{S_A}} \right) dz \\ \qquad \qquad \qquad - \frac{2(S_A - s_A)}{r} \int_{-\infty}^{+\infty} k'(z)k' \left(\sqrt{S_B}(x_p - x_q) + z\sqrt{\frac{S_B}{S_A}} \right) dz \end{array} \right.$$

Cas $S_A, S_B \ll r$: faible déséquilibre de liaison (cont.)

► Le système a donc la forme

$$\begin{cases} x'_p = c_p + \frac{2S_B}{r} f_1\left(\sqrt{S_A}(x_q - x_p)\right) + \frac{2S_B}{r} f_2\left(\sqrt{S_A}(x_q - x_p)\right) \\ x'_q = c_q + \frac{2S_A}{r} f_3\left(\sqrt{S_B}(x_p - x_q)\right) + \frac{2S_A}{r} f_4\left(\sqrt{S_B}(x_p - x_q)\right) \end{cases}$$

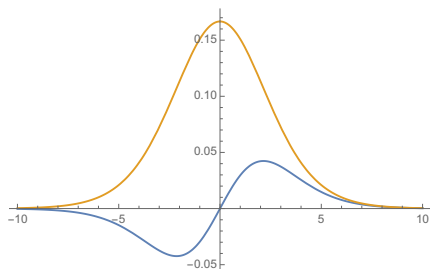


Figure: Termes f_1 en bleu et f_2 en jaune

- Permet de répondre à la question de l'agrégation des fronts ?
- Nécessité d'un développement asymptotique

Le cas de N gènes et le lien avec le modèle de Kuramoto

- ▶ Avec N locus, en première approximation, si $s_{A_i} = s_i$ et $S_{A_i} = S$, on obtient

$$x'_i = c_i + \frac{S}{r} \sum_{j=1}^N f_1(x_j - x_i) + \frac{1}{r} \sum_{j=1, j \neq i}^N s_j f_2(x_j - x_i), \quad i \in \{1, \dots, N\}$$

- ▶ (Forte) similitude avec le modèle de Kuramoto

$$\theta'_i = \omega_i + \frac{K}{N} \sum_{j=1}^N \sin(\theta_j - \theta_i), \quad i \in \{1, \dots, N\}$$

- ▶ Différences : espace des phases, facteur $1/N$

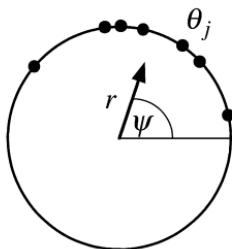


Figure: $re^{i\psi} = \frac{1}{N} \sum e^{i\theta_j}$



Merci pour votre attention !



LBI2M



CNRS • SORBONNE UNIVERSITÉ
Station Biologique
de Roscoff

