

Modelling Evolution

From stochastic spatial models to mean curvature flow

Alison Etheridge

Nic Freeman, Sarah Penington

Hybrid Zones

A hybrid zone is a narrow geographic region where two genetically distinct populations are found close together and hybridise to produce offspring of mixed ancestry.

They are maintained by a balance between selection and dispersal.



With thanks to Nick Barton and his group

A mathematical model

Individuals carry two copies of a gene that occurs as a or A .

Hardy-Weinberg proportions: \bar{w} = proportion of a -alleles,

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline \bar{w}^2 & 2\bar{w}(1 - \bar{w}) & (1 - \bar{w})^2 \end{array}$$

Relative fitnesses:

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline 1 & 1 - s & 1 \end{array}$$

Reproduction

- ▶ Each heterozygote produces $(1 - s)$ times as many germ cells (cells of same genotype) as a homozygote;
- ▶ Germ cells split into effectively infinite pool of gametes (containing just one copy of gene), with proportion of type a

$$\begin{aligned}\frac{\bar{w}^2 + \bar{w}(1 - \bar{w})(1 - s)}{1 - 2s\bar{w}(1 - \bar{w})} &= (1 - s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2) \\ &= \bar{w} + s\bar{w}(1 - \bar{w})(2\bar{w} - 1) + \mathcal{O}(s^2).\end{aligned}$$

In an infinite population, if $s = \frac{\alpha}{N}$ (where N is large), measuring time in units of N generations,

$$\frac{d\bar{w}}{dt} = \alpha\bar{w}(1 - \bar{w})(2\bar{w} - 1).$$

Reproduction

- ▶ Each heterozygote produces $(1 - s)$ times as many germ cells (cells of same genotype) as a homozygote;
- ▶ Germ cells split into effectively infinite pool of gametes (containing just one copy of gene), with proportion of type a

$$\begin{aligned}\frac{\bar{w}^2 + \bar{w}(1 - \bar{w})(1 - s)}{1 - 2s\bar{w}(1 - \bar{w})} &= (1 - s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2) \\ &= \bar{w} + s\bar{w}(1 - \bar{w})(2\bar{w} - 1) + \mathcal{O}(s^2).\end{aligned}$$

In an infinite population, if $s = \frac{\alpha}{N}$ (where N is large), measuring time in units of N generations,

Add dispersal:

$$\frac{\partial w}{\partial t} = \frac{m}{2}\Delta w + \alpha w(1 - w)(2w - 1).$$

The Allen-Cahn equation: zooming out

$$\frac{\partial w}{\partial t} = \frac{m}{2} \Delta w + \alpha w(1-w)(2w-1).$$

Applying a diffusive rescaling $t \mapsto \frac{t}{\epsilon^2}$, $x \mapsto \frac{x}{\epsilon}$, the Allen-Cahn equation becomes

$$\frac{\partial w}{\partial t} = \frac{m}{2} \Delta w + \frac{\alpha}{\epsilon^2} w(1-w)(2w-1).$$

For convenience, set $m = 2$, $\alpha = 1$.

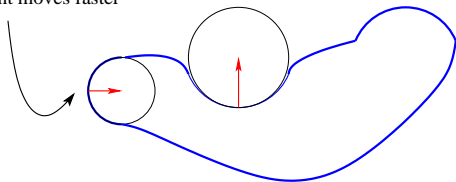
For sufficiently regular initial conditions, as $\epsilon \rightarrow 0$, the solution converges to the indicator function of a region whose boundary evolves according to *curvature flow*.

Curvature flow

- ▶ $\Gamma_t : S^1 \rightarrow \mathbb{R}^2$ smooth embeddings;
- ▶ $\mathbf{n}_t(u)$ unit (inward) normal vector to Γ_t at u ;
- ▶ $\kappa = \kappa_t(u)$ curvature of Γ_t at u .

$$\frac{\partial \Gamma_t(u)}{\partial t} = \kappa_t(u) \mathbf{n}_t(u). \quad \text{Defined up to fixed time } T$$

This point moves faster



Curvature Flow by Matt Dunlop (Warwick)

The Allen-Cahn equation and curvature flow

Let $d(x, t)$ be the signed distance from x to Γ_t . Choose w_0 such that $\Gamma_0 = \{x \in \mathbb{R}^2 : w_0(x) = \frac{1}{2}\}$, $w_0 < \frac{1}{2}$ inside Γ and $> \frac{1}{2}$ outside.

$$\frac{\partial w}{\partial t} = \Delta w + \frac{1}{\varepsilon^2} w(1-w)(2w-1).$$

Theorem (Chen 1992)

Fix $T^* \in (0, T)$. Let $k \in \mathbb{N}$. There exists $\varepsilon(k) > 0$, and $a(k), c(k) \in (0, \infty)$ such that for all $\varepsilon \in (0, \varepsilon(k))$ and t satisfying $a\varepsilon^2 |\log \varepsilon| \leq t \leq T^*$,

1. for x such that $d(x, t) \geq c\varepsilon |\log \varepsilon|$, we have $w(t, x) \geq 1 - \varepsilon^k$;
2. for x such that $d(x, t) \leq -c\varepsilon |\log \varepsilon|$, we have $w(t, x) \leq \varepsilon^k$.

Adding noise?

Hairer, Ryser & Weber (2012), $d = 2$ ($v = 2w - 1$)

$$dv = (\Delta v + v - v^3)dt + \sigma dW,$$

W a mollified space-time white noise.

- ▶ If the mollifier is removed, solutions converge weakly to zero;
- ▶ if intensity of W simultaneously converges to zero sufficiently quickly, recover the deterministic equation.

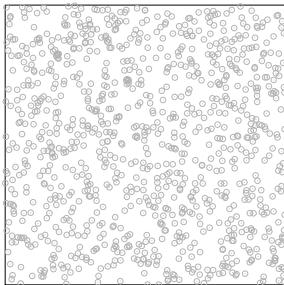
Will hybrid zones still evolve approximately according to curvature flow in the presence of random genetic drift?

... additive white noise is not a good model of genetic drift

The pain in the torus

The Wright-Malécot model

Average one offspring per individual; location of each offspring independent Gaussian pick around position of parent

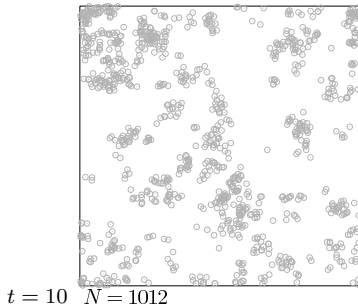
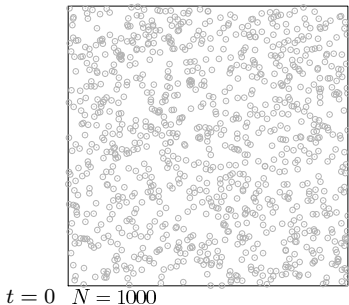


$t = 0$ $N = 1000$

The pain in the torus

The Wright-Malécot model

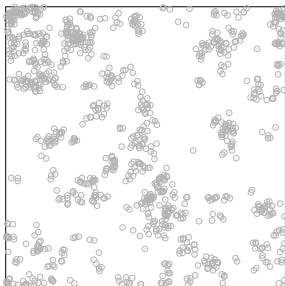
Average one offspring per individual; location of each offspring independent Gaussian pick around position of parent



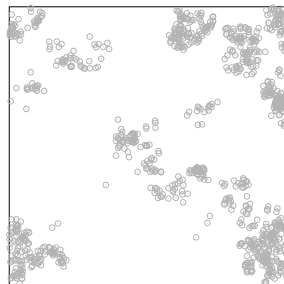
The pain in the torus

The Wright-Malécot model

Average one offspring per individual; location of each offspring independent Gaussian pick around position of parent



$t = 10$ $N = 1012$

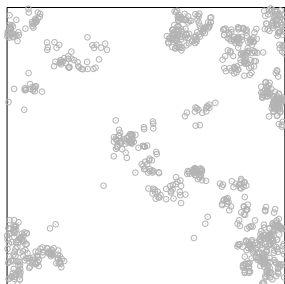


$t = 100$ $N = 972$

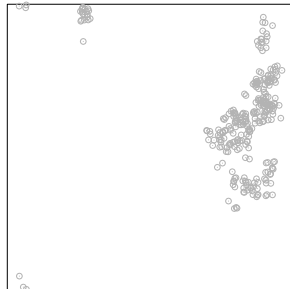
The pain in the torus

The Wright-Malécot model

Average one offspring per individual; location of each offspring independent Gaussian pick around position of parent



$t = 100$ $N = 972$

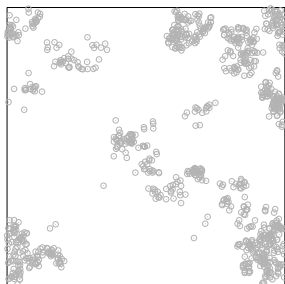


$t = 1000$ $N = 293$

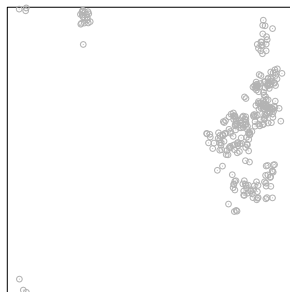
The pain in the torus

“The pain in the torus” Felsenstein (1975)

In $d = 1, 2$, need *local regulation* to prevent clumping



$t = 100$ $N = 972$



$t = 1000$ $N = 293$

The pain in the torus

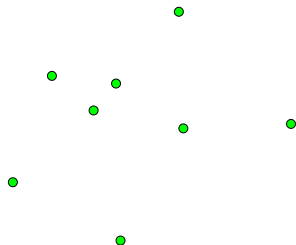
“The pain in the torus” Felsenstein (1975)

In $d = 1, 2$, need *local regulation* to prevent clumping

Nonetheless, over intermediate scales, the Wright-Malécot formula describes the way that correlations in frequencies of different genetic types decay with sampling distance.

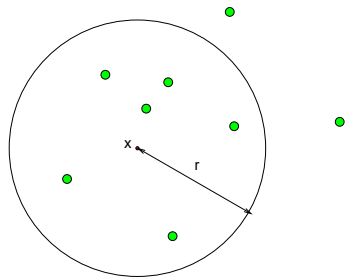
An individual based model

- ▶ Start with Poisson point process intensity λdx ;



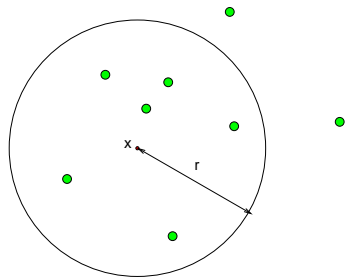
An individual based model

- ▶ Start with Poisson point process intensity λdx ;
- ▶ At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r ;



An individual based model

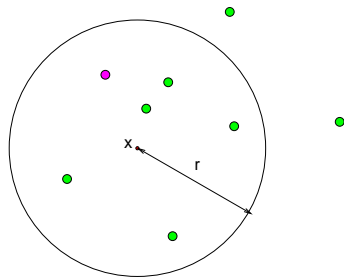
- ▶ Start with Poisson point process intensity λdx ;
- ▶ At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r ;



If ball empty, do nothing

An individual based model

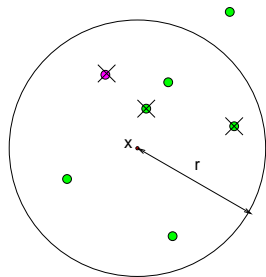
- ▶ Start with Poisson point process intensity λdx ;
- ▶ At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r ;
- ▶ Choose parent uniformly at random from individuals in $B(x, r)$;



If ball empty, do nothing

An individual based model

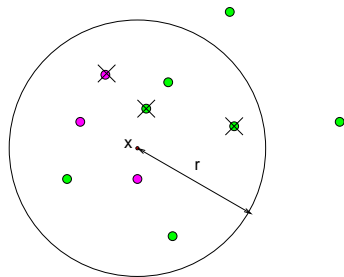
- ▶ Start with Poisson point process intensity λdx ;
- ▶ At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r ;
- ▶ Choose parent uniformly at random from individuals in $B(x, r)$;
- ▶ Each individual in region dies with probability $u \sim \nu_r(du)$;



If ball empty, do nothing

An individual based model

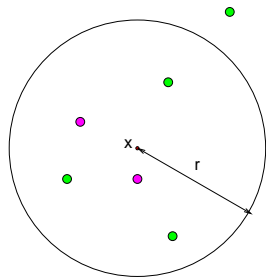
- ▶ Start with Poisson point process intensity λdx ;
- ▶ At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r ;
- ▶ Choose parent uniformly at random from individuals in $B(x, r)$;
- ▶ Each individual in region dies with probability $u \sim \nu_r(du)$;
- ▶ New individuals born according to a Poisson $\lambda u \mathbf{1}_{B(x,r)} dy$.



If ball empty, do nothing

An individual based model

- ▶ Start with Poisson point process intensity λdx ;
- ▶ At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r ;
- ▶ Choose parent uniformly at random from individuals in $B(x, r)$;
- ▶ Each individual in region dies with probability $u \sim \nu_r(du)$;
- ▶ New individuals born according to a Poisson $\lambda u \mathbf{1}_{B(x,r)} dy$.



If ball empty, do nothing

Some remarks

- ▶ Population size is locally regulated.
- ▶ Individual's probability to reproduce declines with local population density.

Theorem If $\lambda > \lambda_c$ there is an ergodic stationary distribution.

Genealogical trees are not so easy to write down.

To overcome this, let $\lambda \rightarrow \infty$, but without scaling the Poisson point process (retains signature of small neighbourhood size).

The (neutral) spatial Lambda Fleming-Viot process

State $\{w(t, x), x \in \mathbb{R}^d, t \geq 0\}$. (Proportion of type a)

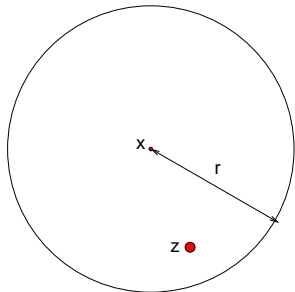
Fix $u \in (0, 1)$, μ a finite measure on $[0, \mathcal{R})$. Π Poisson point process rate $dt \otimes dx \otimes \mu(dr)$ on $[0, \infty) \times \mathbb{R}^d \times [0, \infty)$.

Dynamics: for each $(t, x, r) \in \Pi$,

- ▶ $z \sim U(\mathcal{B}_r(x))$
- ▶ $K \sim \text{Ber}(w(t-, z))$.

For all $y \in \mathcal{B}_r(x)$,

$$w(t, y) = (1 - u)w(t-, y) + uK.$$



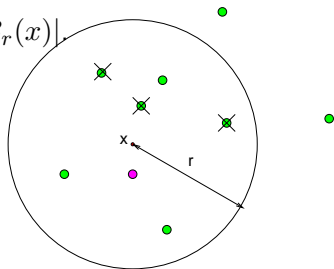
Backwards in time

- ▶ A *single* ancestral lineage evolves in series of jumps with intensity

$$dt \otimes \int_{(|x|/2, \infty)} \int_{[0,1]} \frac{L_r(x)}{C_d r^d} u \nu_r(du) \mu(dr) dx$$

on $\mathbb{R}_+ \times \mathbb{R}^d$ where $L_r(x) = |B_r(0) \cap B_r(x)|$.

- ▶ Lineages can coalesce when hit by same 'event'.



Adding selection: majority voting

Recall proportion of germ cells that are type a is

$$(1 - s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2)$$

First observe that

$$3\bar{w}^2 - \bar{w}^3 = (w^3 + 3w^2(1 - w))$$

\leadsto *majority voting rule.*

Adding selection: majority voting

Recall proportion of germ cells that are type a is

$$(1 - s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2)$$

First observe that

$$3\bar{w}^2 - \bar{w}^3 = (w^3 + 3w^2(1 - w))$$

\leadsto *majority voting rule.*

- ▶ (i) Neutral events rate $\propto (1 - s)$, selective events rate $\propto s$.
- ▶ At **selective** reproduction events, sample **three** potential parents. If types aaa or aaA , then an a reproduces, otherwise an A does.

Noisy hybrid zones

With thanks to Nic Freeman and Jerome Kelleher



A first scaling result

Specialise to $\mu(dr) = \delta_R(dr)$, for some fixed $R > 0$.

$$u_n = \frac{u}{n^{1-2\beta}}, \quad s_n = \frac{\rho}{n^{2\beta}} \quad \text{and} \quad w_t^n(x) := w_{nt}(n^\beta x).$$

Theorem (Modification of Etheridge, Véber & Yu, 2014)
Suppose that $\beta \in (0, 1/3)$, and that w_0^n converges weakly to some w^0 . Then $(w_t^n)_{t \geq 0}$ converges weakly to (weak) solution of

$$\frac{\partial w}{\partial t} = \frac{m}{2} \Delta w + \alpha w(1-w)(2w-1).$$

where m and α are explicit, with

$$m \propto u, \quad \alpha \propto u\rho.$$

What if $\rho \rightarrow \infty$ as $n \rightarrow \infty$?

Main result

$$u_n = \frac{u}{n^{1-2\beta}}, \quad s_n = \frac{1}{\varepsilon_n^2} \frac{1}{n^{2\beta}} \quad \text{and} \quad w_t^n(x) := w_{nt}(n^\beta x).$$

Theorem (Etheridge, Freeman & Penington, 2016)

Suppose $\beta \in (0, 1/4)$ and ε_n is such that $\varepsilon_n \rightarrow 0$ and $(\log n)^{1/2} \varepsilon_n \rightarrow \infty$ as $n \rightarrow \infty$. Set $w_0^n(x) = w_0(n^{-\beta} x)$.

For $k \in \mathbb{N}$ there exist $n(k) < \infty$, and $a(k), d(k) \in (0, \infty)$ such that for all $n \geq n(k)$ and all t satisfying $a\varepsilon_n^2 |\log \varepsilon_n| \leq t \leq T^*$,

1. for almost every x such that $d(x, \sigma^2 t) \geq d\varepsilon_n |\log \varepsilon_n|$, we have $\mathbb{E}[w_t^n(x)] \geq 1 - \varepsilon_n^k$.
2. for x such that $d(x, \sigma^2 t) \leq -d\varepsilon_n |\log \varepsilon_n|$, we have $\mathbb{E}[w_t^n(x)] \leq \varepsilon_n^k$.

where σ^2 is known explicitly.

Majority voting in (Historical) BBM

$\mathbf{W}(t)$ = historical ternary BBM.

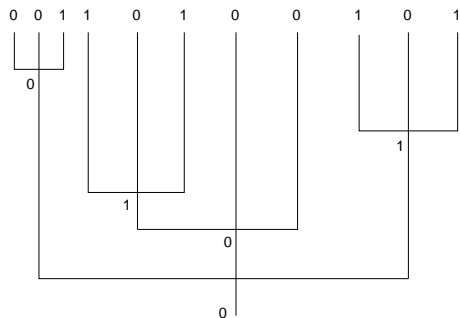
For a fixed function $p : \mathbb{R}^2 \rightarrow [0, 1]$, define a voting procedure on $\mathbf{W}(t)$ as follows.

1. Each leaf, independently, votes 1 with probability $p(W_i(t))$ and otherwise votes 0.
2. At each branch point the vote of the parent particle i is the majority vote of the votes of its three children.

This defines an iterative voting procedure, which runs inwards from the leaves of $\mathbf{W}(t)$ to the root.

Define $\mathbb{V}_p(\mathbf{W}(t))$ to be the vote associated to the root.

Majority voting and the Allen-Cahn equation



$\mathbf{W}(t)$ = historical BBM, branching rate $\frac{1}{\varepsilon^2}$; $p : \mathbb{R}^2 \rightarrow [0, 1]$.

$$w(t, x) = \mathbb{P}_x^\varepsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1]$$

solves

$$\frac{\partial w}{\partial t} = \Delta w + \frac{1}{\varepsilon^2} w(1-w)(2w-1), \quad w(0, x) = p(x).$$

Probabilistic proof of Allen-Cahn result

Representation reduces result to

1. for x such that $d(x, t) \geq c\varepsilon |\log \varepsilon|$,
 $\mathbb{P}_x^\varepsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \geq 1 - \varepsilon^k$;
2. for x such that $d(x, t) \leq -c\varepsilon |\log \varepsilon|$, $\mathbb{P}_x^\varepsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \leq \varepsilon^k$.

Proof in two steps:

- ▶ a one-dimensional analogue in the case $p(x) = \mathbf{1}_{\{x \geq 0\}}$,
(symmetry, monotonicity for this p and amplification of bias through majority voting)
- ▶ for two-dimensional BM W and one-dimensional BM B , couple so that $d(W_s, t - s)$ is well approximated by B_s when W_s is close to Γ_{t-s} . (uses regularity assumptions on initial condition)

Approach parallels approximation of solution by one-dimensional standing wave in Chen (1992).

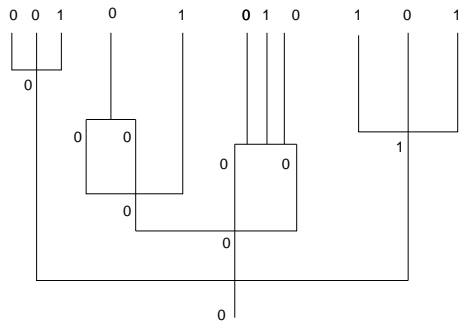
Proof for the 'stochastic' hybrid zones

Π is a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}^2 \times (0, \infty)$ with intensity measure $dt \otimes dx \otimes \mu(dr)$. Dual process of branching and coalescing 'ancestral' lineages. **Tracing backwards in time:**

- ▶ Start with a single individual;
- ▶ at event $(t, x, r) \in \Pi$, independently mark each lineage in $\mathcal{B}_r(x)$ with probability u ;
- ▶ **if at least one lineage is marked,**
 1. if event is neutral, all marked lineages coalesce into a single lineage, whose location is drawn uniformly at random from within $\mathcal{B}_r(x)$.
 2. if event is selective, all marked individuals are replaced by *three* offspring individuals, whose locations are drawn independently and uniformly from within $\mathcal{B}_r(x)$.
- ▶ In both cases, if no individual is marked, then nothing happens.

Majority voting procedure defined as before.

Recasting the result



1. for x such that $d(x, \sigma^2 t) \geq d\varepsilon_n |\log \varepsilon_n|$, we have $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \geq 1 - \varepsilon_n^k$.
2. for x such that $d(x, \sigma^2 t) \leq -d\varepsilon_n |\log \varepsilon_n|$, we have $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \leq \varepsilon_n^k$.

The choice of scaling

Recall

$$u_n = \frac{u}{n^{1-2\beta}}, \quad \text{and} \quad s_n = \frac{1}{\epsilon_n^2} \frac{1}{n^{2\beta}}.$$

- ▶ Each lineage is affected by selective events at rate η/ϵ_n^2 ;
- ▶ number of lineages bounded by that in ternary branching;
- ▶ $\epsilon_n^{-2} = o(\log n)$, so for any $\delta > 0$, $o(n^\delta)$ pairs of lineages;
- ▶ each such pair is in the region affected by some event (neutral or selective) at most $\mathcal{O}(n)$ times in $[0, T^*]$;
- ▶ chance that we see any coalescence events is $o(nu_n^2 n^\delta)$ for any $\delta > 0$.

Since $nu_n^2 = n^{4\beta-1}$, $\beta \in (0, 1/4)$, no coalescence events before time T^* .

Now use the proof in the Allen-Cahn case (and approximate the motion of ancestral lineages by BM).

A more general setting:

If homozygotes not equally fit

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{1}{2}\Delta u + su(1-u)(2au - 1) \\ &= \frac{1}{2}\Delta u + sau(1-u)(2u - 1) + s(a - 1)u(1 - u).\end{aligned}$$

Travelling wave (over faster timescale than curvature flow).

A more general setting:

If homozygotes not equally fit

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{1}{2}\Delta u + su(1-u)(2au - 1) \\ &= \frac{1}{2}\Delta u + sau(1-u)(2u - 1) + s(a - 1)u(1 - u).\end{aligned}$$

Travelling wave (over faster timescale than curvature flow).

