



HAL
open science

Couplage des dynamiques démographiques et évolutives hors équilibre

Guillaume Martin

► **To cite this version:**

Guillaume Martin. Couplage des dynamiques démographiques et évolutives hors équilibre. Génétique des populations [q-bio.PE]. Université de Montpellier, 2021. tel-03764533

HAL Id: tel-03764533

<https://hal.science/tel-03764533v1>

Submitted on 30 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HABILITATION A DIRIGER DES RECHERCHES

Soutenu le 16/11/2021 délivrée par l'Université de Montpellier

Couplage des dynamiques démographiques et évolutives hors équilibre

Coupling demographic and evolutionary dynamics out of equilibrium

GUILLAUME MARTIN

CRCN CNRS

INSTITUT DES SCIENCES DE L'EVOLUTION – MONTPELLIER

ISEM UMR 5554

Jury:

Sepideh MIRHAIMI (Rapporteuse)

Sylvain GANDON (Président du Jury)

Virginie RAVIGNE (Rapporteuse)

Daniel WEINREICH (Rapporteur)

Amandine WEBER (Membre)

Delphine SICARD (Membre)

Table des matières

I.	Abbreviations:	5
II.	Dossier en Français	5
	A. Curriculum Vitae	5
	B. Stages	6
	C. Financements et collaborations majeures	7
	D. Autres collaborations.....	10
	E. Enseignement	11
	F. Encadrement d'étudiants	12
	G. Travaux scientifiques et projet.....	17
	1. Développements théoriques :	20
	2. Développements expérimentaux :	22
	3. Des pistes à explorer pour les années à venir :	23
	4. Une remarque sur l'encadrement et l'avenir de nos étudiants... ..	24
	H. Liste exhaustive des publications.....	25
III.	Aim of my research.....	28
IV.	Rationale for the approaches used	30
	A. Minimal requirements of the models:.....	30
	1. Transient dynamic	30
	2. Evolution of demographic traits.....	30
	3. Theory for the distribution of mutation effects on FCs	31
	4. Theory for G x E on fitness components	33
	5. Handling multiple mutants with epistatic effects	35
	6. Handling stochastic events.....	36
	B. Minimal requirements of the experiments:.....	36
	1. Replicate trajectories.....	36
	2. Measuring distinct FCs	36
	3. Controlled, measured or inferred environmental conditions	37
V.	Empirical approach and projects	37
	A. Existing methods.....	37
	1. Colony forming units (CFUs) the "gold standard"	37
	2. Time kill-curves and pharmacodynamic functions	38
	3. General issues with CFU time series dynamics	40
	4. Special issues when studying killing factors	40

5.	Overcoming part of the logistic issue by the Start Growth Time (SGT) method	41
6.	Microfluidics	43
7.	Fluoro-luminometry	44
B.	Applying FLU	45
1.	Estimating various rates of the Birth Death process.....	45
2.	Calibrations	47
3.	Controls in the absence of stress	48
4.	Kill curves under Ampicillin stress.....	49
5.	ER under ampicillin stress.....	52
C.	Other Experimental projects.....	53
1.	Measuring fitness distributions over time	53
2.	Mutation accumulation experiment	54
3.	3D printed microplates to reproduce simple pharmacodynamics.....	54
VI.	Theoretical approach	55
A.	Context and state of the art	55
1.	Origin-fixation models.....	55
2.	Integro-differential models.....	57
B.	Common assumptions and definitions	60
1.	Notations	60
2.	Population and timescale.....	61
3.	Fitness.....	62
4.	Central assumptions	62
5.	Neutral or semi-neutral components characterizing a genotype.....	63
6.	Laplace transforms.....	64
C.	Properties of the ELT of genotype densities.....	64
1.	Expected total size	65
2.	ELT for the distribution of a particular FC and of fitness itself	65
3.	Density of the fittest genotype	66
4.	Linear operators.....	66
5.	Trajectory of mean fitness and mean FCs	67
D.	Effect of evolutionary forces on the system.....	68
1.	Demography and selection	68
2.	Mutation	70
3.	Migration	72
4.	Sexual reproduction.....	72

E.	Approximation methods to obtain explicit solutions	72
1.	Diffusion approximation ($U \gg \lambda$)	73
2.	Accurate approximation for the SSWM regime that I cannot justify	74
VII.	REFERENCES:	76

Je déclare avoir respecté, dans la conception et la rédaction de ce mémoire d'HDR, les valeurs et principes d'intégrité scientifique destinés à garantir le caractère honnête et scientifiquement rigoureux de tout travail de recherche, visés à l'article L.211-2 du Code de la recherche et énoncés par la Charte nationale de déontologie des métiers de la recherche et la Charte d'intégrité scientifique de l'Université de Montpellier. Je m'engage à les promouvoir dans le cadre de mes activités futures d'encadrement de recherche.

I. Abbreviations:

ER: evolutionary rescue

FGM: Fisher's geometrical model

FLU: fluoro-luminometry

LTEE: Long Term Evolution Experiment

EDP: Equations aux dérivées partielles

AB: antibiotic

CFU: colony forming unit

MIC: minimum inhibitory concentration

CFP: Cyan Fluorescent Protein

OD: optical density

DFE: distribution of fitness effects

FC: fitness component

MGF: moment generating function

ELT: expected Laplace transform

SSWM: strong mutation weak selection

WSSM: weak selection strong mutation

HC: house of cards

II. Dossier en Français

My university imposes me to write 10-20 pages in French for the HDR document if the rest is written in English. My opinion on this rule is quite negative but irrelevant, and it does allow me to use previous activity reports done for regular CNRS evaluation. I simply apologize to the non - French readers: "pardon my French". I have left the figures and their legends in English so that they may be of some use.

Cette partie résume mon dossier d'HDR, elle est relativement succincte j'espère. En effet, s'il me semble évidemment important d'évaluer mon travail passé, surtout mes co-encadrements, je pense plus intéressant, pour moi comme pour le jury, de se pencher sur les sujets que je compte aborder lors de futurs encadrements.

A. Curriculum Vitae

Né le 31/10/1977, pacsé, 2 enfants (10 ans et 5 ans).

Cursus scolaire :

1995: Lycée Balzac, Tours : Baccalauréat série S, option mathématiques (Mention Bien).

1995 – 1996: Lycée Descartes, Tours: Classes préparatoires MPSI (Mathématiques, Physique), admis en 2^{nde} année (MP*).

1996 – 1998: Lycée Clemenceau, Nantes: Classes préparatoires BCPST 1 et 2 (Biologie – Sciences de la Terre, Mathématiques, Physique-Chimie). Equivalence License de Biologie.

1998 – 2001: Agro Paris (INA-PG), Paris. Formation d'ingénieur en sciences de la vie. Double diplôme avec DEA.

2000 – 2001: Diplôme d'Etudes Approfondies (DEA = M2) : Biologie de l'Evolution et Ecologie, Université de Montpellier II, Agro Montpellier, rang 6/21.

Parcours professionnel :

2000 – 2005 : Thèse en Biologie de l'Evolution et Ecologie (bourse MENRT) sous la direction de Thomas Lenormand et Philippe Jarne (CEFE Montpellier), Titre: "Dynamique de l'adaptation: théorie et confrontation aux données expérimentales".

2006 – 2007: Post – Doctorat au Dpt d'Ecologie et Evolution de l'Université de Lausanne (Suisse), responsable : Jérôme Goudet. Génétique quantitative évolutive.

2007 – 2008: Post – Doctorat CNRS au laboratoire MIVEGEC (UMR 5290) puis au CEFE (UMR 5175), Montpellier, responsable: Sylvain Gandon. Epidémiologie évolutive, évolution virale.

2008 – 2012 : Chargé de recherche (CR2) au CNRS au laboratoire ISEM (UMR 5554), équipe Métapopulations, directrice : Isabelle Olivieri.

2012 – : Chargé de recherche (CR1 puis CRCN) au CNRS au laboratoire ISEM (UMR 5554), équipe Métapopulations, directrice : Sandrine Maurice puis Ophélie Ronce (depuis 2020).

B. Stages

La notion de « stage » n'est pas claire pour moi, j'ai donc décidé d'inclure tous les stages de formation réalisés avant de mon affectation actuelle, c'est donc l'occasion de décrire mon parcours jusqu'à mon arrivée à l'ISEM.

- 1999 (1^e année Agro, Langeac Haute – Loire): stage en élevage laitier (2 x 1 mois) sous la direction de Georges Pegon (exploitant). Je le mentionne car ce stage m'a fait réaliser à quel point des éléments d'écologie et d'évolution influençaient le fonctionnement d'une ferme (peuplement d'herbes et sa qualité fouragère, comportements en troupeau, pédologie, sélection des donneurs de spermes sur la base de mesures de génétique quantitative etc.)
- 2000 (2^e année Agro, INRA Moulon) : mini-projet (1 mois) de recherche sous la direction d'Isabelle Godringer sur la gestion dynamique des ressources génétiques du blé. J'ai pu y jouer avec matlab et les notions de Fst, découvrir ce qu'est un laboratoire de recherche et la génétique quantitative.
- 2000 (2^e année Agro, Univ. Lafayette, USA): Stage court (2 mois) sur la structure génétique de l'iris dans un bayou de Louisiane (allozymes) sous la direction de

Susan Mopper. J'ai pu y découvrir les joies d'un terrain exotique en hydroglisseur, les Fst encore, et les interactions de prédation structurant la diversité spatiale (aligator – cerf – iris).

- 2001 (DEA, GEMI devenu MIVEGEC Montpellier): stage court (15 jours) sur la manipulation de leur l'hôte arthropode par des nématomorphes sous la direction de Frédéric Thomas. J'y ai étudié un système fascinant et découvert la parasitologie évolutive.
- 2001 (DEA, CEFE Montpellier) : stage long (6 mois) sur l'évolution de la recombinaison en populations structurées sous l'effet de l'interaction dérive selection (Hill-Robertson effect), sous la direction de Thomas Lenormand. J'y ai découvert l'interaction dérive selection et sa modélisation analytique, joué avec Mathematica, et initié une collaboration qui dure encore aujourd'hui.
- Thèse (2001 – 2005, CEFE Montpellier): Etude de la distribution de l'effet des mutations et paysages adaptatifs, sous la direction de Thomas Lenormand. J'y ai initié des travaux qui constituent le cœur de mon projet actuel : paysages adaptatifs et leur test via l'évolution expérimentale.
- 2006 (Univ. Lausanne) : Postdoctorat FNS (1 an) en génétique quantitative avec Jérôme Goudet. J'y ai travaillé sur les modèles d'évolution multivariée et découvert les matrices aléatoires (et R un peu).
- 2007 (MIVEGEC puis CEFE Montpellier): Postdoctorat CNRS (1 an) en épidémiologie évolutive avec Sylvain Gandon. J'y ai initié une autre partie de mon projet actuel : le lien entre évolution, démographie et paysages adaptatifs à plusieurs pics (composantes de fitness), et découvert les virus...

C. Financements et collaborations majeures

J'ai eu la chance (i) de ne pas dépendre trop des appels d'offres (faire des modèles et élever des bactéries ça ne coûte pas trop cher), et (ii) de profiter de divers financements, grâce à l'aide d'autres, comme c'est souvent le cas je suppose. Plutôt qu'une liste d'acronymes, je profite donc de l'occasion pour décrire certaines collaborations fortes impliquées dans ces projets, dont ils sont la conséquence souvent, et non l'inverse, n'en déplaise à l'ANR. C'est aussi tout de même l'occasion de retracer l'apport de ces projets financés dans mon parcours de recherche depuis mon recrutement.

Tout d'abord, ma principale source de financement (celle qui a permis le gros de mon travail) est le salaire que le CNRS me verse, et l'ensemble des conditions et libertés de travail que m'offrent mes tutelles (CNRS, Université de Montpellier, IRD), notamment

via le salaire qu'elles versent à tous ceux qui m'aident. Il me semblait important de le rappeler : même si c'est évident, je crains que ça ne le reste pas à l'avenir.

Les principaux projets financés dont j'ai bénéficié sont décrits ci-dessous, par ordre chronologique en 4 grandes phases, avec leur apport, notamment en termes d'encadrement. Sauf mention d'un porteur j'étais le porteur du projet.

1/ Premiers financements et mise en place de mon système expérimental :

- Financement Université Montpellier II (2008, 16 k€) : Appel à Projet du conseil scientifique pour l'installation de nouveaux entrants sur le site.
- PHC Picasso avec l'Espagne du MAE (2008-2009, 21k€ de missions sur 2 ans) : avec S. Elena (CSIC, Valencia) et R. Sanjuan (U. Valencia)
- PEPH CNRS joint INSB-INSMI-INEE (2010, 21 k€): partenaires A. Lambert (U. Paris VI), L. Prezeau (CNRS, IGF Montpellier), F. Campillo (INRIA Montpellier)
- RTRA 'BIOPHYS' de la fondation Agropolis (2009-2014, budget total oublié, budget reçu 12k€) : porteur A. Estoup (INRA CBGP Montpellier)

Ces financements m'ont permis d'initier des collaborations (modélisation et expérimentation) et de mettre en place mon système expérimental. J'ai aussi été chaleureusement accueilli par l'équipe voisine de la mienne (Nicolas Mouquet, Mike Hochberg, Oliver Kaltz, Claire Gougeat-Barbera), qui disposait déjà d'un laboratoire de bactériologie. J'ai ensuite pu m'équiper d'appareils de mesure pour la fluoro-luminométrie ('**FLU**') : lecteur multimode. J'ai aussi pu financer les gratifications des premiers étudiants (M1 et L3) que j'ai encadrés, sur ce système expérimental, pour explorer la méthodologie.

2/ Participation forte à deux ANRs sur l'interaction démographie - évolution :

- ANR 'EVORANGE' (2009-2013, 534k€): partenaire (40%), porteuse O. Ronce (CNRS ISEM, Montpellier)
- ANR 'SILENTADAPT' (2013-2017, 359k€): partenaire (30%), porteur T. Lenormand (CNRS CEFE Montpellier)

Ces projets m'ont permis de développer la modélisation comme l'expérimentation sur l'interaction evolution/démographie face à un stress environnemental : c'était le thème central de ces deux ANRs.

Avec Thomas Lenormand (SILENTADAPT), nous avons prolongé (MARTIN AND LENORMAND 2015b) le développement de modèles de mutation entre environnements et leur test par evolution expérimentale initié pendant ma thèse (MARTIN AND LENORMAND 2006a; MARTIN AND LENORMAND 2006b). Cela a donné lieu à mon premier co-encadrement de thèse (Noémie Harmand) avec lui, dont les expérimentations étaient financées par cette ANR (HARMAND *et al.* 2017; HARMAND *et al.* 2018).

Avec Ophélie Ronce (EVORANGE) nous avons initié la modélisation de la rescousse évolutive (Evolutionary Rescue '**ER**') dans un contexte multi-allèles en collaboration avec un postdoctorant (Robin Aguilée) recruté sur cette ANR (MARTIN *et al.* 2013). Ces développements ont mené directement à la deuxième thèse que j'ai co-encadrée, avec Ophélie cette fois (Yoann Anciaux) : les modèles d'ER sur un paysage adaptatif

(Fisher's Geometrical Model '**FGM**') qui résultent de cette thèse (ANCIAUX *et al.* 2018; ANCIAUX *et al.* 2019) sont au cœur de mon projet actuel.

Ces deux projets m'ont aussi permis de prolonger divers tests préliminaires de la méthode de FLU, avec l'encadrement de plusieurs stagiaires rémunérés (M1) et le recrutement d'un assistant ingénieur pour ces développements.

3/ Projets à forte interaction avec les mathématiques :

- Financement inter-Labex CEMEB-NUMEV-AGRO (2018, Salaire postdoc 18 mois) : co-porteurs M. Alfaro (U. Montpellier) & B. Cloez (INRA Supagro Montpellier)
- ANR 'RESISTE' (janvier 2019 - 48 mois, 396k€) : co-porteur L. Roques (INRA BioSP Avignon) : page web : <https://informatique-mia.inrae.fr/resiste/>

Ces projets m'ont permis de collaborer de façon régulière avec plusieurs collègues mathématiciens, et de participer à l'encadrement d'étudiants mathématiciens (Alvaro Gonzales Mateos, postdoc inter-Labex) et Florian Patout (postdoc RESISTE).

Le projet RESISTE est le fruit de ma principale collaboration actuelle (initiée en 2013) avec Lionel Roques, et a un équilibre ~50/50 entre mathématiques et expérimental (il a d'ailleurs été obtenu via La CE 45 de l'ANR (interaction mathématiques/biologie). Cette interaction avec des mathématiciens me permet de développer des modèles plus complexes, et d'améliorer les méthodes statistiques pour leur test expérimental.

Cette collaboration avec Lionel (et François Hamel) m'a aussi permis de participer à l'encadrement de deux autres doctorants, mathématiciens cette fois : Florian Lavigne (co-encadrement officiel) et Marie – Eve Gil (collaborations sur ses 2 articles de thèse (GIL *et al.* 2017; GIL *et al.* 2019)).

4/ projets de FLU 'haut-débit' :

- Projet exploratoire Labex CEMEB 'BACTPHY' 2018-2021, 23k€) : co-porteur avec R. Froissart (CNRS MIVEGEC, Montpellier)

Ce projet et le volet expérimental de RESISTE ont permis d'équiper le laboratoire pour la gestion du plus haut débit (compteur automatique de colonies, incubateurs de microplaques, distributeur pour microplaques).

Enfin ces deux projets m'ont permis d'initier une collaboration régulière avec un spécialiste de l'évolution expérimentale microbienne (phages et bactéries) en la personne de Remy Froissart (MIVEGEC Montpellier). Il implique aussi le recrutement d'un postdoctorant dédié à la production de données expérimentales par FLU. Je suis maintenant quotidiennement en interaction avec Remy pour tous les volets expérimentaux que je mets en place. Cela est d'une grande aide pour penser, réaliser et analyser les expériences, mais aussi pour former/encadrer les étudiants au laboratoire (en présentiel donc !).

Au final ces collaborations, notamment avec Lionel et Remy, me permettent d'envisager de continuer à ménager la chèvre et le chou (expériences et

modélisation), grâce à l'aide régulière d'un spécialiste de chaque (je ne préciserai pas qui s'occupe de la chèvre et qui du chou).

D. Autres collaborations

Les principales collaborations que j'ai actuellement sont décrites ci-dessus mais il en existe d'autres, moins régulières et plus ou moins récentes que j'espère voir s'enrichir avec le temps.

Dans mon UMR : Oliver Kaltz et Emmanuel Fronhoeffler sont partenaires de mon ANR RESISTE et nous devons développer des méthodes de mesures de la prédation sur les bactéries *E.coli* par FLU. Leurs modèles d'études étant des prédateurs du mien, cette collaboration est assez évidente et me donne l'occasion d'étudier un autre stress biotique dans mes expériences d'ER. Ils apportent aussi leur expertise en écologie expérimentale et théorique. Nous avons récemment initié des discussions (et répondu à un appel d'offre local) avec un spécialiste de la phagocytose (Guillaume Charrière, U. Montpellier, microbiologiste) et j'espère que cela nous mènera à une analyse plus fine des mécanismes en jeu dans nos microplaques.

Je suis aussi impliqué (5%) dans l'ANR toute neuve d'Alison Duncan sur l'analyse de ses expériences d'évolution d'acariens sur différents hôtes végétaux, de 'vrais' animaux, sexués et multicellulaires.

Je collabore avec Samira Sarter (CIRAD) sur le développement de la FLU pour étudier l'effet d'huiles essentielles comme alternatives aux antibiotiques. Nous avons fait de premiers essais intéressants sur coli, mais il faut maintenant réussir à transformer ses bactéries (pathogène de poisson, non modèle) pour aller plus loin en termes d'applications (suivi *in vivo* etc.).

Enfin, j'ai tout récemment entrepris de tester mes modèles d'adaptation sur un paysage adaptatif sur les trajectoires non seulement de fitness mais aussi de séquences génomiques. Deux collègues de l'ISEM, évolutionnistes moléculaires, ont accepté de m'aider en cela (Nicolas Bierne et Pierre - Alexandre Gagnaire).

Hors de mon UMR : Je vais reprendre mes collaborations avec Sylvain Gandon (évolution et épidémiologie virale) et Luis Miguel Chevin (évolution de la plasticité, évolution expérimentale des algues unicellulaires), tous deux partenaires de mon ANR et chercheurs au CEFÉ (Montpellier). Nous avons plusieurs tâches en cours avant que celle-ci ne se termine mais n'avons rien commencé encore. Par ailleurs, le recrutement récent de Rafael Forien dans l'équipe de Lionel Roques (INRA BioSP Avignon) a apporté une expertise en processus stochastique à notre collaboration, qui je pense va s'avérer cruciale à l'avenir.

International : Phil Gerrish (U. New Mexico, USA) est venu en visite pendant 2 mois en 2015, il est partenaire de mon ANR, et spécialiste de l'évolution des asexués et développe également des approches basées sur la dynamique de fonctions génératrices pour ces dynamiques. J'ai collaboré avec Matthew Osmond (U. Toronto, Canada) pour des modèles d'ER en plus d'un pas mutationnel durant sa thèse (OSMOND *et al.* 2020). Il est maintenant en poste et nous comptons étendre ces

modèles à divers situations (densité dépendance, recombinaison). Hildegard Uecker (MPI Plön, Allemagne) est spécialiste de l'ER, après plusieurs visites mutuelles nous venons de commencer un projet de modélisation de l'ER dans mes expériences avec des phages. Enfin Helen Alexander (U. Edinburgh), avec qui j'ai déjà collaboré sur l'ER (ALEXANDER *et al.* 2014), nous aide maintenant dans l'analyse de modèles de pharmacodynamique pour mes expériences d'ER en présence d'antibiotiques. Enfin, nous avons initié des discussion avec Robert Noble (U. London, UK) qui travaille sur les modèle d'evolution de la résistance aux chimiothérapies des cancers, il m'a d'ailleurs convié à une conférence récente sur le thème (en visio). L'interaction est d'autant plus intéressante qu'il semble y avoir une 'culture' différente, dans le monde de la cancérologie, de celle des microbiologistes, sur l'evolution de la résistance aux traitements. Du point de vue de la seule génétique des populations, il s'agit par contre de cellules asexuées évoluant par naissance et mort dans les deux cas.

Plus généralement, je pense avoir créé des liens avec la communauté internationale travaillant en évolution expérimentale (depuis ma thèse) et celle travaillant en théorie, sur la rescousse évolutive et/ou les paysages adaptatifs (FGM et autres), et plus récemment avec une partie de la communauté (Française) des mathématiciens intéressés par les modèles d'évolution. Il me reste je pense à trouver des collaborateurs plus spécialisés en microbiologie, pour m'aider dans l'analyse de ce qu'il se passe dans mes expériences.

E. Enseignement

Je n'ai vu nulle mention du travail d'enseignement dans les directives officielles pour rédiger le rapport d'HDR. Cela me semble fort étonnant (pour rester diplomate), alors qu'on nous demande de lister nos brevets éventuels : je n'en ai pas, donc je vais parler de mes enseignements.

Je n'ai que peu enseigné durant ma thèse (pas de monitorat) mais j'ai pu participer activement aux enseignements de statistiques et de génétique quantitative durant mon postdoc à Lausanne grâce à Jérôme Goudet (2006).

Je donne chaque année quelques heures de cours sur l'évolution expérimentale et la modélisation de la selection - mutation - dérive, pour le parcours Ecologie – Evolution de l'UM (M1 et M2). Par ailleurs, profitant de l'opportunité offerte par ma collègue d'équipe Sandrine Maurice, j'ai pu me frotter à la conduite d'une UE entière, de modélisation (30h) au niveau M1 dans le master international Erasmus Mundus en Evolution (master MEME). J'ai assuré cet enseignement durant 6 ans, avec divers projets tutorés en petits groupes etc. C'était intéressant mais difficile au niveau pédagogique, pour gérer les niveaux très hétérogènes en mathématiques que j'avais en classe. J'ai laissé la main cette année à plus expérimentée que moi (Sandrine en l'occurrence). Une autre collègue EC (Céline Devaux) m'a demandé de m'impliquer l'an prochain dans l'enseignement de modélisation de l'UM au niveau License avec elle. Cela me semble plus pertinent pour moi de par la présence d'EC expérimentés pour me guider, et d'étudiants plus homogènes car tous issus du parcours scolaire

Français, *a priori*. Le fait de commencer la modélisation dès la License de biologie me semble intéressant, lorsque les étudiants n'ont pas encore oublié les choses qu'ils ont apprises au lycée, en physique par ex.

Quoiqu'il en soit je pense tout de même que cette expérience d'enseignement m'a appris sur la façon d'aborder la modélisation avec les étudiants biologistes, selon leur profil par rapport à l'outils mathématiques : certains l'aiment (c'était mon cas), d'autres le craignent (voire le haïssent carrément) et il faut alors trouver des moyens de dépasser cette peur, pour qu'ils le voient comme un outil parmi d'autres de la biologie moderne.

La simulation numérique est évidemment une option fort prisée, mais je pense qu'elle permet parfois à l'étudiant de ne pas vraiment faire de modélisation. J'entends par là de se permettre de ne pas prendre le risque de négliger des choses etc. pour pouvoir avancer (on peut simuler un problème très compliqué relativement vite). Or c'est dans ce choix que je crois se situe la modélisation, car une simulation d'un système très complexe ne pourra jamais être explorée correctement pour donner des conclusions générales, l'espace des paramètres explose. Dans ce sens, les logiciels de calcul formel (Mathematica, Maple, Maxima etc.) sont un moyen d'entrée je crois. J'ai pu en voir les effets positifs parfois sur des étudiants de cette UE (et aussi dans ma pratique de modélisation personnelle). Le logiciel permet de shunter partiellement le calcul (ne pas être bloqué face à l'équation), pour aller plus directement à la modélisation, qui se situe avant et après le calcul mathématique lui-même, à mon avis. Bien entendu, il est toujours mieux de comprendre ce que fait un logiciel, mathématiquement, et aussi de savoir si on a même le 'droit' de poser une équation sous une forme donnée, ce que seules les mathématiques permettent. Mais il vaut mieux avancer avec un logiciel, même mal compris, qu'esquiver toute approche formelle par simple peur que ce soit trop compliqué.

F. Encadrement d'étudiants

De 2008 à 2013 je me suis borné à encadrer des étudiants de niveau M1 et L pour réaliser des tests préliminaires sur mon système expérimental de FLU. J'ai continué à le faire ensuite : à ce jour j'ai encadré 4 étudiants en écologie / évolution (3 M1 + 1 L3, U. Montpellier), 8 étudiants L3 en microbiologie en stages d'été de 2 mois (groupe de 5 puis groupe de 2, U. Montpellier), un étudiant de M1 médecine (stage de découverte de la recherche de 2 mois, U. Rennes), un groupe de 4 étudiants en TIPE de BCPST (niveau L1, Lycée Joffre).

Ces stages étaient l'occasion d'avancer tant que la technique n'était pas mûre pour donner lieu à des publications potentielles ou à un rapport de M2 cohérent. Ils ont aussi été l'occasion de former des étudiants aux notions de génétique des populations et démographie impliquées dans ces expériences : naissance et mort, processus de mutation – sélection – dérive. Je pense que le type de données cinétiques obtenues par cette technique (cinétique de vivants et de morts à long terme, haute réplication)

permet d'aborder de façon simple, et sur des données réelles, ces notions : il y a donc une utilité pédagogique en soit.

J'ai co-encadré trois doctorants au total (voir ci-dessus), je détaille rapidement chaque thèse, avec une ou deux figures tirées d'un de leurs articles à chaque fois pour illustrer.

Noémie Harmand (avec T. Lenormand, voir plus haut) a travaillé le test de modèles de paysages pour l'adaptation de bactéries à des doses variables d'antibiotiques. Sa thèse nous a permis d'établir les bases d'une vision 'paysage' (FGM) pour conceptualiser l'effet de la dose antibiotique. On a donc pu relier ces problèmes de résistance antibiotique aux concepts classiques de l'écologie évolutive : la sélection se fait pour un phénotype continu sous-jacent, avec un optimum qui dépend de l'environnement (la dose), la mutation crée une variation symétrique du phénotype, dont l'effet sur la fitness est obtenu via la fonction reliant fitness et phénotype, selon l'environnement. Je reproduis ci-dessous une figure de son premier article (HARMAND *et al.* 2017), qui illustre la façon dont nous avons tenté de conceptualiser l'effet de la dose sur ces paysages.

Figure 1 :

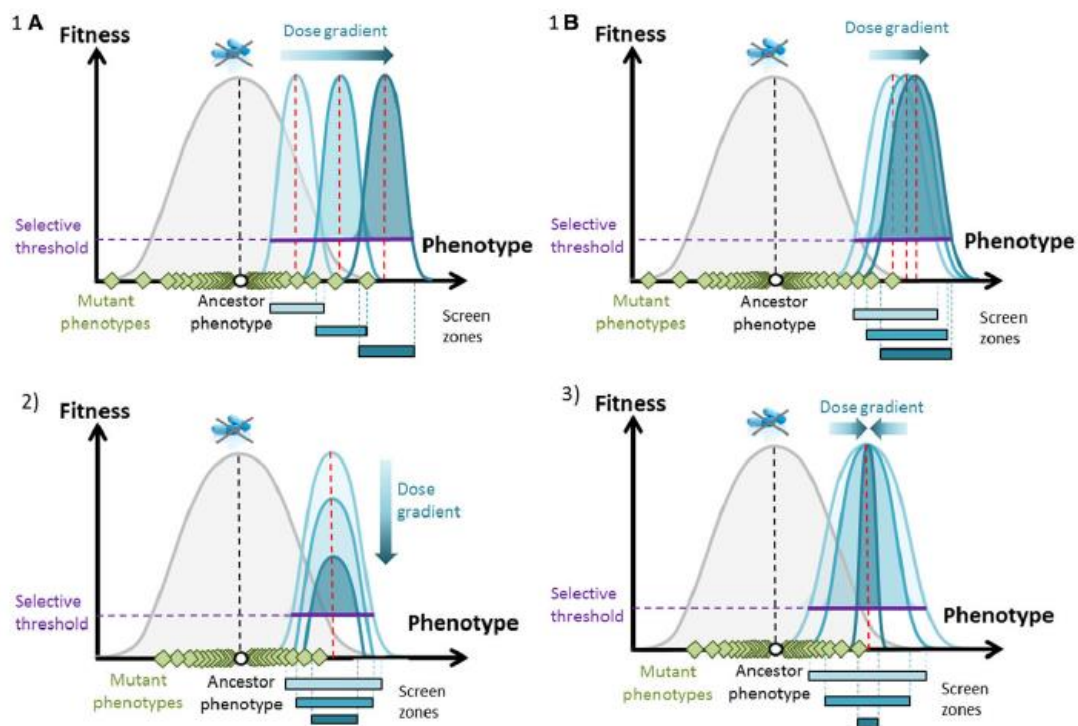


Figure 1. Scenarios of antibiotic dose effects on simple fitness landscapes. The ancestor phenotype (white dot) is optimal in the environment without antibiotic (fitness function given by the gray curve). Increasing antibiotic doses generate different fitness functions, indicated by increasingly dark blue curves. Screen dose may affect 1A–B the optimum, (2) the maximum, or (3) the curvature of the fitness function. The horizontal segment shows the threshold of positive growth. Phenotypic effects of random mutations (diamonds) are normally distributed around the ancestor phenotype. Resistant phenotypes can be screened within the phenotypic interval delimited by the growth threshold: the “screen zones”, indicated by horizontal segments below the x-axis, with the corresponding color of each dose.

Un apport central de sa thèse, à mon sens, a été de montrer que, sur le long terme, l'évolution à divers doses d'antibiotiques conduit à de l'adaptation locale pour

chaque dose. Or, initialement, les mutants criblés sur doses fortes sont plus fits que ceux issus de toute dose plus basse, et ce à toutes les doses. Cette observation classique après criblage amène je crois beaucoup de microbiologistes à avoir une vision assez différente de celles des écologistes/évolutionnistes (« saturés » de notions de niches et d'optimums). Ils supposent souvent que certains mutants résistants sont simplement « meilleurs » dans toutes les doses, et qu'il n'y a pas d'optimum dose-dépendant. Ci-dessous la figure qui illustre cette adaptation locale à long terme (HARMAND *et al.* 2018) : dans chaque dose de mesure (« measure dose ») la lignée la plus fit est celle qui a évolué à cette dose (« evolution dose » en abscisse). Après le criblage seul (avant évolution ultérieure) ce sont les mutants sélectionnés à la dose maximale (200, points bleus) qui étaient les plus fits partout.

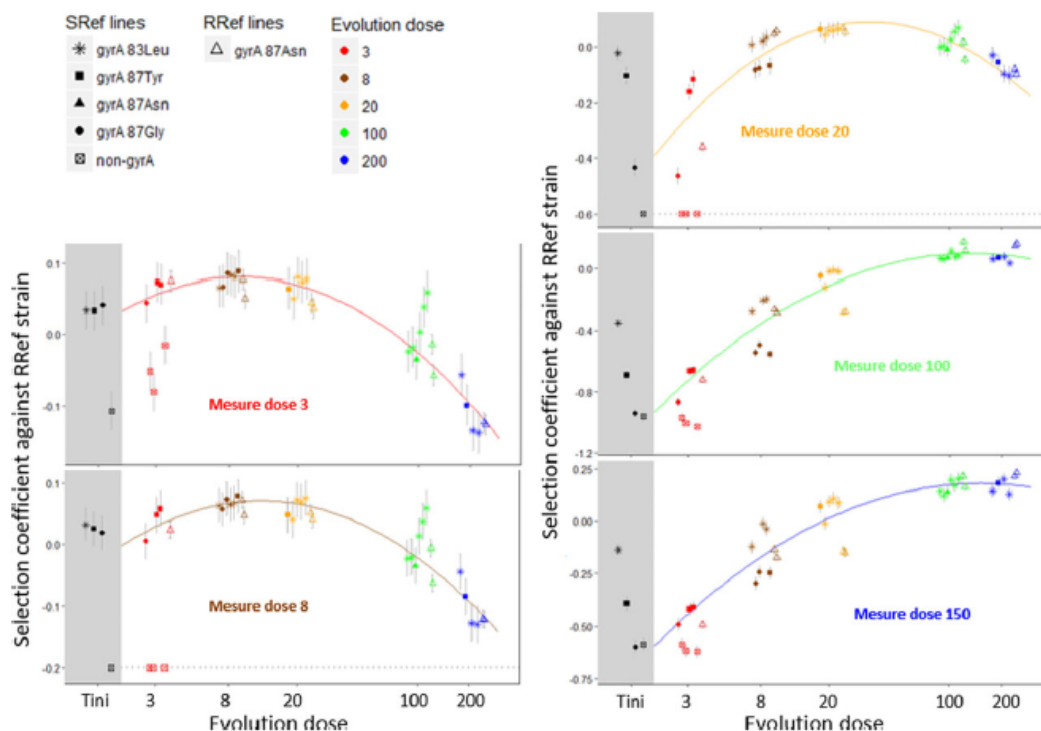


Figure 2. Fitness profiles of the SRef and RRef lines across evolution doses after the screen (within the grey area, only mutants screened in one step are shown) and after 400 generations of experimental evolution (within the white area). The y-axis indicates the fitness difference between SRef and RRef lines and the fitness of the RRef strain used as the competitor in all comparisons. Competitions were performed at different measure doses. Each panel corresponds to one measure dose. Mean selection coefficients across replicates are indicated with standard errors. Fitness measures were made for lines evolved at different doses (x-axis). Values of selection coefficients are not comparable across measure doses (because the fitness of the RRef strain varies among measure doses) but fitness ranks are meaningful. For better readability, points are slightly spread on the x-axis at each dose. Each dot corresponds to a different evolved line, with the color corresponding to the evolution dose of that line (see legend) and the symbol to the identity of the initial mutation (different *gyrA* resistance mutation or non-*gyrA* mutations are indicated in the figure legend). On each panel, a curve of the corresponding measure dose color shows the fitted parabolic relationship between the log-evolution dose and the selection coefficients of the *gyrA* lines. The dots placed on the horizontal gray dotted lines at MD8 and MD20 where set artificially at this threshold value for the readability of the figure, as they have selection coefficients that were measured to be largely lower than this threshold. Representing their exact value would expand the range of values on y-axis, which would significantly reduce readability for the bulk of the data.

Je dirais que pour cette thèse j'ai participé à 20-30% de l'encadrement, Noémie était basée au CEFE.

Yoann Anciaux (avec O. Ronce voir plus haut) a travaillé à relier les modèles d'effet des mutations dans les paysages adaptatifs (FGM) à la probabilité de rescousse

évolutive (ER), lorsque l'environnement change de façon abrupte. Le challenge était que le FGM, malgré sa simplicité, génère un patron complexe de distribution d'effets des mutations sur la fitness, selon l'environnement. Il fallait donc dériver les probabilités d'ER dans un cadre où une variété de taux démographiques se côtoient et sont produits stochastiquement par mutation, avec une distribution qui dépend du stress imposé (lequel affecte aussi la démographie donc la production de mutants eux-mêmes). Une figure de son premier article (ANCIAX *et al.* 2018) illustre comment FGM et ER sont reliés dans le cas de selection forte et mutation faible :

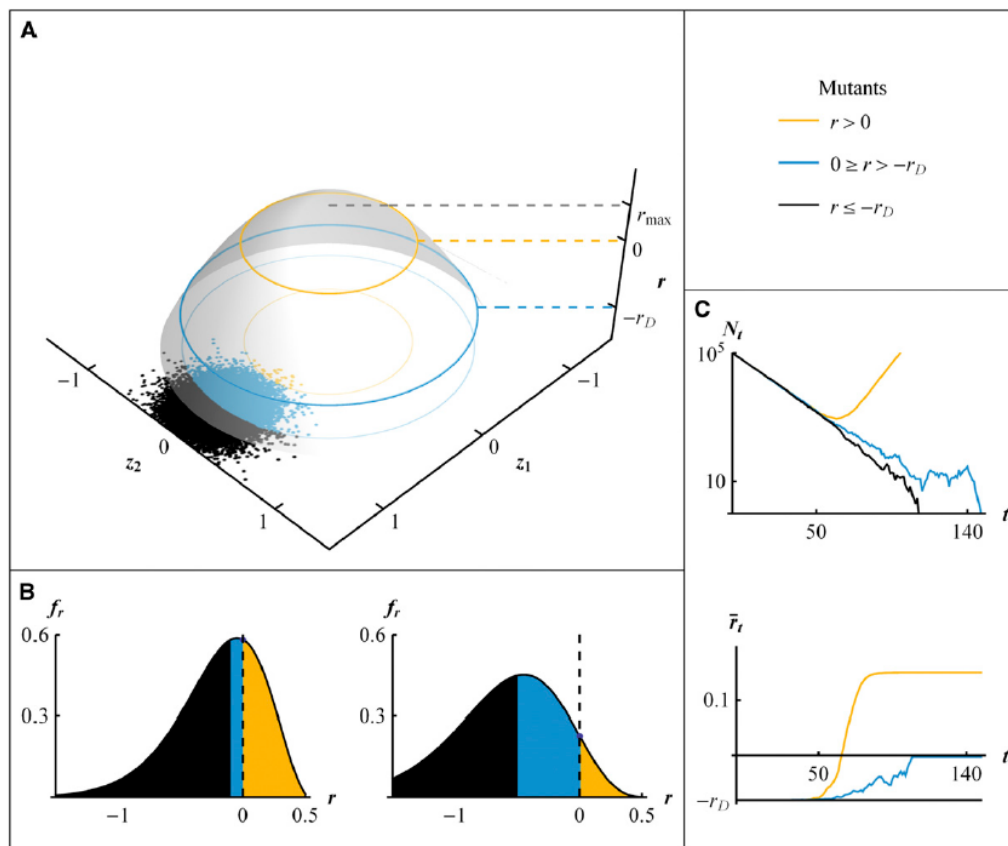


Figure 1 Evolutionary rescue in Fisher's geometric model (FGM). In all panels, black refers to deleterious and neutral mutations ($r_D \geq r$), blue to beneficial but not resistant mutations ($-r_D < r < 0$), and orange to resistant mutations ($r > 0$), around the dominant genotype of the ancestral population with phenotype $\mathbf{z}_A \neq \mathbf{0}$. (A) Fitness landscape (FGM) with growth rate r (z-axis) determined by two phenotypic traits z_1 and z_2 . Dots represent the distribution of random mutant phenotypes around the dominant genotype of the ancestral population. The growth rate of this dominant genotype, in the stressful environment, is $-r_D$, and r_{max} is the maximal fitness at the phenotypic optimum. (B) Distribution of growth rates among random mutants arising from the dominant genotype (distribution of mutation effects on fitness) for two decay rates $r_D = 0.1$ (left) and $r_D = 0.5$ (right). (C) Dynamics of the population size N_t and mean fitness \bar{r}_t of a population starting from a clone at $-r_D = -0.083$ at size $N_0 = 10^5$. The black line represents the case without fixation of a beneficial mutation, the blue line the case with extinction in spite of the fixation of a beneficial, but nonresistant, mutation, and the orange line the case of a rescue. Parameters for the simulations are $r_{max} = 1.5$, $U = 2 \times 10^{-5}$, $n = 4$ and $\lambda = 5 \times 10^{-3}$.

La génétique des populations sous-jacente à ce processus est complexe car des mutants divers sont constamment générés avant de s'éteindre stochastiquement ou d'envahir. Nous avons pu identifier deux régimes extrêmes où un traitement analytique assez simple était possible, ce qui a donné lieu à deux articles :

- un régime très peu polymorphe où l'ER est causée par un mutant résistant (unique, mais variable entre répliquas), c'est le régime Strong Selection Weak Mutation (**SSWM**)
- Un régime fortement polymorphe où la dynamique évolutive peut être approximée de façon déterministe et où seule la stochasticité démographique

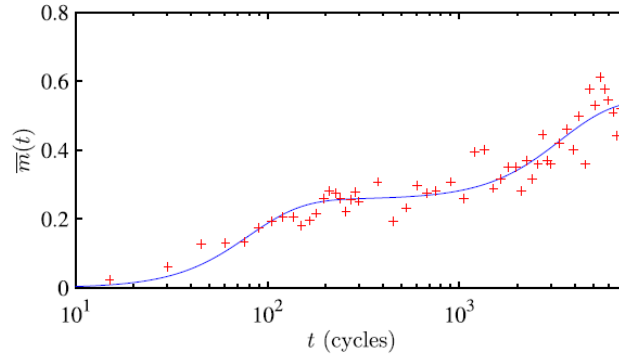
induit une probabilité d'extinction non nulle. C'est le regime Weak Selection Strong Mutation (**WSSM**).

Je dirais que la principale contribution de la thèse de Yoann a donc été de dériver et valider ces deux régimes et d'en ressortir des prédictions testables en terme de relation entre force d'un stress (taux de décroissance induit initialement) et probabilité d'ER. Ce test est l'objet central de mes expériences d'ER par FLU actuelles, et c'est Yoann qui s'en charge puisque je viens de le recruter (Février 2021) pour le postdoc expérimental de RESISTE après une expérience à l'étranger de 2 ans (Thomas Bataillon, U. Aarhus).

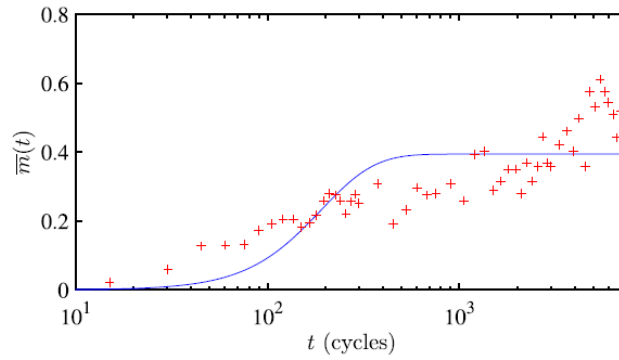
Je dirais que pour cette thèse, ma part dans l'encadrement était plus proche de 70 - 80%, puisque j'ai écrit le sujet, encadré Yoann tout du long et qu'il était basé à l'ISEM.

Florian Lavigne (avec L. Roques et F. Hamel, voir ci-dessus), pour la part (majoritaire) de sa thèse où j'étais impliqué, a analysé mathématiquement des modèles qui étendent le FGM simple : une source et un puit connectés (deux environnements et migration), un modèle anisotrope où les directions phénotypiques sont non – équivalentes. Son premier article (LAVIGNE *et al.* 2020) a permis d'étendre l'interaction démographie / évolution en incluant un processus de migration récurrente depuis un premier environnement (stable démographiquement où la population est adaptée) vers un environnement puits où la population décroît en moyenne en arrivant. En plus d'étudier les propriétés mathématiques du modèle (pas moi !), nous avons dérivé des expressions explicites pour le temps d'attente (moyen) avant que le puit ne soit envahi.

Son deuxième travail principal publié a consisté à étudier la trajectoire de la fitness moyenne sous selection – mutation – dérive (négligée) sous un FGM anisotrope. L'anisotropie entraîne que plusieurs plateaux de fitness successifs peuvent être atteints, avec l'adaptation commençant le long de la direction de variance mutationnelle maximale etc. Nous avons pu montrer qu'un tel patron était cohérent avec les données d'adaptation à long terme de Lenski (Long Term Evolution Experiment **LTEE**), qui est la plus longue évolution expérimentale jamais entreprise à ce jour. Or, la dynamique observée, après avoir montré un plateau de fitness après 5000 generations (suggérant un optimum atteint) a de fait continué à monter à plus long terme (suggérant l'absence d'optimum). Nous avons illustré comment un modèle anisotrope avec optimum était aussi cohérent avec un tel patron. Le figure ci-dessous de son article (HAMEL *et al.* 2020) illustre cet ajustements aux données du LTEE :



(a) Anisotropic model



(b) Isotropic model

Fig. 2. Trajectory of adaptation, anisotropic and isotropic model versus LTEE data. The function $f(t) = \bar{m}(t+2) - \bar{m}(2)$, with $\bar{m}(t)$ given by (34) was fitted to the LTEE data. The functions $f(t)$ with the parameter values corresponding to the best fit are depicted in blue. The red crosses correspond to the LTEE data (population Ara-1, $\ln[w(6.64t)]$, with w the original data). The values leading to the best fit are (a, anisotropic model) $\mathbf{x}_0 = (0.73, 0.76)$, $\mu_1 = 1.3 \cdot 10^{-2}$ and $\mu_2 = 3.0 \cdot 10^{-4}$; (b, isotropic model) $\mathbf{x}_0 = (0.44, 0.78)$ and $\mu_1 = \mu_2 = 5.3 \cdot 10^{-3}$.

Je dirais que le principal apport de la thèse de Florian (du point de vue modélisation, et non des mathématiques, hors de mon expertise), a été de produire un ensemble d'expressions explicites et testables pour des extensions non triviales des dynamiques que nous avons dérivées avec Lionel (MARTIN AND ROQUES 2016). Ma contribution à son encadrement est sûrement autour de 10 – 15 % puisqu'il était basé à Avignon et déjà encadré par deux mathématiciens expérimentés. Mon apport a plus été dans le fait de proposer des problèmes biologiques et l'approche pour les modéliser ou les tester, que dans la production des résultats eux-mêmes (hormis l'écriture bien entendu).

A noter que si je n'ai pas réussi à faire publier Yoann et Noémie ensemble (pour des raisons de décalage de timing et non de thème ou d'entente mutuelle), Florian et Yoann ont une publication en commun.

G. Travaux scientifiques et projet

Mon activité de recherche s'est principalement orientée sur une thématique centrale de mon projet de recherche CNRS initial (et de mon rapport à vague 2014) : **la dynamique de l'adaptation à des environnements changeants, hors équilibre, et ses conséquences démographiques (croissance, extinction, éventuellement épidémiologie)**. J'étudie ces dynamiques de façon théorique (surtout via les

paysages phénotype-fitness), et empiriques via l'observation en direct de cinétiques bactériennes *in vitro*.

Ce projet est pensé comme un duo expérimental et théorique, dans le but de proposer une théorie idéalement générale (entre espèces, entre environnements etc.), et testable quantitativement, idéalement validée sur des données produites par moi et mes collaborateurs, ou en utilisant des données existantes. La partie théorique nécessaire a bien progressé durant ces 5 dernières années. La partie expérimentale, jusqu'à récemment, ne reposait que sur mon travail et celui de stagiaires M1 et L de courtes durée et n'a donc pas avancé aussi vite. Elle n'a pas été publiée et c'est le but du postdoctorat de Yoann Anciaux commencé en Février 2021. La technique que j'utilise (FLU) est toutefois maintenant fonctionnelle et robuste, dans mon laboratoire. Je dispose donc maintenant de prédictions testables sur des mesures expérimentales que je sais comment produire, ou que nous avons déjà produites en assez grand nombre. Reste à réaliser d'autres expériences et surtout à tester les prédictions, les améliorer etc.

Ce projet a des visées fondamentales comme appliquées. En effet l'étude quantitative des processus de sélection naturelle, mutation dérive et migration permet de tester les théories qui sont à la base de notre discipline et que nous enseignons quotidiennement. Pour autant, c'est mon sentiment (en tant que chercheur et par mon expérience d'enseignement) que nous ne disposons pas de beaucoup de démonstrations expérimentales directes de ces théories. Alors que l'idée d'une terre plate remet autant en cause certaines croyances que la sélection naturelle, cette dernière est beaucoup plus attaquée. Il me semble qu'une des raisons est que les équations de la physique sont illustrées voire, démontrées quantitativement en classe, et appliquées quotidiennement, alors que celles de la biologie des populations (et surtout de l'évolution) ne le sont pas. Je ne dis pas que nous ne possédons pas des illustrations expérimentales de la sélection naturelle, que nous enseignons d'ailleurs. Mais celles-ci sont d'abord qualitatives : on documente des relations structure fonction à foison à toutes les échelles (de la molécule au groupe social), mais cela ne montre pas que les équations de la sélection, mutation, recombinaison que nous enseignons sont valides. On documente l'histoire d'une adaptation dans le registre fossile ou dans les génomes, *a posteriori*, mais nous aurions été bien en mal de les prédire à l'avance.

En mécanique, on peut illustrer les lois de Newton pendant un TP de lycée (on mesure la position d'une bille qui tombe au cours du temps par ex.). En écologie, nous disposons de quelques modèles démographiques qui peuvent être ajustés sur des dynamiques observées de taille de population ou structures d'âge (même en nature parfois). Toutefois, les paramètres (moyens) de ces dynamiques changent avec le contexte (environnement biotique et abiotique) et même avec le temps, de façon autonome (plasticité, mutation, recombinaison, migration). Nous ne savons pas où va tomber la bille, à quelle vitesse, ni même si ce sera encore une bille quand elle arrivera au sol. De ce fait, il me semble assez réaliste de dire que nous n'avons pas de théorie générale, validée empiriquement, sur la réponse des populations, encore moins des écosystèmes, aux changements qu'ils subissent. Et nous n'avons donc pas une trame prédictive forte. Du point de vue appliqué, il est évident qu'une telle capacité prédictive nous serait utile, à plein de niveaux. Citons juste deux « serpents de mer »

des appels à projets : la réponse aux changements globaux et aux traitements en médecine et en agronomie.

Ce sujet concerne de nombreux collègues, des écologues des communautés aux paléontologues (je côtoie les deux dans mon unité), et nous l'attaquons chacun à une certaine échelle d'étude, orientés par une certaine expertise, un savoir-faire. Mon entrée est celle de la micro-évolution et la démographie, à relativement court terme, sur une ou deux espèces en interaction maximum, et de fait jusqu'à maintenant, dans les microbes, au laboratoire. La raison à ce choix est un mélange de contingences (liées à ce que je sais ou ne sais pas faire), de stratégie et de visée appliquée.

Pour la stratégie :

- L'évolution et la démographie expérimentales microbiennes donnent une possibilité de mesurer, en grande réplication et en environnement contrôlé, les dynamiques décrites par nos équations de base (sélection, mutation, dérive à minima). On a donc un fort pouvoir de test : on ne peut invoquer le fait qu'une équation ne prédit pas car on ne connaît pas bien le système observé. En gros, avant de passer en météorologie, on a validé les équations simples de la mécanique des fluides au laboratoire.
- Le contexte microbien jouit de développements nombreux de mesure expérimentale (biologie cellulaire et moléculaire, spectrométrie, micro-fluidique), et d'un corpus riche sur les mécanismes proximaux (métabolisme, génétique etc.) : on peut donc s'attaquer aux ultimes avec un certain confort.
- Le coût du travail expérimental en microbiologie est faible (il suffit d'un peu de sucre et d'un appareil pour mesurer les dynamiques), et il peut être interrompu (pas d'élevage à maintenir à long terme), ce qui est un atout dans un contexte de précarité du financement des projets et de l'aide technique.
- La modélisation qui va avec nécessite du temps (que j'ai grâce à un poste permanent) et des collaborateurs mathématiciens, dont nous ne manquons pas en France.

Pour la visée appliquée :

- Les microbes (et les cellules cancéreuses) constituent un élément clé de notre interaction avec le monde, en pathologie (humaine, animale, végétale), en pédologie, en physiologie de la digestion etc.
- Le problème de la dynamique démo-évolutive, monospécifique, est un élément clé de l'évolution de la résistance aux traitements. Le contexte écologique et épidémiologique, les transferts horizontaux entre espèces sont tous importants, mais ils se greffent sur un mécanisme de base : l'évolution de la résistance dans un hôte (ou un lieu) donné, par mutation, sélection, dérive. Il semble donc utile de prédire quantitativement ce mécanisme de base pour apporter des prédictions valides et robustes aux collègues travaillant aux échelles supérieures d'organisation.

Pour parvenir à produire et tester ces théories, divers verrous théoriques, mathématiques, et expérimentaux doivent être levés. J'ai donc principalement travaillé à lever ces verrous depuis pas mal d'années. Je résume ici ces travaux.

1. Développements théoriques :

a) Paysages adaptatifs, dynamique de l'adaptation et rescousse évolutive :

J'ai développé depuis ma thèse l'utilisation du modèle géométrique de Fisher (FISHER 1930) (FGM), qui est je pense le cœur de mon expertise de départ. Ce modèle suppose que la fitness dépend d'un ensemble de traits sous-jacents affectés de façon pléiotrope par la mutation, typiquement suivant une distribution multi-variée gaussienne (KIMURA 1965; LANDE 1980), mais pas nécessairement. Un environnement (biotique ou abiotique) donné détermine un optimum pour ces traits, via une fonction de fitness concave.

Ce modèle est l'outil central de l'écologie évolutive, lorsque celle-ci s'intéresse à l'adaptation à des environnements changeants, dans l'espace ou le temps (pour une revue voir KOPP AND MATUSZEWSKI 2014; TENAILLON 2014). J'ai montré (MARTIN AND LENORMAND 2006b) qu'on pouvait paramétrer le modèle avec des données de fitness, sans avoir à connaître ou mesurer tous les traits sous-jacents à l'adaptation (comme on le fait en génétique quantitative évolutive). Par ailleurs j'ai ainsi que d'autres auteurs montré que le modèle, bien que considéré jusqu'ici comme purement heuristique, pouvait prédire divers patrons observés dans la distribution de l'effet des mutations, leur forme générale (MARTIN AND LENORMAND 2006b) et leur variation entre contextes : épistasie (MARTIN *et al.* 2007 ; PERFEITO *et al.* 2014 ; SCHOUSTRA *et al.* 2016), environnement abiotique (MARTIN AND LENORMAND 2006a ; SOUSA *et al.* 2011 ; TRINDADE *et al.* 2012). J'ai aussi montré (MARTIN 2014) que, si ses hypothèse peuvent sembler simpliste, le FGM émerge en fait comme une approximation locale de modèles beaucoup plus complexes et généraux. J'ai aussi étendu ce modèle (MARTIN AND LENORMAND 2015a) à la distribution jointe des effets des mutations dans deux voire n environnements, résultats que j'utilise abondamment dans mon projet.

Toutefois, pour utiliser le modèle ou le tester sur des dynamiques évolutives et démographiques, il faut pouvoir décrire ces dynamiques à partir du modèle mutationnel de Fisher. Les outils existants à ce jour sont ceux de la génétique quantitative évolutive (e.g. LANDE 1984; VIA AND LANDE 1985; LYNCH AND LANDE 1993) qui supposent que la variance génétique des traits est à l'équilibre à tout temps, en présence de recombinaison forte. Cette approche ne peut donc se transposer directement aux microbes asexués. Le point central de mon travail depuis 5 ans a été d'implémenter les dynamiques adaptatives et démographiques, chez des asexués, lorsque le modèle de mutation est ce paysage adaptatif. Une première publication (MARTIN AND ROQUES 2016) a permis de modéliser cette dynamique de la distribution de la fitness chez des asexués, en réponse à un changement abrupte de l'environnement (introduisant nouvelle méthode d'analyse basée sur les fonctions génératrices). J'ai ensuite publié plusieurs articles théoriques prolongeant ces modèles, avec deux doctorants que j'ai co-encadrés Y. Anciaux (2016-2019, avec O. Ronce) et F. Lavigne

(en cours, avec L. Roques) et un doctorant en visite (M. Osmond, directrice S. Otto, UBC Canada). Nous avons implémenté :

- L'extension à un modèle anisotrope (où les traits de sont pas équivalents (HAMEL *et al.* 2020)) : nous avons montré ici que cette extension permettait de rendre compte des dynamiques à très long terme de la plus longue expérience d'évolution à ce jour (60 000 générations sur la bactérie *E.coli*, (BARRICK *et al.* 2009))
- Le couplage avec la démographie, en étudiant la probabilité d'extinction/rescousse évolutive dans le FGM, le long d'un continuum de bases génétiques : par une seule mutation (ANCIAUX *et al.* 2018), quelques-unes (OSMOND *et al.* 2020) ou un grand nombre (ANCIAUX *et al.* 2019)
- L'évolution d'un système source-puits (LAVIGNE *et al.* 2020) par adaptation du puits.

J'ai aussi produit un modèle pour l'adaptation à un environnement changeant avec une position arbitraire de l'optimum, l'article est en cours de rédaction. Ce dernier modèle sera fort utile pour les applications expérimentales : même *in vitro* et avec un stress abiotique simple (antibiotique) on ne peut réduire le changement environnemental à une dynamique simple (abrupt, ou linéaire etc.), voir ci-dessous.

b) *Modélisation de la pharmacocinétique in vitro :*

Les développements empiriques que je décris plus bas impliquent que je peux, dans mon système expérimental, mesurer en résolution fine les dynamiques de naissance et mort de bactéries soumises à des stress (antibiotiques, huiles essentielles, bactériophages, à ce jour). Le but est d'étudier la rescousse évolutive en direct, donc sur du moyen terme, mais il faut d'abord pouvoir caractériser la dynamique écologique pure de mes souches bactériennes dans le stress que je leur impose. Je pensais initialement qu'une dynamique très simple émergerait à court terme, au moins pour les stress abiotiques : croissance ou décroissance exponentielle selon la dose. En effet les études de « time-kill-curve » (patron de décroissance au cours du temps *in vitro*) en microbiologie sont nombreuses et les données sont ajustées par des modèles très simples. Toutefois les modèles donnent en fait typiquement un ajustement assez imparfait aux cinétiques observées. Cette imperfection est de fait ignorée en première approximation, et ne se voit pas trop sur des dynamiques expérimentales classiques, basées sur des comptages de colonies, et donc à relativement faible résolution temporelle.

Avec la fluoro-luminométrie, ce sont des points toutes les 5 minutes, sur une 20 aine de doses, qui sont observés de façon très répétables, on ne peut donc plus ignorer le mauvais ajustement des modèles de base. Il apparait en fait qu'il y a une dynamique de l'entrée et l'action de la drogue dans les cellules (pharmacocinétique) même dans le contexte *in vitro*, qui n'est à ma connaissance pas décrite ni modélisée (les modèles pharmacocinétiques s'appliquent plutôt aux problèmes *in vivo*, avec une injection de drogue, différents organes etc.). J'ai donc dû développer mes propres modèles mécanistes sur l'échelle cellulaire, diffusion de l'antibiotique vers son compartiment actif, efflux vers l'extérieur etc. Je pense être tout récemment arrivé à produire un modèle qui peut reproduire les time-kill-curve que j'observe au

laboratoire, je dois maintenant tester ces modèles quantitativement : ce sera une de mes premières missions pour 2020.

Ce passage est obligé pour ensuite aborder les dynamiques à plus long terme, lorsque des mutants résistants émergent et modifient ces cinétiques (mutants qui ont eux-mêmes leurs propres paramètres pharmacocinétique et pharmacodynamiques). J'ai donc dû intégrer une nouvelle littérature, mais ce fut très intéressant et stimulant. J'espère pouvoir obtenir des résultats qui parlent aux microbiologistes de cette manière. Aucun de ces travaux n'est encore publié, je veux un ensemble de données robustes et convaincantes (diversité d'antibiotiques) pour soumettre ces résultats (modèles + données), car la technique comme le modèle sont nouveaux donc il faut convaincre d'autant plus la communauté pertinente (microbiologique en l'occurrence). Je pense profiter de quelques interactions nouvelles avec des collègues microbiologistes de l'Université Montpellier pour avoir un retour de personnes habituées à cette communauté et cette littérature, nouvelles pour moi.

c) *Interdisciplinarité :*

Ces 5 dernières années ont été marquées par le renforcement de mes collaborations avec des mathématiciens, notamment avec Lionel Roques (INRA BioSP Avignon), avec qui j'ai obtenu une ANR (math-biologie B11 axe 1) et avec qui je co-encadre la thèse d'un étudiant mathématicien. J'ai de fait publié avec plusieurs mathématiciens, pour des modèles stochastiques (MARTIN AND LAMBERT 2015) ou de systèmes dynamiques et équations aux dérivées partielles (ROQUES *et al.* 2017; GIL *et al.* 2019). J'ai aussi été associé à des publications dans des journaux de mathématiques où nous donnions un cadre mathématique rigoureux aux modélisations que nous avons réalisées pour des applications en évolution (GIL *et al.* 2017; ROQUES *et al.* 2017; GIL *et al.* 2019). Enfin j'ai co-encadré un post-doctorant mathématicien (2017-2019) avec deux mathématiciens Montpelliérains (M. Alfaro et B. Cloez) dans le cadre d'un appel inter labex avec mon labex (CEMEB, écologie évolution) et les labex NUMEV (mathématique/informatiques) et AGRO (agronomie). Malheureusement la publication de ce travail est plus lente que nous espérons et je vais devoir rédiger moi-même l'article du post doctorant.

Je pense disposer maintenant d'un réseau fort de collaborateurs mathématiciens pertinents pour mon sujet d'étude (systèmes dynamiques, processus stochastiques) qui me permet d'aborder les enjeux analytiques comme numériques de mes modèles de façon plus rigoureuse et approfondie, je pense.

2. Développements expérimentaux :

a) *Paysages de Fisher et antibiotiques :*

Ma doctorante Noémie Harmand (2014-2018, co-encadrement avec T. Lenormand) a étudié l'adaptation d'E.coli à diverses doses d'antibiotique, à long terme. Notre but était de tester si la vision classique de l'écologie évolutive s'appliquait aux antibiotiques. Cette vision est celle d'un ensemble d'environnement sélectionnant pour des optimums différents (ici des optimums différents pour chaque dose). Cette vision très populaire (mais peu testée) en écologie/évolution est par contre très peu utilisée pour aborder la réponse aux antibiotiques, et le dosage de ceux-ci. Nous avons pu montrer que les données, à la fois de résistance à court terme (screen) et

de long terme (évolution expérimentale sur 1500 générations) étaient compatibles, qualitativement avec l'existence d'optimums à chaque dose d'antibiotique (HARMAND *et al.* 2017; HARMAND *et al.* 2018). Des tests plus quantitatifs doivent être faits maintenant, mais ce premier lot de résultats établit qu'on peut aborder la question de l'adaptation aux antibiotiques comme on le fait classiquement en écologie évolutive, avec des paysages de type FGM. Ils permettent aussi de tester l'existence de tels paysages, dans le contexte microbien, sur du long terme.

b) Fluoro-luminométrie :

Durant ces 5 dernières années et surtout depuis 2016, mon système expérimental de fluoro-luminométrie s'est beaucoup développé. Il s'agit d'une technique quasi inconnue. Je cite le terme employé par les premiers à l'avoir utilisée, à ma connaissance les seuls (LEHTINEN *et al.* 2003; LEHTINEN *et al.* 2006). Le principe (voir aussi le projet ANR en annexe) est de suivre des bactéries transformées pour exprimer la luminescence (seules les vivantes luminescent) et la fluorescence (vivantes et mortes fluorescent). J'ai ajouté à la technique un marqueur fluorescent de mortalité pour aussi suivre le compartiment cellules mortes. Ce système permet d'observer 3 signaux indépendants (basés sur des processus biologiques distincts) en temps réel, en milieu liquide (donc à forte réplication. On a donc 3 données en série temporelle fine, mesure automatisée, et haute réplication : une expérience typique (qui tourne sur un weekend automatiquement) peut suivre les 3 signaux dans 96*10 populations pendant 100 heures toutes les 30 mn, ou peut suivre 96 populations toutes les 5 mn. De ce fait on possède un fort pouvoir de rejet des modèles proposés, puisqu'on doit ajuster les trois signaux, dans toutes les populations. On a donc un fort pouvoir de test du modèle, et s'il est valide une grande précision d'estimation des paramètres.

La technique est maintenant fonctionnelle au laboratoire, notamment depuis ma collaboration renforcée avec R. Froissart (CNRS, UMR MIVEGEC), et les stages successifs d'étudiants en masters et L3 depuis 2017. Nous avons observé de nombreuses cinétiques, rescousse évolutives, sur des antibiotiques, des bactériophages et des huiles essentielles. Cela a permis de montrer la fonctionnalité de la technique et sûrement a aidé à obtenir le financement ANR. Ces développements ont été rendus possible notamment par un PEP II CNRS qui a permis de se lancer.

Je suis donc au stade où je peux comparer données haut-débit et modèles divers, le but du projet global évidemment, et ce sera ma tâche principale pour ces 5 ans.

3. Des pistes à explorer pour les années à venir :

Mon projet de recherche à long terme est de travailler à ce qui me semble des verrous importants à ce stade :

- Publier et développer le volet test expérimental de ce duo en testant les prédictions (ER ou FGM + ER) existantes, mais aussi en étendant le système expérimental pour pouvoir mesurer des situations plus complexes (stress dynamiques, migration). Les analyses des données existantes ont tout juste commencé. Il s'agira aussi de relier ces méthodes aux pratiques classiques

de microbiologie (en clinique etc. etc.) pour savoir comment mieux utiliser les mesures classiques ou les améliorer sans développer tout un système expérimental dédié justement.

- Tester d'autres prédictions (FGM sans ER) sur les données existantes de l'évolution expérimentale (notamment LTEE), à la fois les dynamiques de fitness mais aussi de séquences génomiques. J'aimerais aussi tester les prédictions du FGM sur les distributions d'effets des mutations (comme je l'ai fait en thèse), en utilisant l'explosion de données expérimentales actuelle (deep mutational scan method).
- Développer des modèles particuliers à chaque stress expérimental (antibiotiques, phages lytiques, prédateurs) au fur et à mesure que les besoins se feront sentir.
- Implémenter la modélisation de l'ER de façon plus 'exacte' en utilisant des outils des processus stochastiques plus généraux (Backward Kolmogorov Equations) et non les approximations que nous avons développées jusqu'ici (thèse de Yoann). J'ai tenté d'explorer cette piste pour la rédaction de cette HDR, sans aller très loin encore donc je ne la décrirai pas.
- Développer des approximations explicites pour les situations de sélection forte (SSWM) dans le FGM. En effet, pour diverses situations, nous parvenons à écrire une équation aux dérivées partielles (**EDP**) qui décrit le processus de sélection – mutation (dérive négligée) sans approximation. Toutefois, cette EDP n'est pas toujours évidente à résoudre, même numériquement. Les résultats analytiques ne sont obtenus que par une approximation d'effets faibles des mutations (WSSM). J'aimerais développer une approximation complémentaire, d'effets forts. J'ai tenté d'en dresser les contours lors de la rédaction de cette HDR (dynamic House of Cards approximation) mais ce n'est qu'embryonnaire je crois.
- Développer des simulateurs (individu centré) efficaces pour les dynamiques bactériennes en grande population polymorphes (les codes existants sont basés sur du temps discret qui n'est pas approprié, ou ne suivent que quelques mutants). Développer des solveurs numériques précis pour les EDP dérivées dans nos modèles.

Ces différents projets devraient pouvoir me tenir occupé, ainsi que mes collaborateurs et étudiants potentiels, pour pas mal d'années je pense.

[4. Une remarque sur l'encadrement et l'avenir de nos étudiants...](#)

Puisqu'il s'agit d'un document visant à évaluer notre capacité à encadrer des étudiants, il me semble logique d'exprimer mon opinion sur pourquoi et dans quelle mesure je devrais encadrer des étudiants dans le futur.

J'hésite fortement à former de nouveaux étudiants au niveau du doctorat, tant que ceux que j'ai formés ne seront pas installés professionnellement (dans l'académique ou pas). Il me semble que le contexte d'embauche ne se prête pas à former des docteurs en nombre dans mon domaine, qu'ils seront aussi bien à s'orienter dès la fin du M2 vers un travail hors recherche académique.

Par ailleurs, je ne vois pas facilement de débouchés dans le privé pour les étudiants que je forme. Il est possible qu'à terme leurs compétences en microbiologie ou modélisation leur servent à une embauche en entreprise, mais ce sera nettement plus facile pour un élève qui sort d'une sacro-sainte école d'ingénieur et n'a fait qu'un petit stage de 3 mois (je le sais j'en viens). Tant qu'Université et Prépas seront séparés je ne vois pas facilement mes doctorants trouver de l'embauche via leur doctorat, à l'opposé de ce que j'ai pu voir en Suisse ou aux USA bien entendu. Si le fait de tourner mon projet expérimental et de modélisation vers l'appliqué (pour d'autres raisons) permet que les étudiants que je forme intéressent des domaines hors académiques tant mieux, mais pour le moment je ne vois pas cela arriver de si tôt.

Ma stratégie est donc de limiter l'encadrement, au fur et à mesure que mes anciens doctorants se 'casent', soit dans une autre vie (Noemie Harmand par ex.) soit dans l'enseignement (Florian Lavigne a priori) soit l'académique. Si on ne remplace même pas les départs à la retraite dans la Recherche, je ne vois pas ce qui justifierait de former 20 docteurs dans une vie.

H. Liste exhaustive des publications

Par ordre chronologique inverse (en rouge les publications avec un(e) doctorant(e) encadré(e) :

1. Roques, L., F. Patout, O. Bonnefon, and **G. Martin**, Adaptation in General Temporally Changing Environments. *SIAM Journal on Applied Mathematics*, 2020. 80(6): p. 2420-2447.
2. Osmond, M.M., S.P. Otto, and **G. Martin**, Genetic Paths to Evolutionary Rescue and the Distribution of Fitness Effects Along Them. *Genetics*, 2020. 214(2): p. 493-510.
3. Lavigne, F., **G. Martin**, Y. Anciaux, J. Papaix, and L. Roques, When sinks become sources: Adaptive colonization in asexuals. *Evolution*, 2020. 74(1): p. 29-42.
4. Hamel, F., F. Lavigne, **G. Martin**, and L. Roques, Dynamics of adaptation in an anisotropic phenotype-fitness landscape. *Nonlinear Analysis: Real World Applications*, 2020. in press.

5. Gil, M.E., F. Hamel, **G. Martin**, and L. Roques, Dynamics of fitness distributions in the presence of a phenotypic optimum: an integro-differential approach. *Nonlinearity*, 2019. 32(10): p. 3485-3522.
6. Anciaux, Y., A. Lambert, O. Ronce, L. Roques, and **G. Martin**, Population persistence under high mutation rate: From evolutionary rescue to lethal mutagenesis. *Evolution*, 2019. 73(8): p. 1517-1532.
7. Harmand, N., R. Gallet, **G. Martin**, and T. Lenormand, Evolution of bacteria specialization along an antibiotic dose gradient. *Evol Lett*, 2018. 2(3): p. 221-232.
8. Anciaux, Y., L.M. Chevin, O. Ronce, and **G. Martin**, Evolutionary Rescue over a Fitness Landscape. *Genetics*, 2018. 209(1): p. 265-279.
9. Roques, L., J. Garnier, and **G. Martin**, Beneficial mutation-selection dynamics in finite asexual populations: a free boundary approach. *Sci Rep*, 2017. 7(1): p. 17838.
10. Kondombo, C.P., M. Deu, J. Chantereau, **G. Martin**, E. Chapuis, P. Letourmy, J.D. Zongo, F. Sagnard, and K. Vom Brocke, Patterns of genetic structure and phenotypic diversity in sorghum landraces in relation to farmers' management in Burkina Faso. *International Journal of Biological and Chemical Sciences*, 2017. 10(4).
11. Harmand, N., R. Gallet, R. Jabbour-Zahab, **G. Martin**, and T. Lenormand, Fisher's geometrical model and the mutational patterns of antibiotic resistance across dose gradients. *Evolution*, 2017. 71(1): p. 23-37.
12. Gil, M.-E., F. Hamel, **G. Martin**, and L. Roques, Mathematical Properties of a Class of Integro-differential Models from Population Genetics. *SIAM Journal on Applied Mathematics*, 2017. 77(4): p. 1536-1561.
13. **Martin, G.** and L. Roques, The Nonstationary Dynamics of Fitness Distributions: Asexual Model with Epistasis and Standing Variation. *Genetics*, 2016. 204(4): p. 1541-1558.
14. **Martin, G.** and T. Lenormand, The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution*, 2015. 69(6): p. 1433-1447.
15. **Martin, G.** and A. Lambert, A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theor Popul Biol*, 2015. 101: p. 40-6.
16. **Martin, G.**, Fisher's geometrical model emerges as a property of large integrated phenotypic networks. *Genetics*, 2014. 197(1): p. 237-255.
17. Alexander, H.K., **G. Martin**, O.Y. Martin, and S. Bonhoeffer, Evolutionary rescue: linking theory for conservation and medicine. *Evol Appl*, 2014. 7(10): p. 1161-79.
18. **Martin, G.**, R. Aguilée, J. Ramsayer, O. Kaltz, and O. Ronce, The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments. *Philos Trans R Soc Lond B Biol Sci*, 2013. 368(1610): p. 20120088.
19. Manna, F., R. Gallet, **G. Martin**, and T. Lenormand, The high-throughput yeast deletion fitness data and the theories of dominance. *J Evol Biol*, 2012. 25(5): p. 892-903.

20. Manna, F., **G. Martin**, and T. Lenormand, Fitness landscapes: an alternative theory for the dominance of mutation. *Genetics*, 2011. 189(3): p. 923-37.
21. **Martin, G.** and S. Gandon, Lethal mutagenesis and evolutionary epidemiology. *Philos Trans R Soc Lond B Biol Sci*, 2010. 365(1548): p. 1953-63.
22. Magalon, H., T. Nidelet, **G. Martin**, and O. Kaltz, Host growth conditions influence experimental evolution of life history and virulence of a parasite with vertical and horizontal transmission. *Evolution*, 2010. 64(7): p. 2126-38.
23. Chevin, L.M., **G. Martin**, and T. Lenormand, Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution*, 2010. 64(11): p. 3213-31.
24. **Martin, G.** and L. Lenormand, The distribution of beneficial and fixed mutation effects close to an optimum. *Genetics*, 2008. 179: p. 907 - 916.
25. **Martin, G.**, E. Chapuis, and J. Goudet, Multivariate QST-FST comparisons: a neutrality test for the evolution of the g matrix in structured populations. *Genetics*, 2008. 180(4): p. 2135-49.
26. Chapuis, E., **G. Martin**, and J. Goudet, Effects of selection and drift on G matrix evolution in a heterogeneous environment: a multivariate Qst-Fst Test with the freshwater snail *Galba truncatula*. *Genetics*, 2008. 180(4): p. 2151-61.
27. **Martin, G.**, S.F. Elena, and T. Lenormand, Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat Genet*, 2007. 39(4): p. 555-60.
28. Goudet, J. and **G. Martin**, Under neutrality, $Q_{st} \leq F_{st}$ when there is dominance in an island model. *Genetics*, 2007. 176: p. 1371-1374.
29. **Martin, G.**, S.P. Otto, and T. Lenormand, Selection for recombination in structured populations. *Genetics*, 2006. 172(1): p. 593-609.
30. **Martin, G.** and T. Lenormand, The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution*, 2006. 60(12): p. 2413-27.
31. **Martin, G.** and T. Lenormand, A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution*, 2006. 60(5): p. 893-907.
32. Thomas, F., A. Schmidt-Rhaesa, **G. Martin**, C. Manu, P. Durand, and F. Renaud, Do hairworms (Nematomorpha) manipulate the water seeking behaviour of their terrestrial hosts? *Journal of Evolutionary Biology*, 2002. 15(3): p. 356-361.

III. Aim of my research

My research in general is an attempt to model and predict the demographic and evolutionary dynamics of populations facing environmental change. This goal is shared with many other researchers, as it is central to many applications (medical, agronomic, in conservation) and it is at the core of, one could say it is the essence of, ecology and evolution. Many issues of current importance are in fact demographic in the end, even when their key underpinnings are evolutionary: whether a pathogen population will die out or grow in the face of a treatment, whether a given population will invade a new habitat (invasion, emerging diseases etc.), whether a population will get extinct due to environmental change (conservation). It is also the essence of evolution and ecology, in that these fields have emerged as a (mature?) "scientific object" by framing biological systems into the concepts of dynamical systems (DEPEW *et al.* 1996). Transient dynamics are also essential, rather than steady states and equilibria, because we can rarely observe the latter in the lab (they take too long to be reached).

My own approach, which again I share with several colleagues (collaborators or not) is to tackle this issue in a simplified context: that of microbial experimental evolution and simple ecological models (asexual populations so far, with the hope to extend to sexuals and to two coevolving species). In exchange for this harsh simplification, the ambition of the experimental and theoretical approach is high, aiming to quantitatively capture and predict observed dynamics at a fine timescale, in controlled conditions. This ambition to quantitatively *predict* (not just *fit*) observed dynamics in controlled lab contexts is akin to what was undertaken in classical physics (mechanics, thermodynamics, magnetism etc.). At least it is so in my, maybe naïve, view of classical physics. I believe that, if we could quantitatively describe, using general mathematical equations, the demographic and evolutionary dynamics that we observe in the lab, it would lead to much more powerful tests of our general theories about mutation, selection, migration or sex to take a few examples. It could also lead to more consensus on experimental observations across labs, as is typical in physics: the quality and calibration of the experimental data would then be judged based on more or less accepted quantitative expectations in simple situations, before using the experimental system to study new scenarios. Finally, it could allow to use these equations as building blocks for a quantitative simulation of more complex natural systems.

I can see two major wagers in this research 'program'. First, several colleagues I have met over the years will argue that even lab eco-evolutionary dynamics are too complex to be captured, in their main aspects, by what will necessarily be simplified models (in order to be general they must be somewhat simplified). Second, other or the same colleagues argue that even if we could capture some general quantitative laws of lab dynamics, their translation into natural systems will not be possible (open systems, too many hardly measurable confounding factors, biotic interactions impossible to reproduce in the lab etc.).

I have no scientific argument to oppose these valid critics, I can only state that it seems too early to tell whether this approach is a waste of time or not. In any case, it is my impression that the majority of biologists I meet are convinced that such quantitative validation of general simple equations is rather doomed, so that only a

few of us will end-up wasting time, in the worst case scenario. I also would like to point that if no simple law can prove sufficiently correct to be *de facto* useful in evolution and ecology, then teaching of these subjects should shift to less quantitative content. Indeed, in most evolution and ecology classes that I know of, many simple equations are used (frequency dynamics under selection and mutation, Fisher's fundamental theorem of natural selection, predator-prey or logistic growth dynamics etc.). Yet typically, a limited amount of quantitative empirical support is provided for these equations, as teaching material. If experiments and modelling developments could at least provide such support in lab conditions, it would already be useful I think.

On the other hand, some simple equations are already well supported by regular observations: exponential growth is the limiting behavior of many stochastic models of population dynamics and is indeed observed on a daily basis in the lab, on microbes (of course for some finite time). The development of microfluidics or various measurement methods in spectrometry also allow to observe population dynamics (again of microbes) at a fine time-scale, up to an individual level, on large amounts of replicates, for a low cost. In the case of spectrometry (which I practice), the cost is that of the machine (25 000€ in our case), which typically lasts for more than 20 years, a bit of sugar, and microplates (~5€) for 96 populations over as many hours as one can (120h typically in our system, with measures every 20-30mn). Having a semi-low cost measurement apparatus is important scientifically, because it makes us relatively independent of funding, which provides liberty and stability of research programs. A minimal research program based on genetic data (sequences, NGS) hardly costs less than several k€, for one experiment, if not tens of k€ (especially at the population scale of evolutionary ecology): it cannot be ran over long periods without funding. It is thus useful that some of us should try to develop these low-cost experimental designs.

This cost-effectiveness is one rationale for the program I propose, even at the risk of eventually being doomed to never achieve its ambition of quantitative prediction. Low-cost means that (i) many of us can do it (ii) that the experiments can be repeated, to provide ample replicated demographic data, and (ii) that the corresponding models are forced to be tailored for such cheap experimental data. They cannot be relying on observations of genotypic, gene-regulatory, metabolic or environmental states, all of which can only be acquired at substantial expense. These are present in the model (more or less implicitly), but they must be shunt, just as the detailed genetic basis of a quantitative trait is shunt in the breeder's equation, on purpose (because we cannot access it all and the biologists certainly could not in the early 20's).

The rationale for aiming at quantitative rather than qualitative insight is the obvious one, testability: there are many ways in which a model can produce a valid qualitative prediction without having captured the key underlying mechanisms of the observation, it is less likely for a quantitative prediction. The other reason is application: a quantitative theory allows for optimization and measurements that are not possible with a qualitative theory, or at least it is my impression.

In the following section I discuss the minimal ingredients that I think are required to come up with an empirically testable quantitative theory for the observations of

(eco) evolutionary dynamics in the simplified lab context, and the minimal requirements of the corresponding experimental systems.

In the next section, I describe the experimental system that I have set-up over the past years and which has been used by several Masters students so far. I then show a set of results that I have obtained so far (and not published yet). This will be the basis of the experimental projects that I intend to pursue in the years to come.

In the last section, I describe the general modelling framework that I intend to use in the years to come, some of its subcases correspond to work that I have published so far, with some of my PhD students. The goal is to illustrate the mathematical objects that I intend to work with, their potential applications to particular questions, and the challenges associated with their analysis.

The goal of these two chapters is to provide new material rather than summarize past work, and to show the jury the basic tools that I intend to propose to future students, in their research projects with me.

IV. Rationale for the approaches used

A. *Minimal requirements of the models:*

For the purpose of modelling the demographic and evolutionary dynamics in a changing environments, there are some minimal elements that the model must contain.

1. Transient dynamic

We must be able to describe the transient dynamics of genotype frequencies and population sizes, or equivalently of genotype densities. Many of the applications we have in mind and the experimental dynamics we can observe occur over timescales smaller than that of equilibria, we must rely on modelling approaches that allows to capture transient dynamics. This eliminates analytical approaches based on solving for equilibria, and can limit the use of long time limit behaviors. The latter may still be useful on occasions where a separation of timescales argument allows to use the long time limit of the fast process.

2. Evolution of demographic traits

Fecundity, survival, birth or death rates, virulence and transmission rate (etc.) are traits that directly enter the mathematical models of demographic dynamics of genotype densities, and hence fitness. They connect the demographic/ecological process and the evolutionary process for populations with multiple genotypes. I denote them Fitness Components (**FCs**), which I think is fairly classic. Several FCs like survival or fecundity are sometimes called life-history traits, but I tend to keep a distinction with these traits as the latter encompass a wider diversity of traits such as weight or size at maturity, which do not directly enter the computation of fitness as part of demographic model. The FCs are those traits that directly govern the demographic dynamics of the populations facing environmental change. In the framework I will use we will see many

examples where they contribute to fitness *additively* (e.g. growth rate $r = b - d$ birth minus death rate).

This distinguishes FCs from 'optimized traits' in phenotype-fitness landscapes (discussed in MARTIN AND LENORMAND 2006b; MARTIN 2014): the former are the y-axis of a FC-landscape while the latter are on the x-space of the landscape (see also below). The former contribute to fitness additively, while the latter contribute to fitness via a concave function (defining an optimum): there is no optimum, in any environment, for birth rate itself (the larger the better), while there may be an optimum for any metabolic rate or gene expression level in given conditions.

3. Theory for the distribution of mutation effects on FCs

Any FC is a continuous trait with a wide genetic basis (most mutations have some effects on fitness or fitness components, even if mild). Therefore, we cannot use two-alleles or few-alleles theory to describe its evolution, we must rely on the population genetics of continuous traits, with many different alleles (pre-existing or mutant) that can arise stochastically by mutation and have a diversity of effects on fitness or multivariate effects on FCs. Since Kimura's (1965) seminal work, the effects of a plethora of alleles on a continuous trait are typically handled through a given distribution of these allelic effects (with a few parameters, instead of infinitely many single types' parameters). When considering effects on fitness this distribution is called the Distribution of Fitness Effects (**DFE**). Choosing a realistic yet mathematically manageable DFE is a central challenge of all models of fitness evolution.

Furthermore, epistasis in mutations fitness effects is empirically pervasive, whether among random mutations (REMOLD AND LENSKI 2004), beneficial mutations (CHOU *et al.* 2011; KHAN *et al.* 2011) or antibiotic resistance mutations (MACLEAN *et al.* 2010). The choice of the DFE model must thus ideally include a shape of single effects and an epistasis pattern both close enough to observed mutant data.

Handling infinitely many alleles imposes to use some form of fitness landscape relating infinitely many genotypes with their fitnesses. There is a wealth of such landscapes which I try to classify according to **Table 1** below. A fitness landscape always produces a correspondence between a genotype and a fitness (because a fitness is only defined for a genotype not for a phenotype). However, it can do so by characterizing the genotype in different ways and by focusing on the individual genotype scale or on a population average over co-segregating genotypes.

Characterization of the genotype / scale of observation	Individual genotype	population's average
Genotypic position \mathbf{h} (or hamming distance h)	NK landscape, Mount Fuji etc. genotype-fitness landscapes: $W = f(\mathbf{h})$	Wright's allele frequency - mean fitness landscape: $\bar{W} = f(\mathbf{p})$
Phenotypic position (Breeding value for phenotype \mathbf{x})	Fisher's geometric model of adaptation: $W = f(\mathbf{x})$	Lande's mean phenotype - mean fitness landscape: $\bar{W} = f(\bar{\mathbf{x}})$

Table 1: various types of fitness landscapes: genotype – fitness landscapes in red, phenotype-fitness landscapes in green.

There are two possible types of quantities characterizing a genotype in the landscape (in the x-space of the landscape):

- A genotype sequence (a position \mathbf{h} in some genotype space, for ex. 4^L possible combinations of 4 nucleotides at L sites or locus. Note that this has the limit that deletions and insertions are not defined here. An alternative is the use of Hamming distance h cumulating all the genetic differences with some origin of sequence space (a reference genotype).
- A phenotypic value associated with a genotype (i.e. a breeding value for that phenotype): this can be typically some trait for which there is an optimum, but it can also be the slope of a reaction norm of another phenotype etc.

There are two possible scales at which both the genotype and the fitness are characterized:

- the individual genotype's scale: its particular sequence or breeding value determines its fitness
- the population scale: the frequency of a given allele at some locus, or the mean breeding value of the population for a phenotype determines the population's mean fitness. Note that the frequency of an allele p is the average value of a random variable $X = \{0,1\}$ characterizing the allelic state of each individual in the population, in a biallelic model ($p = \bar{X}$), hence it is a population's average just as \bar{x} .

Together, these two criteria allow to classify various famous landscapes into distinct classes : Wright's (1932) landscape relating genotype frequency and mean fitness, Lande's (1976) landscape defining selection gradients acting on phenotypic means, various landscapes relating sequence and fitness (NK, Mount Fuji etc.) and Fisher's (1930) geometrical model of adaptation (**FGM**). For a historical perspective on genotype-fitness landscapes see (MCCANDLISH 2011), for one on Fisher's geometrical model see (TENAILLON 2014).

The population-scale landscapes (right column of Table 1) lead to clarifying insight on the process of natural selection, but they may not gather sufficient information to close a system of equations in a context where new genotypes arise regularly (by mutation or recombination). The individual-scale landscape's dynamics can be used to derive the population – scale landscape's dynamics but not the other way around. Therefore, we will use individual scale landscapes here.

Among individual genotype-scale landscapes, the genotype-based ones allow to implement potentially any form of epistasis. This advantage comes at the cost of having no clear relationship to environments or environmental changes (see next point). Phenotype-fitness landscapes impose a particular form of epistasis (depending on the fitness function used), but it has the main advantage of providing a natural link to the environment (assuming that it sets the optimum's position, a *minima*). Hence I

have been studying these landscapes mostly (although some of my work has aimed to handle more general landscapes (MARTIN AND ROQUES 2016)).

4. Theory for $G \times E$ on fitness components

Because we consider the response to environmental changes, we must have a theory for how the environment will change, in a predictable manner, the genetic effects (of mutations, of segregating alleles) on FCs. Hence we must have a tractable theory for how the joint distribution of fitness or multiple FCs changes as the environment changes, and we must have such a theory for a continuum of environments, to be able to tackle continuous change in abiotic conditions over time or space, or coevolution with a continuously changing biotic partner. This removes most existing fitness landscape models from the options: any genetic landscape (NK landscapes and the likes) does not give any rule for how allele fitnesses (heights of the peaks) change with external conditions. Thus, the most natural tool to use in this context is a phenotype-fitness landscape, where external conditions set the optimum of the underlying phenotypes, hence Fisher's geometrical model of adaptation (FGM). Indeed, although it is not always cited this way, most general theories of adaptation to changing environments in space and time use the FGM framework (evolution of a species range, invasion of a spatial gradient, adaptation to 'climate change' etc.). In this framework, fitness is a concave function of a set of n phenotypic traits, which vary by mutation (with some distribution typically centered on the parental type) and potentially recombination. Several simplifying assumptions are most often made regarding the genetic basis of the phenotypic traits. Mutation effects are additive across loci/sites, along a chromosome (no epistasis) or between chromosomes (no dominance). Mendelian inheritance is assumed upon sexual reproduction with either an explicit genetic basis, L biallelic loci or L loci with infinitely many alleles (KIMURA 1965; LANDE 1976), or in an infinitesimal limit ($L \rightarrow \infty$), e.g. (BULMER 1971), under the so-called infinitesimal model (for an explicit derivation and historical perspective, see BARTON *et al.* 2017). An infinite allele model is used to describe mutation effects, which are drawn *de novo* into a given multivariate distribution.

A sketch of the FGM for two FCs (birth and death rates) is given in **Figure 1** below.

The central assumptions in (all versions of) the FGM are :

- (i) Frequency-independent selection: so that a given genotype with given breeding value for phenotype has fixed fitness (or FCs values) whatever the composition of the population with which it is interacting.
- (ii) Existence of a local optimum (concave fitness function) which may depend on external conditions.
- (iii) Continuous unbiased variation in breeding values among mutants, that is centered around the parental value

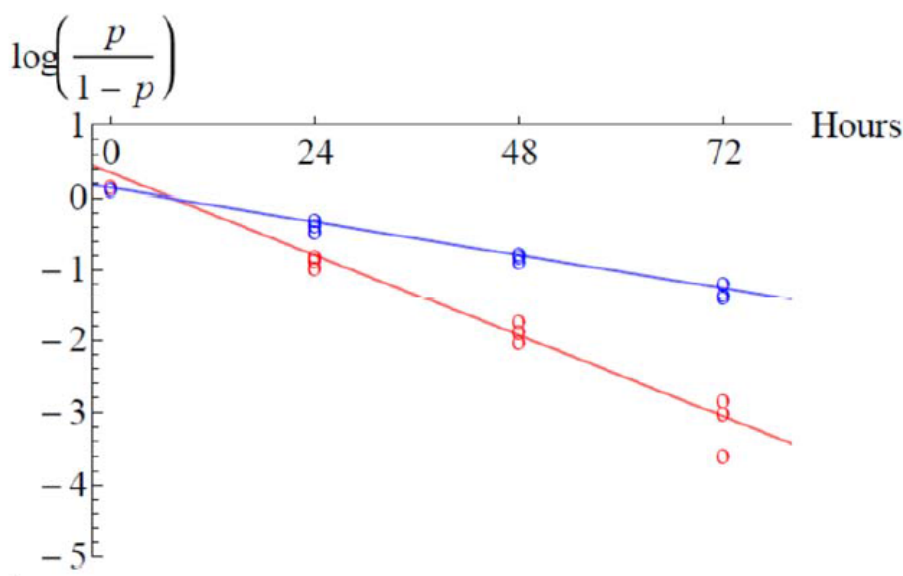
The applicability and biological realism of this framework is amply discussed in, for ex. (MARTIN AND LENORMAND 2006a; MARTIN AND LENORMAND 2006b; MARTIN 2014; TENAILLON 2014). I quickly discuss the realism of each assumption below.

Assumption (i) is regularly tested in experiments measuring fitness relative to a reference competitor by following the relative frequency of each type over several time-points (not just two): either by head to head competitions (e.g. LIND *et al.* 2010; HARMAND *et al.* 2018) or by pooled types competition with regular re-sequencing of the co-segregating types (e.g. HIETPAS *et al.* 2013). Whenever the frequencies $p_i(t)$ of types $i \in [1, K]$ has a trajectory well fitted by the equation:

$$\begin{aligned} \partial_t p_i(t) &= (m_i - \bar{m}(t)) p_i(t) \\ \bar{m}(t) &= \sum_{i=1}^K m_i p_i(t) \end{aligned}$$

Then frequency-independent selection is *a priori* valid: there is a fixed growth rate m_i of type i irrespective of the frequency of any of the co-segregating types. The classic subcase is selection between two types $i = \{1, 2\}$ with $s = m_2 - m_1$ leading to the famous equation $dp/dt = s p (1 - p)$ for the frequency p of type 2. The finding of a straight line over time for $\log(p/(1 - p))$ is a validation of the frequency-independent equation (and its slope estimates relative fitness s).

This is illustrated in **Figure 1** below by data from one of Noémie Harmand's experiments (former PhD student):



Log (CFP/YFP) between two fluorescently marked strains of E.coli, one selected for resistance to nalidixic acid (CFP) and the other its ancestor. The points show three replicates for two distinct resistant strains, in the absence of drugs.

Figure 1: frequency – independent selection yields linearity of logit frequencies over time.

The frequency-independent assumption is so widely accepted in experimental evolution that, quite often, authors directly estimate s from the change in $\text{logit}(p)$ between two time points, shunting the test of frequency-independence itself (e.g. LENSKI *et al.* 1991). Overall, frequency-dependent selection does occur on occasions, even in the lab (e.g. cross-feeding mutants in VASI AND LENSKI 1999). However, it is rather the exception than the rule, at least among spontaneously arising mutants of the same species, in the lab conditions where it has been measured so far. While some biologists tend to find the simplified frequency-independent selection 'inelegant', it is good news that it should be so easily testable and so frequently validated. Should frequency-dependence be the rule, we would need to estimate the relevant parameters of this more complex interaction (from competitions at various starting frequencies). We would require a model for how the frequency-dependence terms change with the environment and the frequencies of all co-segregating types, at any time and across environments: a daunting task whenever more than a few genotypes co-segregate in a population. Frequency-independent selection is thus a key null assumption to build up a process from microscopic individual genotypes' characteristics to population-scale competition between arbitrarily many types.

Assumption (ii): the existence of a local optimum for phenotypes is at the core of many evolutionary theory, from Darwin's definition of a species niche (the famous Galapagos finches examples), to the concept of trade-offs (between environments, between life history traits). In fact, if we accept the existence of some arbitrary function relating fitness and an underlying multivariate phenotype, and that mutations are small perturbations on this phenotype, then the *mean* effect of random mutations on fitness can only be deleterious if the fitness function is concave (hence tends to define an optimum). As we observe such mean deleterious effects in almost all experimental measures of random mutation fitness effects, there is substantial support for the existence of concave fitness functions, and thus a single local optimum (MARTIN AND LENORMAND 2015b).

Assumption (iii) is not critical in itself: we could imagine to derive a modified model where mutations tend to have some biased effect on the underlying phenotypes of the FGM. However, the shape of observed DFEs in optimal conditions suggests otherwise: if there was a non-negligible bias, among random mutants, from their parent's optimal phenotype (in optimal conditions), then DFEs would also be shifted and could not encompass continuous deleterious variation up to zero (discussed in MARTIN 2014).

5. Handling multiple mutants with epistatic effects

Predicting the dynamics of fitness over time with any given fitness landscape is not straightforward. Indeed, one must follow a dynamical system with an ever changing set of variables (newly formed genotypes) undergoing selection, mutation and drift *a minima* (for isolated asexual populations in constant conditions). There is, to my knowledge, no general mathematical framework to *exactly* model these dynamics, even in the simplest non-epistatic fitness landscape. However, many approximate analytical approaches have been developed over the years, and I will present one such approach, based on tracking generating functions of the distribution

of polygenic traits (BURGER 1991; HANSEN 1992; TURELLI AND BARTON 1994). One could say that, in the method as we implemented it so far (since MARTIN AND ROQUES 2016), the least well treated force is genetic drift. The reformulation proposed in this document is an attempt to correct for this limit as much as possible.

6. Handling stochastic events.

The demographic process is inherently stochastic at the scale of individual life histories (reproduction, migration, mutation and death events). This stochasticity may sometimes be ignored but genotype densities will always be transiently small at some point, a *minima* when a type newly arises (by a single mutation, recombination or migration event). Therefore, a minimal requirement for models of evolution and demography is to explicitly account for this stochasticity, even if to later ignore it in some explicit limit. One of the central attempts of the modelling presented here is, like in most models of adaptation, to explicitly derive a deterministic mathematical object (a PDE or coupled PDE-ODE) from these microscopic stochastic events.

B. Minimal requirements of the experiments:

These are determined partly by the requirements of the theory.

1. Replicate trajectories

As the processes involved in the evo-demographic response to environmental change can be highly stochastic, it is necessary to measure a large number of replicates, and to produce time-series of their dynamics, not just endpoints. This suggests using microbes (for which many generations elapse over a few tens of hours), and either time-lapse microscopy (e.g. microfluidics) or spectrometric measurements of incubated microplates in a microplate reader. Microfluidics is more precise than microplate reading, in that it gives information at the scale of individual cells, over very numerous replicates. However, it can be expensive and technical to set up from scratch, and it can constrain strongly the ecological setting (colibaciles dividing along a single line, in the basic setting): for example I am not sure that bacteriophage infections are easy to perform in a microfluidics context. As I was a beginner in microbiology, I thus decided to set for microplate reading techniques. Note that nowadays, single-cell information in microplates can be obtained by microplate time lapse microscopy: the Cytation® reader (BioTek) does this, and one is now available next door in Emmanuel Fronhoeffler's lab. We will likely do some tests with bacteria for our collaborations in the context of my current ANR RESISTE.

2. Measuring distinct FCs

In order to test predictions on the evolution of FCs, they must be measurable, which means that distinct demographic traits must be available to measurement, not just fitness itself. In the experiments I will present we seek to access the live and dead cell compartments in order to have a measure of the birth and death rate separately (not just the growth rate which is classically measure in microbial demography). I do this using a mixture of three markers: for live, dead and total cells. The use of three markers

allows an internal check of the underlying model: all three markers must yield trajectories consistent with each other, under a common model of the birth and death process induced by the stress imposed (antibiotics or phages so far).

3. *Controlled, measured or inferred environmental conditions*

The populations considered must be closed so that replicates are truly independent (i.e. avoid contamination). The environmental variables (temperature, stress etc.) must be either fixed (thermostat, chemostat etc.) or variable but measurable (e.g. drug dose in a variable dose regime), or we must be able to infer them from the measured dynamics, via a model (e.g. phage density when only measuring bacterial dynamics in phage–bacteria coevolution, dynamics of antibiotic concentration within cells by diffusion and binding etc.). Plate readers allow to control conditions quite substantially (temperature, agitation), additional elements (drug, phages) can be added by injection within wells without opening the system (but the wells must then be open, with a risk of contamination). I use antibiotics (passive elements easy to control), lytic phages and protist predators (dynamic elements which density must be inferred via a model). I plan to use fluorescent antibiotics in 3D printed microplates so that controlled and monitored diffusion of the drug is possible (to better reproduce pharmacokinetics in vitro).

V. Empirical approach and projects

A. Existing methods

1. *Colony forming units (CFUs) the “gold standard”*

The broad goal of my experiments is to measure the birth and death process the bacterial cells undergo, as it happens, in the face of various stressors (chemical, virus, predator etc.). We further wish to see the effect of various levels of a given stress, and to have access to many replicates, in order to capture the effect of demographic stochasticity and rare mutational events. This requires to be able to measure live and dead cells at regular time intervals, in many parallel culture vials, which is not straightforward.

In microbiology, Colony Forming Units (CFU) are the gold standard to measure the number of live cells in a liquid medium. One takes a sample of the culture at time t , dilutes it by some factor F , and plates it on a Petri dish with agar and sugar. After some incubation time (typically 24h), any live cell that was sampled and dropped somewhere on the surface of the agar will have divided to form a ‘colony’ (typically a visible colony will consist of $\sim 10^6$ cells). A colony is thus really a population of many aggregated cells all descending from one original live cell from the sample, that dropped at the center of the now visible colony. To produce a time series of the density of live cells in a culture, one thus takes regular samples of the culture, dilutes and plates each on a dish, incubates plates for ~ 24 h, then counts the number of CFUs on each plate, to infer the density of live cells (#CFUs/ml) that was present in the culture at each time point.

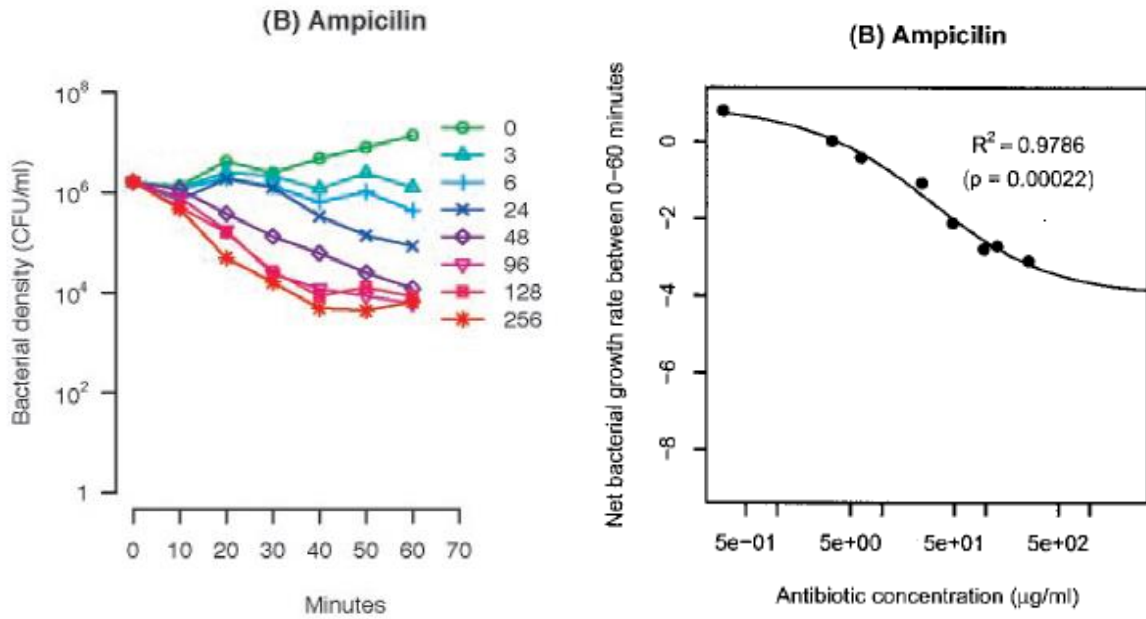
There are several issues with the use of CFU time dynamics for the study of the effect of demographic stressors (especially those that induce killing) and of the evolutionary response to such stressors. I detail them below.

[2. Time kill-curves and pharmacodynamic functions](#)

One classic application of CFU time series is in pharmacology, where one wishes to measure the effect of a given antibiotic (**AB**) at a given concentration c on the dynamics of the density of live cells. Let $n(t, c)$ be the density of live cells in the culture at time t and concentration c . The observed dynamics of $n(t, c)$ (live cells vs. time) are called time-kill curves. Typically, one seeks to fit a single constant growth rate $\overline{n'(t, c)/n(t, c)} \equiv r(c)$ to the time kill curve and then plots the relationship between growth rate $r(c)$ and AB dose c .

This relationship is the central characterization of the effect of an AB on a given bacterial species or strain within species. The function $r(c)$ is often called the “pharmacodynamic function”. Pharmacodynamic is the part of pharmacology that describes the biological effect of a given drug depending on its concentration. In the case of anti-infectious drugs (antibiotics, antifungic, anticancer, antiviral) the key biological effect of interest is the pathogen's growth rate (it's population dynamics). The minimum inhibitory concentration (**MIC**) is the dose at which growth is exactly stopped ($c_* | r(c_*) = 0$) and the overall function allows to predict the full dynamics of the pathogen as the drug doses changes (due to ingestion and elimination) in the place of application (pharmacokinetics), e.g. the blood plasma, an organ etc.

Below is a typical measurements of pharmacodynamic functions (REGOES *et al.* 2004), here the example on Ampicilin which is the AB I have been using the most so far. A set of time kill curves is obtained by regular CFU counts in vials with various drug concentrations (left panel). From this, an early net growth rate $\overline{n'(t, c)/n(t, c)} \equiv r(c)$ is estimated and plotted against drug dose (right panel):



The fitted relationship in the left panel is a Hill - Function which was introduced in this paper, mostly based on phenomenological observation, and became the standard function used for this purpose since. Citing the original paper:

Modeling the relationship between bacterial net growth and antibiotic concentration. As our model of the net growth rate of the bacterial population, ψ (which can be positive or negative), exposed to an antibiotic concentration, a , we assume the following relationship:

$$\psi(a) = \psi_{\max} - \mu(a) \quad (1)$$

Hereby, ψ_{\max} is the growth rate of the bacterial population in the absence of antibiotics, and $\mu(a)$ is the death rate of the bacterial population exposed to an antibiotic concentration a , which we assume to be a Hill function:

$$\mu(a) = E_{\max} \frac{(a/EC_{50})^{\kappa}}{1 + (a/EC_{50})^{\kappa}} \quad (2)$$

E_{\max} designates the maximum antibiotic-mediated death rate; EC_{50} is the antibiotic concentration at which the death rate is at half of its maximum, $E_{\max}/2$; and κ denotes the Hill coefficient, which is a measure of the steepness of the sigmoid relationship between μ and a . All the rates above [$\psi(a)$, ψ_{\max} , $\mu(a)$, and E_{\max}] are measured as the change per hour in the logarithm of bacterial density to the base 10.

To me, it always seems quite elegant that in spite of the fairly blunt technique used to estimate $r(c)$ (mean over the first 60 minutes here), the observed functions are consistent with a simple pharmacodynamic functions (in all AB studied, the Hill - function gives a good fit to $r(c)$). Yet, it remains obvious (even for other AB in this paper) that the observed CFU dynamics cannot be summarized as a single $r(c)$

(straight lines for each dose), so that there is in fact a pharmacodynamics during an *in vitro* time – kill curve itself. I also think that the underlying dynamics of live and dead cells are more complex than what is pictured in these CFU count dynamics, but that the very use of CFUs might in fact blur the pattern. I explain why below.

3. General issues with CFU time series dynamics

1/ The obvious issue of CFU counts is logistic: many samples must be taken in several culture vials, then many petri dishes must be plated and incubated and counted. Furthermore, a range of dilution factors F must be attempted for each sample (several logs) unless we already know the order of magnitude of the population density that we are seeking to measure. With F too high there are so few CFUs to count that the measure is unreliable, with F too low, confluent colonies will arise instead of clear isolated CFUs.

I have acquired a spiral plater to overcome the dilution and counting issues: this small “robot” pours the culture sample on the surface of the round petri dish as a spiral, at a constant rate. This creates a continuous dilution gradient from the center towards the rim of the dish. After incubation, colony counting is performed on photos of the dish (taken by a dedicated illumination platform), the analysis software seeks the zone of dilution (along the spiral) where counting is optimally reliable, then corrects for the dilution along the spiral to find the correct density. Yet, even with this nice apparatus, there are still many samples to take in the system, many agar dishes to prepare, plate and incubate.

2/ Another general issue of the CFU time series system is that it requires to open the system to sample in it, which may or may not be problematic. For example, when using a volatile compound in the culture, or when studying anaerobic growth conditions.

These are general issues but there are particular issues that apply when studying the effect of stressors on the dynamics of live and dead cells, which is my case and probably the case of most people studying live cell density dynamics.

4. Special issues when studying killing factors

3/ There is a particular issue when using CFU counts on a petri dish containing a killing factor. This is typically done to characterize the frequency of mutants that resist a given treatment, which is applied to the petri dish itself, in the widely-used Luria-Delbruck experiment. The probability to form a colony from a single cell corresponds to the non-extinction or explosion probability of the stochastic birth-death ('**BD**') process occurring for each cell on the petri dish. In standard media with cells well adapted to the lab, there is little death occurring during. The growth process is a pure birth process and explosion is certain (although it will take variable times between cells): all live cells present will form a colony. However, if the medium on the dish induces some killing

(e.g. a drug), then the probability that a live cell will yield a colony (an exploding and large population) is less than one. For example, it is equal to the ratio of birth/death rate for the linear BD process. This ratio may further depend on the drug concentration.

⇒ Direct CFU counts are not reliable if the petri dish medium induces some killing.

To correct for this, one must know how the medium on the dish affects the birth/death rate ratio of the strain. This can actually be used to test stochastic BD process models.

4/ There is a particular, and I think underestimated, issue when using CFU time series to measure the effect of a treatment that enters the cells and induces some killing (a virus or a drug for example). Say the medium on the dish is now standard: dilution of the past medium upon sampling will have removed most stress factors, or one can use an additional treatment on the petri dish to destroy this factor e.g. (betalactamase to destroy the betalactam AB). Still, the cells sampled might keep a memory of the conditions they were facing just before sampling, during the treatment. They may have accumulated defects due to the drug, be infected with the virus etc., and these defects do not vanish instantaneously once plated on a permissive medium. They may alter the birth and death rate of the cells, at least during the first few generations, which is when explosion vs extinction is decided. Thus, a cell may be doomed to die ultimately after $t = 10$ minutes of exposure, but still be perfectly alive by this time. If so, the CFU counting technique will count doomed cells as dead cells, not reflecting the true dynamics ongoing by time $t = 10$.

⇒ Direct CFU counts are not reliable if there is some memory effect of the cell's past environment before plating, that induces killing.

To correct for this, one must know how the memory of the past environment will affect the BD process on the petri dish, and how this memory will vanish or not during growth on the dish. Again this can actually be a way to test models of such memory effect, of drug action etc.

5/ Finally, by definition CFU counts only count live cells, and give no information on the density of dead cells.

For all these reasons (but mainly the logistic one) I have had to develop my own system to measure high – replication live and dead cell dynamics, which ended up being fluoro-luminometry (**FLU**).

[5. Overcoming part of the logistic issue by the Start Growth Time \(SGT\) method](#)

To measure live cells dynamics without as much logistics as CFU count, I tried another technique that uses the same principle as quantitative PCR. After we had started to experiment the technique, we found out that it had been proposed, validated and even had a name: the set growth time (**SGT**) method (HAZAN *et al.* 2012).

Several wells of a microplate culture (source plate) are monitored by regular sampling over time. However, instead of plating them (several replicate petri dishes

per well x time point), the samples are simply diluted into another microplate (SGT plate) containing standard medium, which is refrigerated. At the end of the sampling phase the source plate the SGT plate contains replicate samples of the original wells at different times. A calibration is added to the SGT plate (a set of wells containing known cell densities). The whole plate is then incubated and monitored for growth by any means available (in our case by fluorescence of cells expressing CFP). The waiting time to reach a given signal after the start of growth is related to the initial cell density.

Let r be the growth rate of the strain used in the experiment, assumed fixed over time and between wells of the SGT plate (standard medium, no evolution at the timescale of the SGT experiment). If the density of the initial sample in well i is not too small ($n_i(0) \gg 1$), growth is exponential (after correcting for some measurable lag t_L): $n_i(t) = n_i(0) e^{r(t-t_L)}$. The waiting time τ_i to reach some threshold n_F ($n(\tau) = n_F$) is

$$\tau_i = t_L + \frac{1}{r}(\log n_F - \log n_i(0))$$

Allowing to infer $n_i(0)$ and the corresponding density in the source plate. This is really like qPCR applied to cellular growth instead of DNA copying by polymerases.

The method avoids plating, incubating and counting many petri dishes, and allows a handy microplate- to – microplate sampling. It proves fairly accurate based on the calibration curves (below a figure from Timo Van Eldijk's M1 report in 2017)

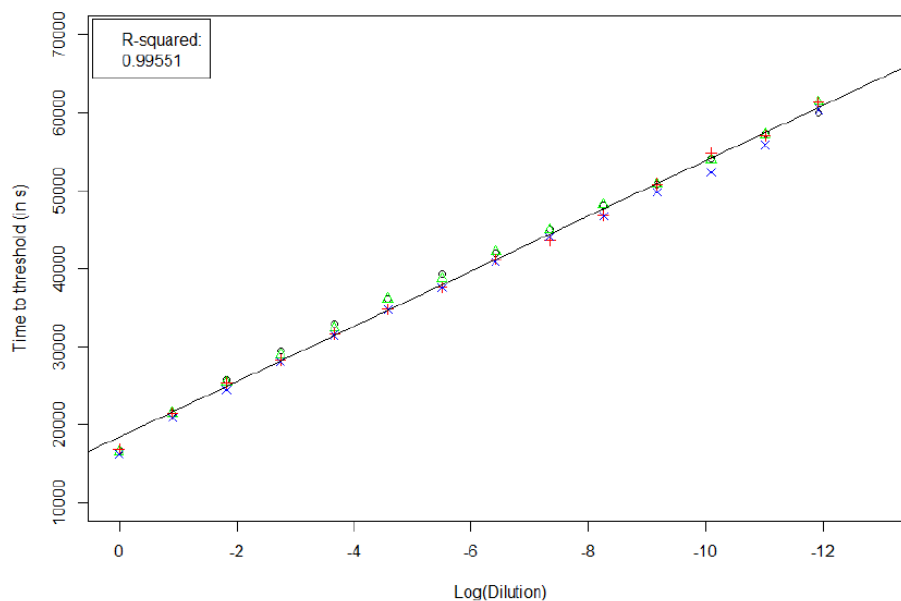


Figure 4. Time to threshold (fluorescence= 8000) plotted against the logarithm of the dilution level of the original well. Replicates shown with different symbols with different colours. All fourteen dilution levels are shown.

We used it to produce a set of time kill curves across doses of Ampicillin (equivalent to the figure above from Regoes); the figure is again from Timo's report

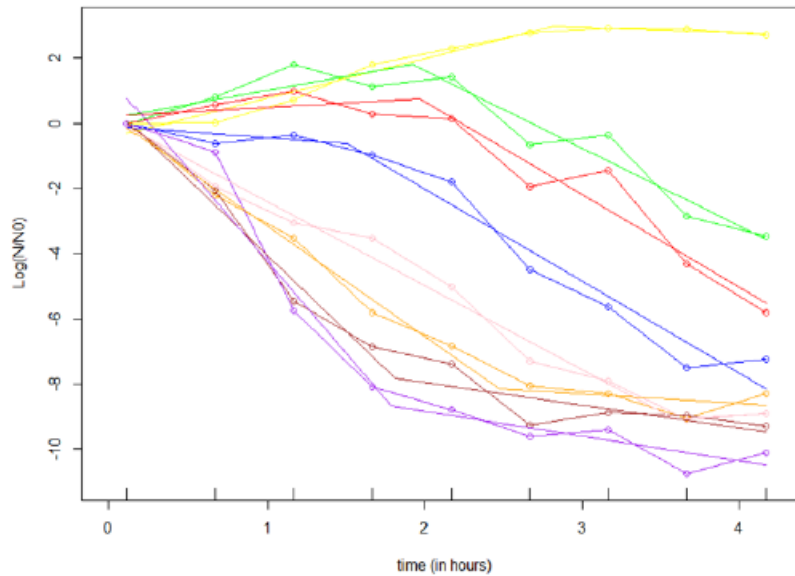


Figure 10. The scaled live cell concentrations over time as obtained from the SGT data from experiment 1A. Each series of coloured points connected by a spline represent a single dose of ampicillin (1 replicate, 7 doses and control). Segmented linear regression lines consisting of two segments are also shown. The following doses (in $\mu\text{g/ml}$) are shown: Control (light yellow), 1 (green), 2 (red), 4 (blue), 8 (pink), 12 (dark yellow), 16 (brown) and 20 (purple).

The SGT method reduces the logistic burden compared to CFU counts, but is still a fairly heavy experiment (it is hard to monitor more than a few tens of populations simultaneously). It also shares the same other issues as CFU dynamics, so we have only used it as an independent means to measure time kill curves so far.

6. [Microfluidics](#)

An alternative to study fine-scale stochastic dynamics of birth and death processes is to use microfluidics with time-lapse microscopy. This is the technique that allows the best precision with respect to the detailed single cell-level stochastic events that happen during cell life. For example, it has allowed to empirically explore the modes of cell elongation and division in colibaciles (yielding the now quite established 'adder model') which I use in a later model (WITZ *et al.* 2019).

However, microfluidics is not a technique that is cheap, or easily implemented in a lab from scratch (at least that is what I have been told). Furthermore, it is not always easy to implement various ecological scenarios into a microfluidics context: for example, it is unclear how phage infections would behave in a tight microfluidic environment, predation by larger unicellulars would most certainly be impossible, non-model bacteria may not grow easily on microfluidics environments etc. This would impede several of my empirical projects in the mid term.

More generally, it is my hope to test the validity of simple techniques (like OD measurements) using the experimental system and models that I develop. The idea being that we may be able to tell a wider community which kind of simple technique is reliable for which question. If my measures were to rely heavily on elaborate microfluidics, I could not test simple techniques in parallel in the same conditions. On

the contrary, FLU is simple to implement (once one has a marked strain) and is performed in classic microplate wells, where a lot of bacteriology is done (OD measurements) so we can study conditions that are typical of a simple microbiology lab. This being said, I would obviously love to find collaborators expert in microfluidics with whom to set experiments and test my models in some conditions (eg antibiotics are a priori amenable to the microfluidics context).

7. Fluoro-luminometry

In the end I elected for the monitoring of triply marked cells, both fluorescent and luminescent.

Live cells = Luminescence. Luminescence is an active process by which the cells produce light by consuming ATP, it is catalyzed by the lux operon (lux CDABE in our case). Light can only be produced by live cells with enough ATP available, it has been used before as a viability marker to measure the effect of AB on bacteria (TENHAMI *et al.* 2001). Because it is an active light production process it has a high signal to noise ratio allowing precise measures of even small populations (e.g. KISHONY AND LEIBLER 2003). One drawback of luminescence is that the signal per cell may change with physiology and conditions (lower/higher ATP per cell).

Total biomass = Fluorescence. On the other hand, fluorescence is catalyzed by a single gene (in our case Cyan Fluorescent Protein or **CFP**), and is a passive process: the marked cells express fluorophores on their membrane, that are excited by a narrow range of photon wavelength and emit another narrow range of photon in return. This is more like a mirror reflecting light in a given color. Because it is passive, fluorophores emit light upon excitation even within dead cells. Unlike luminescence, fluorescence is a less precise measure (lower signal/noise ratio) in general but is much more robust in that the signal per unit of biomass is fairly stable, even over 120h kinetics.

After making my first attempts and reading some bibliography I realized that this method had also been proposed and published (LEHTINEN *et al.* 2003) by a finish team, and had a name: fluoro-luminometry (FLU). Oddly enough (in my opinion), it has not been widely used since. I had the chance to meet, in Turku in 2018, the senior authors of this paper (Esa-Matti Lilius). He was a gold mine and a very fit and active 75 year old scientist. I later invited him to visit in Montpellier, he came all the way from Finland, with his neighbor, in his camping car, and went back via Munich to attend the Oktoberfest !

Dead cells = SYTOX. Because we are interested in measuring live and dead cells, I have updated the technique to add a third marker: a dead cell fluorescent stain. There are various such stains, the famous BacLight© system is based on such a staining protocol (green fluorescence for live cells and red for dead ones). However, most of the dead cell stains are actually toxic to live cells, which is problematic for kinetic monitoring. Of all the stains we tried (again Timo's M1 work), only the SYTOX Orange

and SYTOX green have proven non toxic and efficient (other SYTOX colors did not). I use SYTOX orange because it is spectrally more distinct from the CFP. These dead cell stains are small molecules that fluoresce upon binding to DNA. However, SYTOX cannot enter a cell with uncompromised membrane permeability and so never marks 'live cells'. It will enter cells that have compromised membranes and fluoresce upon binding to their DNA.

Optical density = live cells ? The most widely used method to measure cell dynamics is optical density (OD) so it seems natural that we compare our observation to this observable, that is often the only one available. In principle, OD is interpreted as live cells, but that may not be the case. In cases where death causes a burst of the cell however (lysis) it may be a good approximation. This is what we intend to test on two killing factors that lyse the cells: ampicillin and phages.

To summarize, luminescence measures 'live cells' in the sense of having enough available ATP and SYTOX measures dead cells in the sense of having compromised membrane. Fluorescence measures biomass in general. Our strain is called 10KClux because it is the 10 000 generations evolved strain from Lenski's LTEE (10K) and carries a CFP and lux operon.

B. Applying FLU

1. [Estimating various rates of the Birth Death process](#)

I start with a short theoretical argument on how we interpret the data from our FLU experiments. Let us see how different quantities characterize the birth death (BD) process over time and for a given level of a treatment (say a drug concentration c). These are approaches I intend to use on the FLU dynamics (future and existing ones) but that have not been applied yet.

Assume a simple linear BD process, the observation of the live cell density alone does not yield an inference for the birth and death rates separately. Denote the birth (resp. death) rate $b_c(t)$ (resp. $d_c(t)$) at time t in a vial with drug concentration c , and denote $N_c(t)$ (resp. $D_c(t)$) the live (resp. dead) cell densities, and denote $B_c(t) = N_c(t) + D_c(t)$ their total which is the total biomass. The dynamics (expected dynamics over replicates or exact dynamics in a deterministic limit) of these three quantities are:

$$\begin{aligned} \dot{N}_c(t) &= (b_c(t) - d_c(t)) N_c(t) \\ \dot{D}_c(t) &= d_c(t) N_c(t) \\ \dot{B}_c(t) &= b_c(t) N_c(t) \end{aligned} \quad , \quad [1]$$

where $\dot{X} = \partial_t X$ is the time derivative of a variable. We can see that live cell densities alone only allow to estimate the growth rate, but not birth and death rate separately.

This further requires measuring dead cell densities and/or total biomass. More precisely, given an observation of $\{N_c(t), D_c(t), B_c(t)\}$ we can estimate :

$$\begin{aligned} r_c(t) &= b_c(t) - d_c(t) = \dot{N}_c(t)/N_c(t) \\ d_c(t) &= \dot{D}_c(t)/N_c(t) \\ b_c(t) &= \dot{B}_c(t)/N_c(t) \end{aligned} \quad , \quad [2]$$

Measuring these three quantities thus allows a validation of a given model of BD vs time and concentration: if the model is valid and the measurements accurate the three quantities should be consistent. Furthermore, it allows to estimate $\{b_c(t), d_c(t)\}$ which is of importance because the probability of stochastic extinction when rare (which is critical for the fate of rare mutants) depends on these two quantities separately. For a linear BD process with arbitrary birth and death rates over time, the probability to be extinct by time t starting from a single copy at $t = 0$ is well – known (KENDALL 1948): it can be written

$$\begin{aligned} p_0(t) &= \mathbb{P}(N_c(t) = 0 | N_c(0) = 1) = \frac{\int_0^t d_c(s)/n_c(s) ds}{1 + \int_0^t d_c(s)/n_c(s) ds} \\ n_c(t) &\equiv \mathbb{E}(N_c(t) | N_c(0) = 1) = \exp\left(\int_0^t (b_c(s) - d_c(s)) ds\right) \end{aligned} \quad , \quad [3]$$

where $n_c(t)$ is the expected density of live cells at time t , starting from a single copy.

With a density-independent growth model, as is expected away from stationary phase, we have $\mathbb{E}(N_c(t) | N_c(0) = k) = k \mathbb{E}(N_c(t) | N_c(0) = 1)$ so that $n_c(t)$ can be estimated from observations of large (deterministic) population size dynamics (at the same concentration c): $n_c(t) = N_c(t)/N_c(0)$. This might be another way to analyze the BD process key quantities such as $p_0(t)$ in the future.

Finally, total biomass gives insight into the relative contribution of de novo vs. preexisting mutations to the ER process (preexisting meaning before the onset of stress). The total amount of mutations produced over a period is Poisson with rate equal to the cumulative number of birth events (with instantaneous rate $b(t) N(t)$). Therefore, consider a preculture is grown in the absence of stress up to a time T ($c = 0$ when $0 \leq t \leq T$). The number of mutations it has produced is Poisson with rate $\Lambda_0 = \int_0^T b_0(s) N_0(s) ds = \int_0^T \dot{B}_0(s) ds = B_0(T) - B_0(0)$: the difference between final and initial biomass at the onset of stress. Typically, the preculture is started as one or a few cells so $\Lambda_0 \approx B_0(T)$. Similarly, the number of (one step) mutations produced by the dominant wild-type after the onset of stress ($c > 0, t \geq T$) up to extinction (when extinction happens in at least some replicates) is $\Lambda_c = \int_T^\infty b_c(s) N_c(s) ds = \int_T^\infty \dot{B}_c(s) ds = B_c(\infty) - B_c(T)$.

Overall, the relative contribution of post-stress and pre-stress mutations to the whole ER process (Λ_0 and Λ_c) can be estimated directly from biomass measures: $B_c(\infty) | N_c(\infty) = 0$ at extinction among extinct populations and $B_0(T) = B_c(T)$ at the onset of stress. Of course, this ignores the fact that some extinct populations may stem not only from the decay of the wild type alone but also from the contribution of rare less subcritical mutants. However, in our experiments and simulations, there is a rather clear divide

between repeatable extinction trajectories and variable ER trajectories so this should not be too problematic.

Another approach we have started to develop is a less phenomenological one, where birth and death rates are explicitly related to the actual stress nature and level c via a semi-mechanistic model. I do not expose these here as it is too early in their development, but the idea is to develop a complementary more precise but less general version of the ER process, that includes some explicit mechanism (i.e. models of antibiotic action or phage infection for example).

Let us now turn to the raw empirical patterns we have observed so far via FLU. All examples are done in either 96 or 24 well plates with glucose rich Davis medium at 30°C and strain 10KClux, with a culture started in early log-phase (after lag, before stationary phase).

2. Calibrations

These are calibration curves (Timo's M1 report) for luminescence (figure 1) and CFP fluorescence (figure 2) against various dilutions of a log-phase culture.

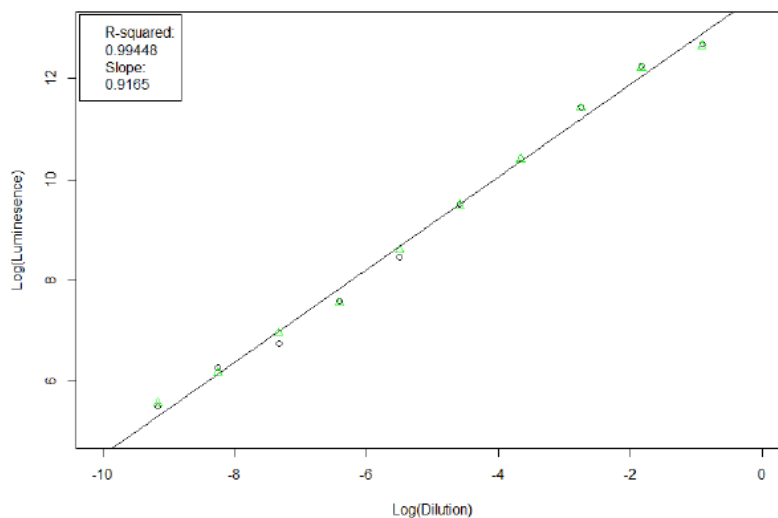


Figure 1. Logarithm of the luminescence plotted against the logarithm of the dilution level. Replicates shown with different symbols with different colours. Only the linear range of dilutions (factor 0.4) is shown.

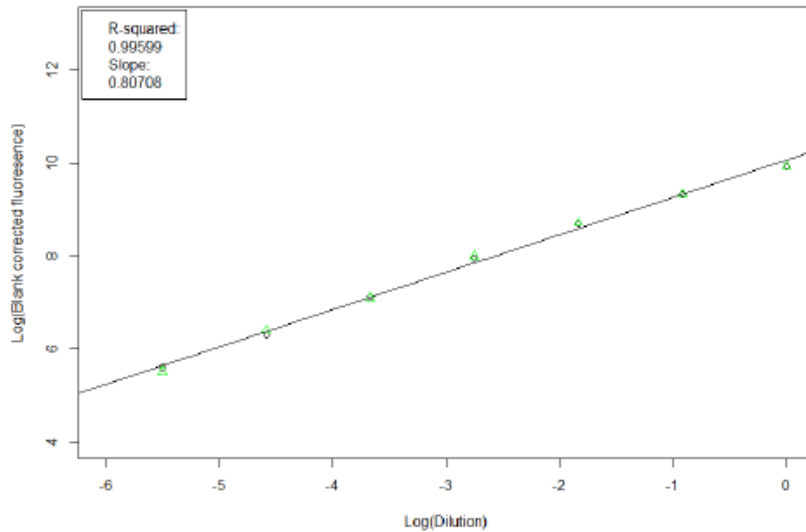


Figure 2. Logarithm of the blank corrected fluorescence plotted against the logarithm of the dilution level. Replicates shown with different symbols with different colours. Only the linear range of dilutions (factor 0.4) is shown.

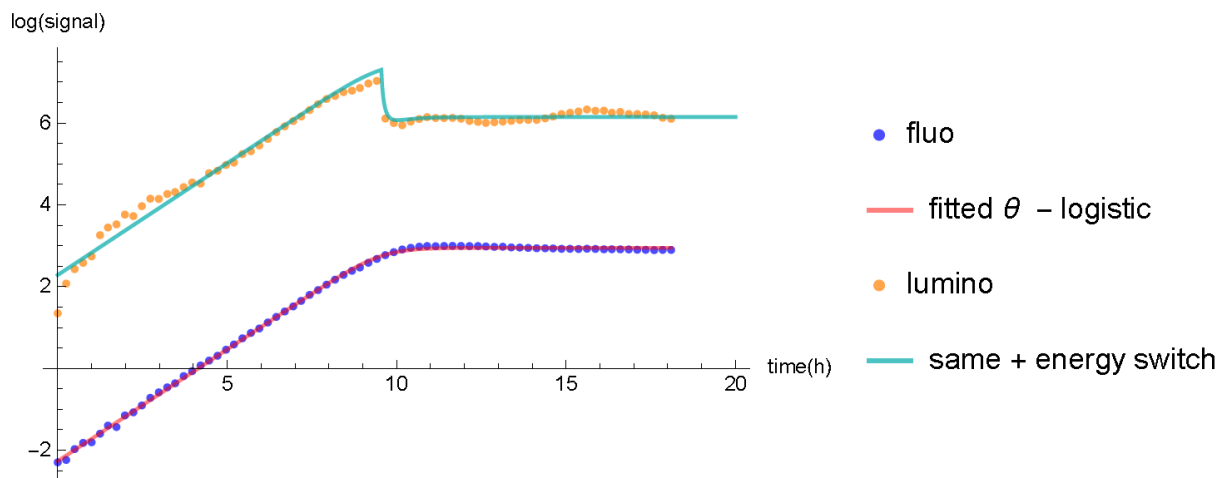
We can see that there is a dynamic range of several logs were both measurements are log-linearly related to cell density.

3. Controls in the absence of stress

Below I illustrate control curves (in the absence of any stress) for CFP fluorescence signal and luminescence with our strain 10KClux, in a 96 well microplate (200uL) agitated at 30°C, starting in log-phase (no lag). The fitted model is a θ -logistic:

$$N'(t) = b N(t) \left(1 - \left(\frac{N(t)}{K} \right)^\theta \right)$$

The fitting routine and cleaned data were kindly provided by my postdoc Yoann Anciaux:



The red curve is fitted to the fluorescence data, and yielded: $\hat{b} = 0.54, \hat{K} = 19.04, \hat{\theta} = 3.45$. The green curve is not fitted, it is the same θ -logistic (same θ parameters) with a different initial value. This shows that the lumino and fluo data are perfectly consistent with each other during the growth phase in controls (as expected if there are no deaths in standard medium: $r_0 = b_0$ at $c = 0$). The late dynamics in the lumino signal is a model of shift in energy per cell as the population reaches near stationary phase. This additional shift is fitted to the lumino data (shift time, rate of decay in energy per cell and ultimate energy per cell if the log phase energy is 1). This shows that although lumino signal is tricky to analyze in stationary phase, it is fairly reliable and consistent with fluo in log phase.

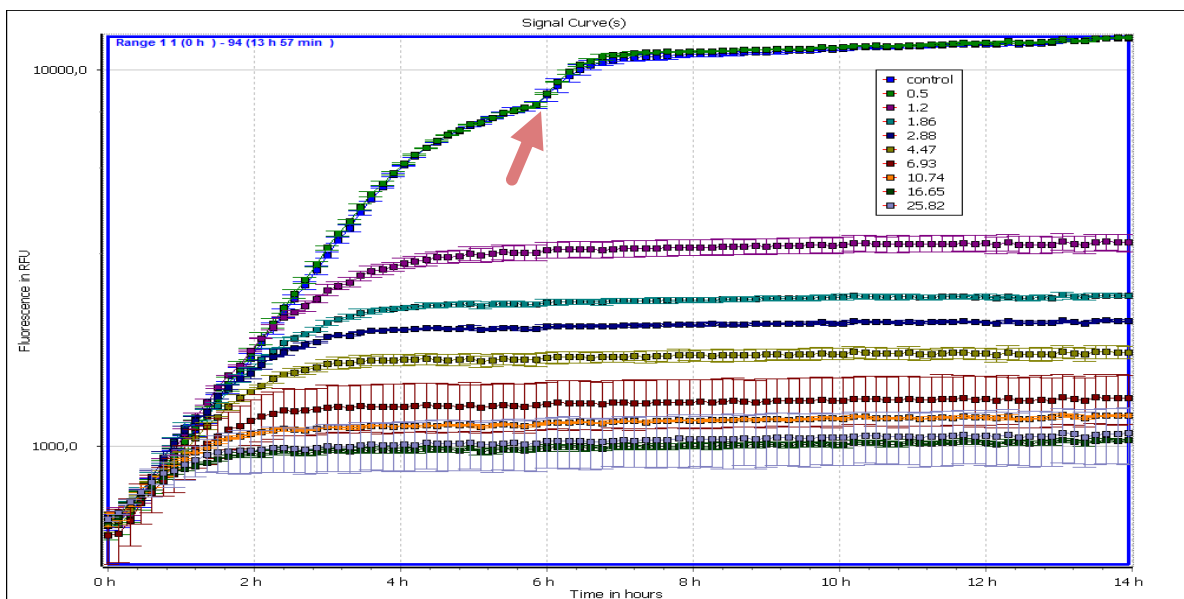
The θ -logistic with a $\theta = 3.45$ is also indicative that density dependence is highly negligible in our system until the population reaches very near stationary phase.

I now turn to example of kill curves and ER in the face of ampicillin measured by FLU system.

4. Kill curves under Ampicillin stress

Here are examples of recent dynamics that we got for live (lumino), dead (SYTOX) and total (CFP) cell densities at various doses of ampicillin. These data are averages from 2 replicates of dynamics in 24 well plates (for higher precision), the observation is similar in 96 well plates with slightly more noise and an apparent late diauxic growth pattern in the controls. I do not have the SYTOX data for this example as we had trouble with our sytox batch. In any case I show these to illustrate the kind of data we can observe. These are raw blank corrected data from the plate reader. The doses ($\mu\text{g/ml}$) of ampicillin are indicated in legend, they are increasing by a constant factor ($\times 2.4$).

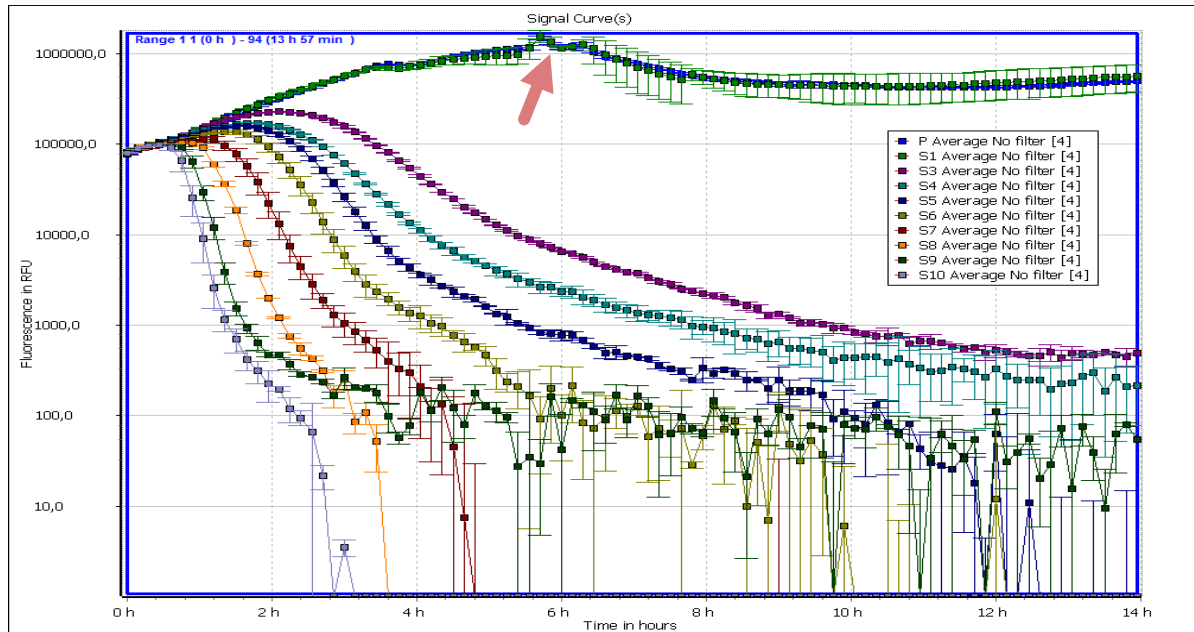
CFP fluorescence: total biomass dynamics



We get very repeatable trajectories with a regular pattern as the dose increases (less and less accumulation of biomass). The control and lowest dose are undistinguished.

Note a slight difference with the 96 well plate data (previous part): there is a second increase in biomass before full stationary phase (red arrow). Our best guess is that this stems from diauxic growth with the late use of another resource. This being very late in the growth period it should not affect the kill curves at substantial doses but it remains to be interpreted.

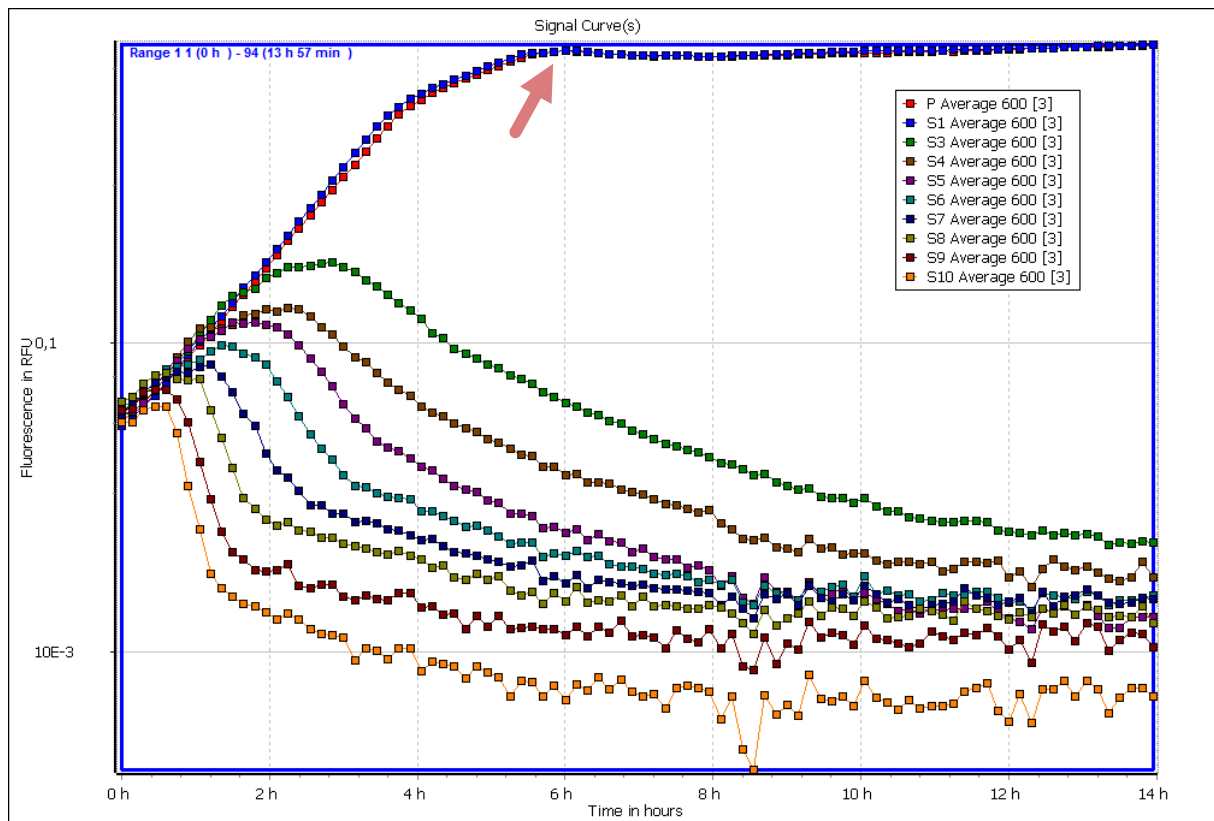
Luminescence: live cells dynamics (the curve label correspond to the same increasing doses from S1 to S10 and P is positive control (dose 0)).



The pattern is quite consistent (visually) with the CFP data, with a decay in live cells earlier and faster for higher doses. Again replicates are very similar (narrow error bars with just two replicates) suggesting that a deterministic model should perform well on these data.

The time corresponding to the supposed "diauxic growth" switch is indicated on the graph, it does correspond to a shift in the luminescence patterns of the controls.

OD: 'live cells' density potentially



This illustrates that, in the case of betalactam AB that induce cell lysis, we may hope to reliably measure live cells with just OD, as the curves are very similar to those from luminescence. One note however: the decay does not go to zero here, which is likely due to the fact that lysed cells still block the light, although less so than when they are in normal shape.

In all three datasets we see that as the concentration becomes very high the curves tend to show limited change, suggesting some saturation pattern with drug dose.

Some verbal argument now:

At very high concentration there is still a fixed delay before the AB seems to be effective, a minima about an hour delay after the addition of AB (time 0). The interpretation of this fact has been quite challenging, I think I have now an assumption for this effect, but it requires a detailed modelling exercise that I will not do here. The main idea is that the effect of the AB follows the causal chain

External AB -> target enzyme -> molecular phenotype -> effect on BD rates

For example, in the case of ampicillin (a betalactam AB) the target enzymes are the Penicillin Binding Proteins. Once bound to Ampicillin (~ irreversibly) they are inactivated and unable to connect glycan strains within the cell wall, the molecular phenotype is thus the proportion of connections between glycan strains. This proportion will in turn affect how cells divide (vs. die upon dividing) and possibly other aspects of cell survival and elongation.

In general, even if the drug is highly concentrated, the molecular phenotype that ultimately causes the decay may take time to build up, in the case of ampicillin it is the accumulation of missing glycan links along the cell wall, which likely occurs at the

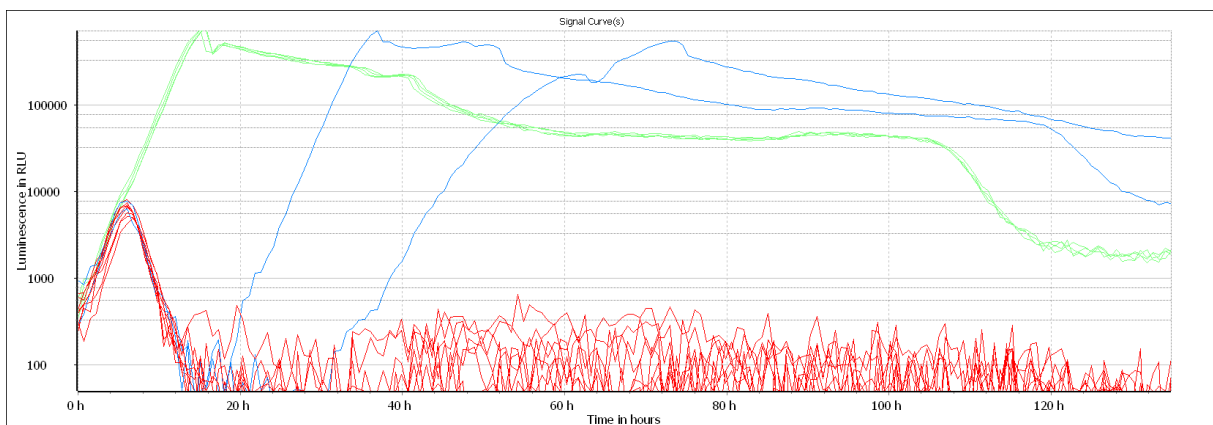
timescale of a cell's elongation and division, hence of the order of the hour. That this pattern differs from the observed kill curves from CFU counts (see above Regoes data) might not be inconsistent. The cells may have fully saturated their target enzymes with drug after ten minutes, still the molecular phenotype is not apparent (in lumino), but most of these cells are doomed and will not form a CFU after plating, so the decay arises earlier on the CFU count data.

Finally, the decay of live cell density (in both lumino and OD data) seems to slow down late in the kill curve. This is sometimes interpreted as 'persister' phenotypes. It will have to be studied in detail as it may affect the likelihood of long term ER emergence.

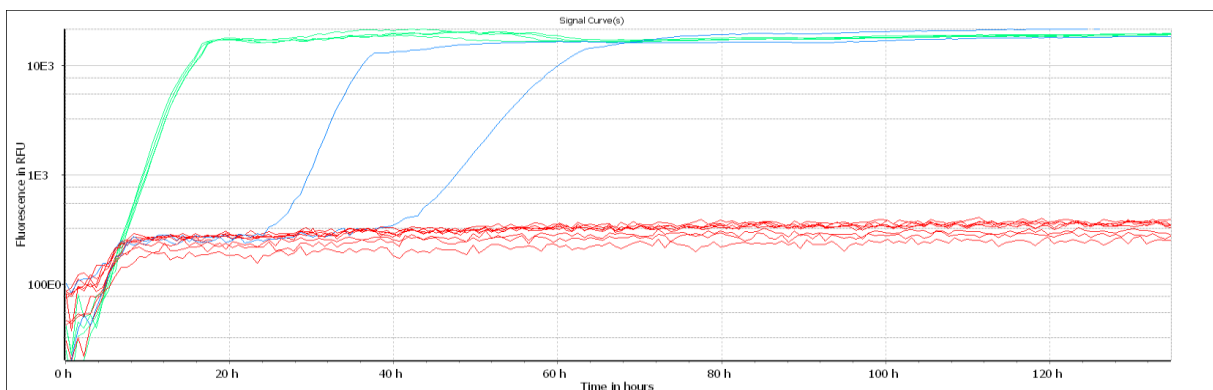
5. ER under ampicillin stress

I now show an example of a longer-term 'ER' experiment where we incubate the cultures for more than 120h. I picked a case for a given dose and a given plate of one such experiments (Julien Vadez's M1 work in 2019).

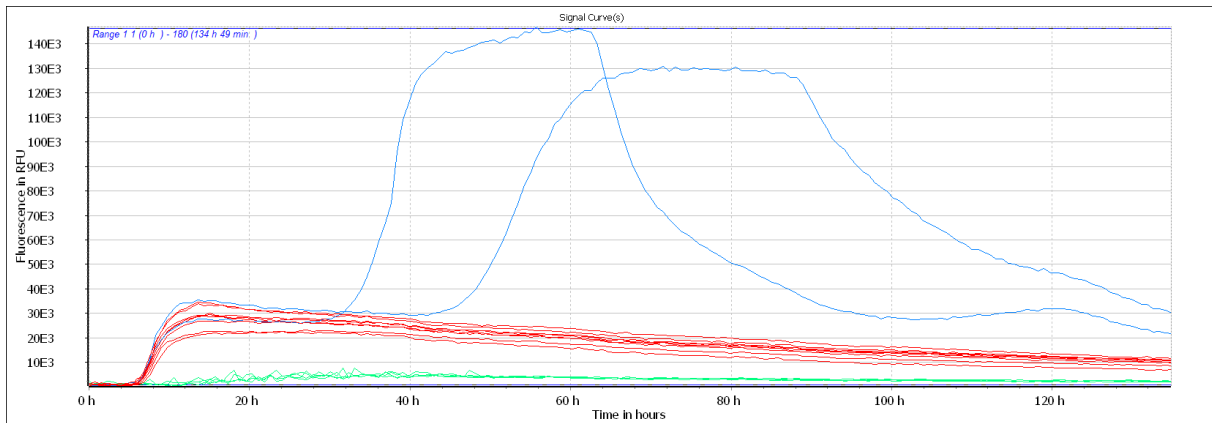
ER in lumino: 2 ER trajectories (blue) 9 extinct (red), 4 controls (green).



ER in fluo: the corresponding curves in terms of CFP signal



ER in sytox: the corresponding SYTOX signal.



There is a clear consistent pattern between the three signals, the two ER trajectories appear to have distinct birth and death rates (distinct live and dead cell density profiles), but that would require proper quantitative analysis. At least it is apparent that the ER lines do not simply avoid death but they grow in spite of substantial death (additional accumulated dead cells).

Note that there is a later drop in sytox signal quite long (20-30h) after the ERs have reached stationary phase. This is likely due to the destruction of DNA molecules by DNAases leaked from dead cells after some time, so that the binding and resulting fluorescence of sytox is lower. Indeed, this late decay could be impeded by the addition of EDTA which is a DNAase inhibitor.

These data are just illustrations of the potential for later analysis with future student (starting with my current postdoc). We also have performed similar experiments with other stressors, in particular lytic phages but I do not detail this here.

C. *Other Experimental projects*

I quickly mention two extension of the system that I wish to implement in the years to come.

1. Measuring fitness distributions over time

Most experimental evolution document only mean fitness trajectories while a lot of information lies in the distribution of fitness over time. I wish to measure these distributions by performing high dilutions of the cultures (this is easily done by pin-replication) then regrowing the resulting diluted samples. The proper statistical treatment of such an experiment is not straight forward, as it requires to implement the effect of early birth and death events on the asymptotic growth pattern. However, BD process theory provides the require elements for this task.

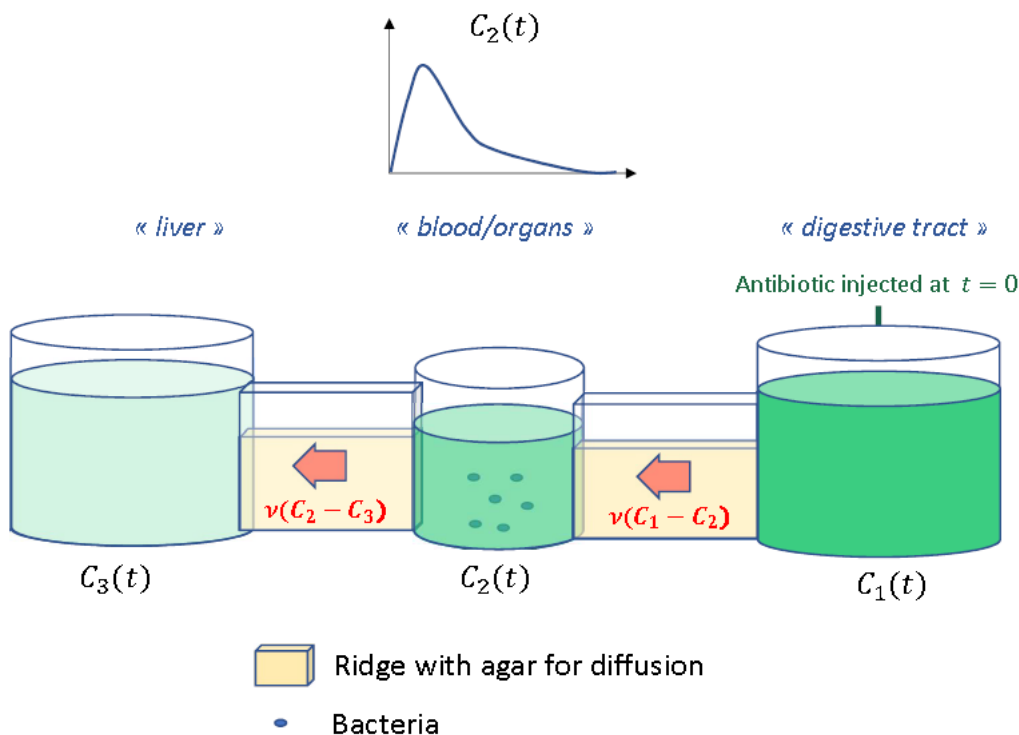
2. Mutation accumulation experiment

In a similar vein, mutation accumulation (MA) experiments will need to be implemented, to measure key mutational parameters (rate and effect distribution). Using the same liquid-based techniques (high regular dilutions) I hope it can be possible without too much burden. However, here too, similar BD modelling will be required to predict the outcome of serial dilutions and growth over substantial periods of time. The classic MA analysis framework does not include these details as it has been tailored for large multicellular organisms where single individuals can be picked every generation.

3. 3D printed microplates to reproduce simple pharmacodynamics

I would like to implement a more complex system of pharmacodynamics in microplate, to be able to follow more realistic dose trajectories over time, and to test models of continuously changing environments. My hope is to 3D print microplates that would allow diffusion of an AB from one well to another to mimick pharmacokinetics (increase and decay in drug dose over time).

The design is illustrated below:



The three adjacent wells are connected by a small ridge filled with agar through which the AB diffuses but bacteria cannot cross. A source ($C_1(t)$) sends AB to the well with bacteria ($C_2(t)$), and this well is depleted of its AB by diffusion to a dump well ($C_3(t)$).

The goal is to keep the handy FLU + microplate setting, to be able to follow replicate time dynamics, while allowing for semi-controlled changes in dose regime.

VI. Theoretical approach

A. Context and state of the art

I now present a theoretical framework that I have been developing for some years now. Because this method sets the dynamics of the system into particular forms of PDEs, I went to discuss this at a workshop (in Lyon, 2015) where I met Lionel Roques (from INRA BioSP in Avignon), and he has been my main collaborator ever since. The idea behind this framework was to fill-in a gap in available theory of asexual evolution (the one most relevant to the data I am interested in). Let us first detail what approaches are available. For simplicity of the illustration, and because it is what has been most detailed, I only illustrate models of fitness dynamics (not multivariate sets of fitness components), in constant conditions (no continuous change in the environment), and without sexual reproduction (haploid asexuals) or migration (isolated population).

1. Origin-fixation models

The most popular tools so far to model asexual adaptation are so-called “origin-fixation models” (reviewed in MCCANDLISH AND STOLTZFUS 2014): one follows the expected outcome of the successive establishment of new mutations in the population. These models have seen important developments, especially regarding the complicated treatment of the fixation process in asexuals (GERRISH AND LENSKI 1998; DESAI AND FISHER 2007). Roughly put, an origin-fixation model considers the expected change in mean fitness as the stationary regime of a renewal-reward process: with a steady flux of new mutations destined to fixation (renewal) and their effect on mean fitness upon fixation (reward). Let \bar{m}_t be the mean Malthusian fitness of a genotype (the expected exponential growth rate of the genotype), if $K(t)$ mutations are present by time t then the mean fitness is $\bar{m}_t = \bar{m}_0 + \sum_{i=1}^{K(t)} s_i p_i(t)$ where the s_i are drawn into the distribution of segregating mutations and $p_i(t)$ are their corresponding frequency by time t . This assumes no epistasis as mutation effects simply add up on mean fitness. This process is stochastic ($K(t), s_i, p_i(t)$ are all random variables): its outcome can only be derived by framing it as a renewal-reward process. To do this, one shunts the frequency dynamics to only consider fixed mutations and their effects (indexed with a ‘*’, so that by the only mutations counted are those with frequency $p_i(t) = 1$). The mean fitness is simplified to $\bar{m}_t = \bar{m}_0 + \sum_{i=1}^{K_*(t)} s_i^*$ where s_i^* are now drawn into the distribution of fixed mutations and $K_*(t)$ counts their number. In fact, at intermediate times, the alleles are polymorphic, but by considering a steady regime of fixations this transient state can be ignored. The process remains stochastic (K_*, s_i^* are stochastic) but we can now compute the expected increase in the number of fixed mutations (substitution rate) and in the mean fitness (adaptation rate).

Figure 2 bellow illustrates the principle of origin-fixation models.

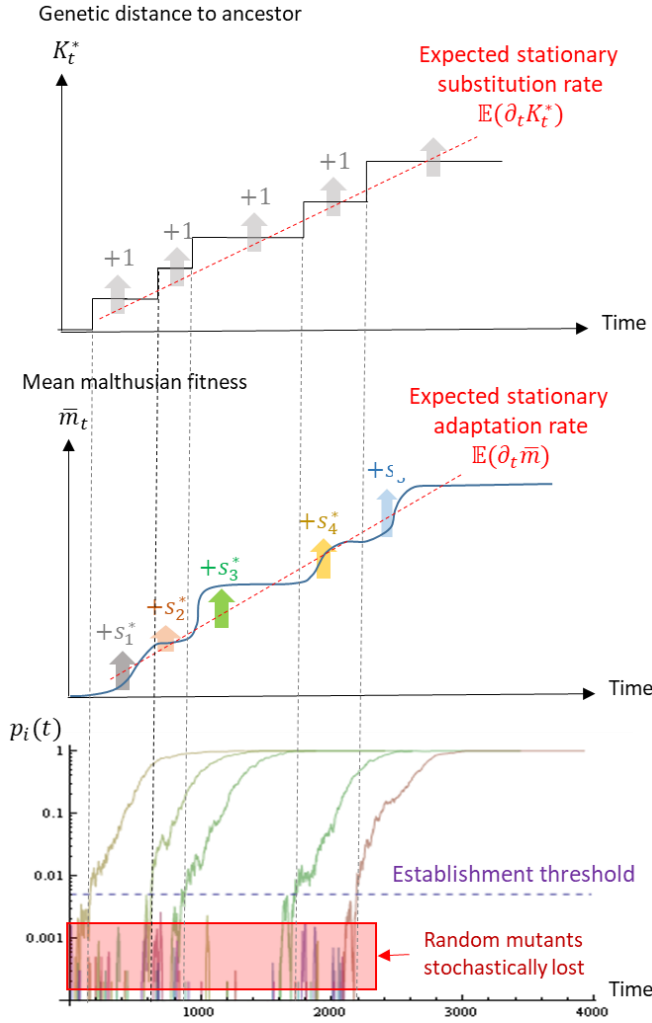


Figure 2: Schematic dynamics of genetic distance (K_t^*), mean fitness (\bar{m}_t) and mutant frequencies ($p_i(t)$ in log-scale) under an origin-fixation model. Mean fitness and genetic distance both rise by the successive establishment (probability $\pi(s_i^*)$) of new mutations with effects s_i^* drawn into a stationary distribution. Other mutations (including all deleterious ones) do not avoid stochastic loss when rare and are discarded from the computation. Here a very simple situation is considered with freely recombining loci so that mutations fix independently.

These expected rates are only computed at steady state, hence they are large time limits in constant conditions. From basic properties of renewal reward processes we have:

$$\begin{aligned} \mathbb{E}(\partial_t K_*(t)) &\xrightarrow{t \rightarrow \infty} NU \int_{s \in \mathbb{R}} \pi(s) f(s) ds \\ \mathbb{E}(\partial_t \bar{m}_t) &\xrightarrow{t \rightarrow \infty} NU \int_{s \in \mathbb{R}} s \pi(s) f(s) ds \end{aligned} \quad [4]$$

Here, NU is the inflow of new random mutations (at rate U per capita per unit time) for a population of fixed size N . $f(s)$ is the density function of the DFE among random mutations, and $\pi(s)$ is the probability of fixation of mutations with effect s . In fact, $\pi(s)$ has a complex (but often explicit) expression in asexuals, because co-segregating mutants hinder one another's fixation (while they sweep independently in freely recombining genomes). A *minima*, $\pi(s)$ depends on the population size, mutation rate, and the DFE itself $f(s)$ (GERRISH AND LENSKI 1998; GERRISH 2001). With larger mutation rates, it further depends on the fitness distribution of resident genotypes (that of $m - \bar{m}_t$), in travelling wave models (DESAI AND FISHER 2007; HALLATSCHKE 2011). To compute these complex expressions, the stationary regime assumption is critical

(waves travelling at steady speed, established mutants arising and rising in frequency at a stationary rate).

A useful property of origin-fixation models is that they provide predictions for both fitness trajectories (\bar{m}_t) and the changes in sequence space ($K_*(t)$). Indeed, with infinitely many alleles, $\partial_t K_*(t)$ gives the rate of substitution over the genome. This allows to compare the predictions to both fitness and sequence data (e.g. in WISER *et al.* 2013).

On the other hand, by construction, origin – fixation models require stationary regimes, which limits the handling of transient dynamics, and of epistasis or environmental change (which change the DFE over time). Some extensions have been proposed that use a form of quasi-stationary state approximation to handle epistasis: updating the DFE and thus the stationary fitness increase as the dominant background changes (KRYAZHIMSKIY *et al.* 2009; WISER *et al.* 2013; GOOD AND DESAI 2015). However, this approach is only accurate with a clear separation of timescales, when rare beneficial mutants sweep through the population very fast (KRYAZHIMSKIY *et al.* 2009). It cannot easily be implemented in the context of more polymorphic ‘waves’ of fitness (DESAI AND FISHER 2007), when there is not a single dominant genotype that would determine the current DFE.

Overall, although probably the most popular framework to model asexual evolution, origin – fixation tools are by construction not ideally suited to handle epistasis, or changing environments. Furthermore, they ignore the contribution from deleterious mutations (which have a negligible fixation probability). However, even without fixing, deleterious mutations can affect the mean fitness while segregating at low frequency in large numbers: the quantitative importance of this effect is unknown as deleterious mutations are typically not implemented in the simulation checks of origin – fixation models. Deleterious mutations are also the most frequent and most empirically available mutations, so it seems a lost opportunity to ignore them in fitness dynamics models. We discussed this issue in (MARTIN AND ROQUES 2016; ROQUES *et al.* 2017).

2. [Integro-differential models](#)

The main alternative to origin – fixation models is to follow the dynamics of the full distribution of fitness, at the natural timescale of the demography (continuous time or generations), not at the scale of fixations. To my knowledge, the first such model of evolution on a “fitness space” was introduced by Tsimring *et al.* (1996), thus fairly simultaneously with origin-fixation models. Let $p_t(m)$ be the frequency, at time t , of a genotype with Malthusian fitness m . The frequency dynamics, by selection between arbitrarily many types, has the form of an integro-differential equation, in continuous time (counted in arbitrary units):

$$\begin{aligned} \partial_t p_t(m) &= (m - \bar{m}_t) p_t(m) \\ \bar{m}_t &= \int_{m \in \mathbb{R}} m p_t(m) dm \end{aligned} \quad [5]$$

This is the exact dynamics for a continuous time demography (e.g. microbes undergoing a birth death process with asynchronous divisions). It is also the result of a continuous time approximation to a discrete time demography (e.g. annual plants with non-overlapping generations and synchronous reproduction). Time is then counted in generations and fitness is the genotype-specific per generation reproductive output W (Darwinian fitness). The corresponding Malthusian fitness is then $m \equiv \log W$ and the frequency change over one generation are:

$$p_{t+1}(m) = \frac{W}{\bar{W}_t} p_t(m) = \frac{e^m}{\bar{W}_t} p_t(m) \quad [6]$$

$$\bar{W}_t = \int_{m \in \mathbb{R}} W p_t(m) dm = \int_{m \in \mathbb{R}} e^m p_t(m) dm$$

Now, assume that the variance in Malthusian fitness is small (a weak selection limit), then $\bar{W}_t \approx e^{\bar{m}_t}$, so that eq.[6] approximately yields $p_{t+1}(m) \approx e^{m-\bar{m}_t} p_t(m)$. Small fitness variance also implies (by Fisher's fundamental theorem of natural selection) that \bar{m}_t hardly changes from t to $t + 1$. The solution of eq.[5], integrated from t to $t + 1$ and with \bar{m}_t constant over this period, is $p_{t+1}(m) \approx e^{m-\bar{m}_t} p_t(m)$. The two models (eqs. [5] and [6]) are thus equivalent and eq.[5] can thus be used for both types of demography. This allows using a single model for a diversity of species.

Mutation introduces a convolution of the distribution of types present, and the DFE. Let $f(s|m)$ be the probability density function of the DFE, for mutations arising in a background with fitness m . When $f(s|m) = f(s)$ does not depend on m , the DFE is non-epistatic as in eq.[4]. The resulting dynamics of the fitness distribution under selection and mutation (at rate U per capita per unit time) is an Integro-differential equation:

$$\partial_t p_t(m) = (m - \bar{m}_t) p_t(m) + U \left(\int_{x \in \mathbb{R}} f(m-x|x) p_t(x) dx - p_t(m) \right) \quad [7]$$

The dynamics of fitness distributions in this approach are illustrated in **Figure 3** below (reproduced from Fig. 1 of MARTIN AND ROQUES 2016):

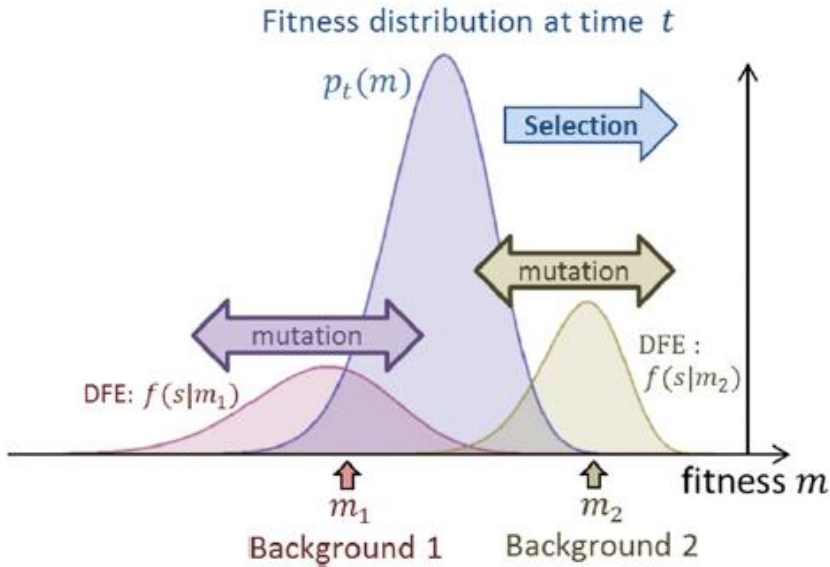


Figure 1 The standing fitness distribution [$p_t(m)$, blue curve] travels to the right by selection. Each genetic background under this distribution (e.g. m_1 and m_2 here) mutates to new genotypes with fitness $m_i + s$, where s has the density $f(s|m_i)$ depending on background fitness (red and brown curves for m_1 and m_2 , respectively).

In the first version of this approach (TSIMRING *et al.* 1996), non-epistatic mutation was assumed ($f(s|x) = f(s)$) and its effect approximated by diffusion: $\Delta_{mut} p_t(m) \approx \Delta t D \partial_m^2 p_t(m) - v \partial_m p_t(m)$ for some positive coefficients D and v . This particular subcase has recently yielded an exact solution (ALFARO AND CARLES 2014), via elaborate techniques borrowed from the study of the Schrödinger equation. Apart from that, exact explicit solutions over time, for eq.[7], seem hard to find. It is likely that more complex scenarios (multiple fitness components, dynamic environmental conditions), will be even harder to tackle. However, numerical analysis of these equations is possible, e.g. in a work by Marie Eve Gill (2019).

Compared to origin-fixation models in eq.[4], eq.[7] has the advantage of including both deleterious and beneficial mutations, predicting a trajectory not just for the mean but for the whole distribution of fitness (which is potentially empirically available although seldom measured). It also naturally allows for epistasis or change in environmental conditions, as no stationarity is not required by construction.

However, eq.[7] is built directly as a deterministic model ignoring stochastic variation (genetic drift and mutational randomness). The model is typically only accurate over some finite time (MARTIN AND ROQUES 2016), then the loss of rare fitter mutants, that is ignored in the model, can become important, especially in non-epistatic models. We have proposed (ROQUES *et al.* 2017) a free boundary approximation to connect the early dynamics governed by an Integro-differential model, with a long term adaptation rate from travelling wave origin-fixation models (GOOD *et al.* 2012). However it has the same limits as the origin-fixation model on which it depends.

Another drawback of the Integro-differential approach, as we have seen, is that it does not easily yield explicit results, even for simple non-epistatic models. Working with genotype frequencies and not densities (as is done in demography) has the advantage that similar dynamics apply under various demographic regimes. However, when it is these demographic dynamics that are the ultimate focus, this advantage is moot. On the other hand, working with frequencies imposes a complex mathematical treatment because frequency dynamics are always non-linear (eq.[5]), even in various contexts where demographic dynamics are linear, as will be apparent in the next sections.

In this chapter, I aim to derive a fitness dynamics model at the natural timescale of the demography and without requiring stationary regimes (as in Integro-differential models), but to build it as an explicit expectation over stochastic replicates (as in origin-fixation models). Instead of genotype frequencies $p_t(m)$, I focus on the expected Laplace transform of genotype densities, an approach akin to that used by previous authors studying asexual adaptation (RATTRAY AND SHAPIRO 2001; GOOD AND DESAI 2013)}. I just use it here in the different context of epistatic mutation, and on densities rather than frequencies. Although focusing on a less intuitive object than densities or frequencies, I think this approach is more amenable to analytical treatment in various contexts, and it still provides the key biological information we are interested in. What I will present is a generalization of a method I have been developing since the years 2010's, but that we elaborated much more since I started collaborating with Lionel Roques in 2014-2015. Several of the corresponding papers that we published since can be seen as subcases of this general model, and I think there are several other interesting subcases to analyze in the future. This is why I directly present the model in its general form.

Only very recently have I tried tackling the problem on densities rather than frequencies and I do not yet fully grasp the limits and accuracy of the approach. In particular, I lack a wealth of simulation checks to evaluate to what extent it does a better job at capturing stochastic biases than the formulation we have used so far, since (MARTIN AND ROQUES 2016). What is certain is that it retrieves all our previous published results (which have been checked by simulations), so it cannot be worse. I think that, compared to mere deterministic limits, this formulation provides a less *ad hoc* treatment of the microscopic and stochastic processes underlying eco-evolutionary dynamics. In any case, the point of presenting projects to future students is that they be partly open, so I decided to present the model in this form.

B. Common assumptions and definitions

1. Notations

All along a real number is given in *italic*, while a vector of real numbers is given in **bold**. I write X_t for a stochastic process over time and $X(t)$ for a deterministic function of time where X may be a variable or a functional (a function of some variable). I denote $\partial_z X$ the partial derivative of the function X with respect to the particular variable z . I denote $\nabla_{\mathbf{z}} = \{\partial_{z_i}\}_{i \in [1,k]}$ the gradient operator: a k -vector containing all the partial

derivatives with respect to each of the variables z_i in \mathbf{z} . The notation $\oint_{\mathbb{R}^k} X \, d\mathbf{m}$ refers to integration of function X over a multivariate variable $\mathbf{m} \in \mathbb{R}^k$. Finally $\mathbf{a} \cdot \mathbf{b}$ is the scalar product of two vectors \mathbf{a} and \mathbf{b} .

Table 2 below summarizes the main notations.

\mathbf{m}	Set of k fitness components (FCs) characterizing a genotype
m_i	Individual fitness component $i \in [1, k]$
$N_t(\mathbf{m})$	Density of genotype \mathbf{m} at time t
$\boldsymbol{\alpha}(t)$	Vector of linear coefficients relating fitness components and Malthusian fitness
$m = \boldsymbol{\alpha}(t) \cdot \mathbf{m} + c(t)$	Malthusian fitness of a genotype: expected growth rate of the genotype
$\mathbf{z} \in \mathbb{R}^k$	Dummy variable of the multivariate Laplace transform of genotype densities
$\oint_{\mathbb{R}^k} X(\cdot, \mathbf{m}) \, d\mathbf{m}$	Integral of a function $X(\cdot, \mathbf{m})$ over all fitness components $\mathbf{m} \in \mathbb{R}^k$
$\ell_t(\mathbf{z})$	Laplace transform of the genotype densities in one stochastic replicate trajectory of genotype densities
$\mathcal{L}_t(\mathbf{z}) \equiv \langle \phi_t(\mathbf{z}) \rangle$	ELT: Expected Laplace Transform of genotype densities over the distribution of their stochastic trajectories ("ensemble mean")
U	Mutation rate per unit time per capita (constant mutation model)
u	Mutation rate per division per capita (birth – dependent mutation model)
$\mathbf{s} \mathbf{m} = \mathbf{m}' - \mathbf{m}$	change in FCs by a single mutation (mutation effect on FCs)
$M_s(\mathbf{z}, \mathbf{m}) = \mathbb{E}(e^{s \cdot \mathbf{z}} \mathbf{m})$	Moment Generating Function of mutation effects on FCs, may depend on the background FC value \mathbf{m} (epistasis)
$\partial_z X(\cdot, z_*)$	Partial derivative of function X with respect to variable z , taken at the particular point $z = z_*$
$\nabla_{\mathbf{z}} X(\cdot, \mathbf{z}_*)$	Gradient vector of X at $\mathbf{z} = \mathbf{z}_*$: all partial derivatives of X with respect to all variables in \mathbf{z} , taken at the particular point $\mathbf{z} = \mathbf{z}_*$

2. [Population and timescale](#)

I present a model for a single isolated population (I implement the term due to migration from another population in section D.3). It is asexual and haploid for most

examples but section D.4 tackles the case of sexual reproduction in diploids. The population consists of a set of genotypes (which number may vary arbitrarily) each characterized by a set of k fitness components $\mathbf{m} = \{m_i\}_{i \in [1,k]} \in \mathbb{R}^k$, which biological nature will depend on the scenarios considered. The number of copies of a given genotype at time t is a random variable denoted $N_t(\mathbf{m})$. I will call it “density of the genotype” for short.

I consider only continuous time dynamics. This is either to model the exact demography of a population with overlapping generations (e.g. cells undergoing a birth-death process), or to take a weak selection approximation of a discrete time demography (e.g. annual plants), see eqs. [5]-[6].

3. Fitness

Malthusian fitness is the only one that can be defined in a continuous time demography, as there are no discrete synchronous stages at which one can come and count the offspring of all genotypes at once. Consequently, I simply refer to “fitness” for the Malthusian fitness $m = m(t, \mathbf{m})$. When not necessary, I drop the reference to both time and genotype for notational compactness, but fitness does depend on both genotype and time in general.

I believe that, contrary to what one may read sometimes), the definition of fitness, in any demographic continuous time model, is unequivocal, especially as we ignore particular interactions like kin – selection or frequency dependent selection. It is the expected instantaneous exponential growth rate of the subpopulation consisting of all copies of a genotype \mathbf{m} . So-called ‘fitness proxies’ may be all that is available empirically, but, unless they can be convincingly connected to exact fitness, their dynamics are difficult to interpret and they definitely cannot be interpreted as fitness.

4. Central assumptions

The key assumptions required in the model are:

Assumption 1: FCs are deterministic and fixed for a given genotype, and do not depend on time or any population or environmental variable.

Assumption 2: FCs contribute additively to fitness via deterministic coefficients:

$$m(t, \mathbf{m}) \equiv m = \boldsymbol{\alpha}(t) \cdot \mathbf{m} + c(t) \quad [8]$$

The vector $\boldsymbol{\alpha}(t) = \{\alpha_i(t)\}_{i \in [1,k]} \in \mathbb{R}^k$ contains the k coefficients relating FCs and overall fitness. For example, the fitness components may be the per capita per unit time rates of birth (b) and death (d) of the microbial genotype ($k = 2$) under a linear birth-death model. We then have $\mathbf{m} = \{b, d\}$ and setting $\boldsymbol{\alpha}(t) = \{1, -1\}$ (deterministic) yields $m = \boldsymbol{\alpha}(t) \cdot \mathbf{m} = b - d$ which is indeed the expected growth rate of the genotype characterized by $\mathbf{m} = \{b, d\}$.

Assumption 3: the non-genetic contribution to growth rate ($c(t) \in \mathbb{R}$) is either deterministic or does not covary with any population characteristics. Thus, absolute fitness is defined to the nearest constant (with respect to genotypes and stochastic events).

Assumption 4: selection is frequency-independent. Over a small time interval δt , the expected change in density of genotype \mathbf{m} , averaged over stochastic events during δt and conditional on its current density $N_t(\mathbf{m})$ is thus:

$$\mathbb{E}(\delta N_t(\mathbf{m}) | N_t(\mathbf{m})) = m N_t(\mathbf{m}) \delta t \quad [9]$$

5. Neutral or semi-neutral components characterizing a genotype

An important remark at this stage: it is not necessary that the components have an effect on the growth rate: they may be purely neutral (have no effect on demography overall) or 'semi-neutral' (having no effect on growth rate but undergoing indirect selection, like modifiers). As long as they are a characteristic property of a genotype, and given a model for their distribution by mutation, they can be implemented as well. It may still be useful to follow such components, if they affect other aspects of the evolutionary process or if they are key measurable quantities in experiments. The components I have in mind are:

h : Hamming distance between the genome's sequence and that of some reference genotype (typically the ancestor strain in asexual evolution experiments), if we wish to predict a trajectory in genetic sequence space.

$\mathbf{y} \in \mathbb{R}^1$ or \mathbb{R}^2 or \mathbb{R}^3 : Position on some physical space in one, two or three dimensions

δ : dispersal rate (if we wish to study the evolution of dispersal).

U or u : mutation rate per generation (discrete time) or per division (continuous time) if we wish to follow the evolution of mutation rates.

σ : selfing rate if we wish to follow the evolution of selfing.

These elements, by assumption, do not contribute to fitness: they have corresponding coefficients $\alpha_i(t) \equiv 0$ in the vector $\boldsymbol{\alpha}(t)$. Yet, introducing them allows to follow the evolution of modifiers of migration, mutation or sex (δ, u or U, σ), to follow an evolutionary-demographic dynamics in space (\mathbf{y}) or to propose an additional testable prediction in terms of trajectories in sequence space (h).

In the case where we follow a position in physical space \mathbf{y} , it would obviously be interesting to let $c(t)$ or $\boldsymbol{\alpha}(t)$ also vary over space (spatially heterogeneous conditions). However, the results I present here do not apply in general in this case as position \mathbf{x} is then also a 'property' of a genotype. Introducing \mathbf{y} here is simply useful to track the spatial spread of an adapting population in a homogeneous environment. It should be possible to extend the approach for simple forms of $\{c(t, \mathbf{y}), \boldsymbol{\alpha}(t, \mathbf{y})\}$, e.g. polynomial in \mathbf{y} , but I have not tried any of this so I will skip the issue in this report.

6. Laplace transforms

Genotype densities $N_t(\mathbf{m})$ will change stochastically by birth, death, mutation, and sometimes reproduction or migration events. The central quantity that will allow us to model the resulting changes of the whole population is an expected Laplace transform of genotype densities (where the expectation is taken over stochastic replicates of the whole process).

The multivariate Laplace transform of genotype densities (of the function $N_t(\mathbf{m})$) is

$$\begin{aligned} \ell_t : \mathbb{R}^k &\rightarrow \mathbb{R} \\ \ell_t(\mathbf{z}) &\equiv \oint_{\mathbb{R}^k} N_t(\mathbf{m}) e^{\mathbf{z}\cdot\mathbf{m}} \mathbb{d}\mathbf{m} \end{aligned} \quad , \quad [10]$$

Because $N_t(\mathbf{m})$ is a stochastic process $\ell_t(\cdot)$ is a stochastic function of \mathbf{z} with undetermined value (depending on the fluctuations in genotype densities). However, we can also define the expected Laplace transform over all possible replicates of the stochastic trajectories of $N_t(\mathbf{m})$. We use a classic notation $\langle X \rangle$ for this expectation of some random variable X over stochastic replicates, a so-called 'ensemble mean'. We thus define the function:

$$\begin{aligned} \mathcal{L} : \mathbb{R}^+ \times \mathbb{R}^k &\rightarrow \mathbb{R} \\ \mathcal{L}(t, \mathbf{z}) &\equiv \langle \ell_t(\mathbf{z}) \rangle = \left\langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) e^{\mathbf{z}\cdot\mathbf{m}} \mathbb{d}\mathbf{m} \right\rangle \end{aligned} \quad , \quad [11]$$

Which we denote Expected Laplace Transform (**ELT**) of genotype densities. This is now a fully deterministic function of time ($t \in \mathbb{R}^+$) and of $\mathbf{z} \in \mathbb{R}^k$ (it is an expectation over a stochastic process). We can derive its dynamics in terms of deterministic partial differential equations (PDEs), or coupled PDE-ODE systems etc., in various situations.

Before deriving these PDEs, let us first answer this question: "what's the point of studying this highly unintuitive function of the actual demographic and evolutionary process?". In the next section, I detail a set of handy properties of the ELT, which allow to compute many biologically relevant quantities from a knowledge of $\mathcal{L}(t, \mathbf{z})$.

C. Properties of the ELT of genotype densities

The most direct use of the ELT is to invert it. Genotype densities and their corresponding Laplace transform are in bijection, so in principle, inverting the Laplace transform will yield a proxy for the expected genotype density distribution, hence the density distribution of fitness components. However, such inverse Laplace transform may be numerically challenging, and it is not necessary: many important observable quantities can be retrieved from the ELT in a much simpler way.

Sometimes, I will consider a particular variable $X = \phi(t, \mathbf{m})$ that depends on the FCs and possibly on time, and define the ELT for this variable (among the genotypes of the population at time t) as $\mathcal{L}_X: \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}$: $\mathcal{L}_X(t, z) = \langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) e^{z \phi(t, \mathbf{m})} \mathbb{d}\mathbf{m} \rangle$ for $z \in \mathbb{R}$.

1. Expected total size

It is easily checked that, from the definition in eq.[11]

$$\mathcal{L}(t, \mathbf{0}) = \langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) \mathbb{d}\mathbf{m} \rangle = \langle \mathcal{N}_t \rangle \equiv \tilde{\mathcal{N}}(t) \quad , \quad [12]$$

where $\mathcal{N}_t = \int_{\mathbb{R}^k} N_t(\mathbf{m}) \mathbb{d}\mathbf{m}$ is the sum of the densities of all genotypes, i.e. the total size of the population. The ELT, taken at $\mathbf{z} = \mathbf{0}$ yields the expected population size over time. In most applications of my models (evolutionary rescue, invasion of new habitats etc.) this is the central quantity of interest.

2. ELT for the distribution of a particular FC and of fitness itself

The multivariate ELT of the FCs yields the ELT of each fitness component individually (marginal distribution). Define the k -vector with all elements equal to zero except at some particular position i where it is equal to one: $\bar{\mathbf{1}}_i \equiv \{a_j\}_{j \in [1, k]}$ where $a_j = 0$ if $j \neq i$ and $a_i = 1$. The ELT for the i^{th} fitness component is given by:

$$\mathcal{L}_{m_i}(t, z) = \mathcal{L}(t, z \bar{\mathbf{1}}_i) = \langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) e^{z m_i} \mathbb{d}\mathbf{m} \rangle \quad . \quad [13]$$

Similarly, consider fitness itself, measured in the conditions of some time t_* which may differ from the current time t : $m(t_*) = \alpha(t_*) \cdot \mathbf{m} + c(t_*)$. From eq.[11], the ELT for fitness, among genotypes present at time t and measured in conditions of time t_* , is given for $z \in \mathbb{R}$ by:

$$\begin{aligned} \mathcal{L}_{m(t_*)}(t, z) &= \mathcal{L}(t, z \alpha(t_*)) e^{z c(t_*)} \\ &= \langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) e^{z (\alpha(t_*) \cdot \mathbf{m} + c(t_*))} \mathbb{d}\mathbf{m} \rangle = \langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) e^{z m(t_*)} \mathbb{d}\mathbf{m} \rangle \quad . \quad [14] \end{aligned}$$

In general, this allows to deduce information on fitness characteristics and trajectories, when only this quantity is measured and not its underlying components \mathbf{m} .

The property in eq. [14] can also be useful when considering evolution experiments where the environment is changed (e.g. coevolution, drug dose regimens etc.), and the fitness of samples from different times is evaluated in various conditions corresponding to past, present and future times.

3. Density of the fittest genotype

Assume that there is, at any given time t_* , a predictable, finite maximum fitness $m_{max}(t_*)$ in the landscape. It is the case in any fitness landscape with a global optimum (that may be moving over time): the optimal genotype in the conditions of time t_* has fitness $m_{max}(t_*) = \max(m(t_*))$. Then the limit

$$\lim_{z \rightarrow \infty} \mathcal{L}_{m(t_*)}(t, z) e^{-z m_{max}(t_*)} = \langle N_t(m_{max}(t_*)) \rangle \quad [15]$$

Yields the expected density, at any time t , of the genotype that would be optimal in the conditions of time t_* . We used a similar argument in (MARTIN AND ROQUES 2016; GIL *et al.* 2019) to derive the frequency of the optimal genotype at mutation-selection balance in the FGM ($t = t_* = \infty$) and the result was derived as a deterministic approximation. Eq.[15] shows that it is a much more versatile tool, and that it can be properly obtained as an exact expectation over stochastic replicates, not as a deterministic limit.

In principle (I do not explore it here), I think we could access the density of any given genotype \mathbf{m}_* by setting the optimum (setting a particular value of $\alpha(t)$) at a position where \mathbf{m}_* is optimal and applying eq.[15].

A rule can be stated to give intuition on the biological meaning of these Laplace Transforms. The shape of $\log \mathcal{L}_X(t, z)$ around $z = 0$ tends to describe the bulk of X values among the most common genotypes (the mean, variance etc. of X), while its shape near $|z| = \infty$ depends on the extrema (on the fittest genotypes at $z = +\infty$ and the least fit genotypes at $z = -\infty$). This also provides a connection between extreme value theory and ELTs: the approximate shape of $\mathcal{L}_X(t, z)$ near $z \rightarrow \infty$ corresponds to the density distribution of the maxima of X for any variable X . Once rescaled as frequency distribution, it should thus pertain to one of the three domains of attraction of the extreme value theory: Gumbel, Fréchet and Weibull domains. I have explored the literature a little bit to find developments in this direction but could not find any, unfortunately (but this is a literature I am not familiar with of course).

4. Linear operators

In deriving the PDEs for $\mathcal{L}(t, \mathbf{z})$, we use the fact that the three following operators are linear operators acting on ℓ_t : (i) differentiation with respect to \mathbf{z} or time t , (ii) integration with respect to \mathbf{m} and (iii) expectation with respect to microscopic stochastic events or to full replicate trajectories of genotype densities. These three operators can thus be swapped at will in the expressions.

We can thus write the following equalities which are handy to compute the dynamics of \mathcal{L} from microscopic events and to extract the moments of the distribution of \mathbf{m} at any time (see next sections).

Among other uses of this property (see later) we can note that the microscopic and stochastic changes in genotype densities, over some small δt (denoted $\delta N_t(\mathbf{m})$) are connected to the deterministic time derivative of \mathcal{L} ($\partial_t \mathcal{L}$) as follows:

$$\partial_t \mathcal{L}(t, \mathbf{z}) \equiv \langle \partial_t \ell_t(\mathbf{z}) \rangle = \lim_{\delta t \rightarrow 0} \frac{\langle \delta \ell_t(\mathbf{z}) \rangle}{\delta t} = \lim_{\delta t \rightarrow 0} \oint_{\mathbb{R}^k} \langle \delta N_t(\mathbf{m}) \rangle \frac{e^{\mathbf{z} \cdot \mathbf{m}}}{\delta t} \mathbb{d}\mathbf{m} \quad , \quad [16]$$

5. Trajectory of mean fitness and mean FCs

So far we have required no particular assumption to extract biological information from the ELT. For the next properties we require one additional assumption:

Assumption 5: *The total population size is approximately deterministic* (across stochastic replicates): $\langle \mathcal{N}_t \rangle \approx \tilde{\mathcal{N}}(t)$ some deterministic function of time given by $\mathcal{L}(t, \mathbf{0})$ (eq.[12]).

This means that, while each genotype's density may still be small and highly fluctuating across replicates, the trajectory of the total population size remains roughly stable across replicates. This is typically a large population assumption. Then the ELT is simply related to the (expected) moment-generating function of the distribution of \mathbf{m} in the population.

Let $p_t(\mathbf{m}) \equiv N_t(\mathbf{m})/\mathcal{N}_t$ be the frequency of genotype \mathbf{m} in the population. In general, $\langle p_t(\mathbf{m}) \rangle \neq \langle N_t(\mathbf{m}) \rangle / \langle \mathcal{N}_t \rangle$, as the expectation of a ratio is not the ratio of the expectations. However, under Assumption 5, $p_t(\mathbf{m}) \approx N_t(\mathbf{m})/\tilde{\mathcal{N}}(t)$ in all replicates so that $\langle p_t(\mathbf{m}) \rangle \approx \langle N_t(\mathbf{m}) \rangle / \tilde{\mathcal{N}}(t) = \langle N_t(\mathbf{m}) \rangle / \langle \mathcal{N}_t \rangle$. This means that

$$\phi(t, \mathbf{z}) \equiv \frac{\mathcal{L}(t, \mathbf{z})}{\mathcal{L}(t, \mathbf{0})} = \frac{\langle \oint_{\mathbb{R}^k} N_t(\mathbf{m}) \nabla_{\mathbf{z}} e^{\mathbf{z} \cdot \mathbf{m}} \mathbb{d}\mathbf{m} \rangle}{\langle \mathcal{N}_t \rangle} \approx \langle \oint_{\mathbb{R}^k} p_t(\mathbf{m}) \nabla_{\mathbf{z}} e^{\mathbf{z} \cdot \mathbf{m}} \mathbb{d}\mathbf{m} \rangle = \langle M_{\mathbf{m}}(t, \mathbf{z}) \rangle \quad , \quad [17]$$

where $M_{\mathbf{m}}(t, \mathbf{z})$ is the Moment Generating Function (MGF) of the multivariate distribution of FCs at time t (which is stochastic as the genotype frequencies are stochastic). The ELT rescaled by its values at $\mathbf{z} = \mathbf{0}$ provides the expected MGF of the joint distribution of all FCs. This means, by properties of MGFs, that derivatives of $\phi(t, \mathbf{z})$ at $\mathbf{z} = \mathbf{0}$ yield the expected raw moments of the joint distribution of \mathbf{m} . For example, for any given pair of components $\{m_i, m_j\}$ and any pair of orders $\{n_i, n_j\}$, we have the following approximation: $\partial_{z_i}^{n_i} \partial_{z_j}^{n_j} \phi(t, \mathbf{z}) \approx \langle \partial_{z_i}^{n_i} \partial_{z_j}^{n_j} M_{\mathbf{m}}(t, \mathbf{0}) \rangle = \overline{\langle m_i^{n_i} m_j^{n_j} \rangle}$. Where ∂_z^n is the derivative of order n with respect to variable z .

The expected MGF of the distribution of fitness itself is also readily available (setting $t = t_*$ in eq.[14]): $\mathcal{L}_m(t, z)/\mathcal{L}_m(t, 0) = \phi(t, z, \alpha(t)) \approx \langle M_m(t, z) \rangle$.

The main straightforward application of this property is to derive mean fitness trajectories. Using the interchangeability of the various linear operators involved (derivation, expectation, integration, average within replicate populations, scalar product) and under assumption 5 (deterministic total size), we have:

$$\begin{aligned} \langle \bar{\mathbf{m}} \rangle &= \langle \nabla_{\mathbf{z}} M_{\mathbf{m}}(t, \mathbf{z}) \rangle \approx \frac{\nabla_{\mathbf{z}} \mathcal{L}(t, \mathbf{0})}{\mathcal{L}(t, \mathbf{0})} \\ \langle \bar{m} \rangle &= \alpha(t) \cdot \langle \bar{\mathbf{m}} \rangle + c(t) \approx \alpha(t) \cdot \frac{\nabla_{\mathbf{z}} \mathcal{L}(t, \mathbf{0})}{\mathcal{L}(t, \mathbf{0})} + c(t) \end{aligned} \quad , \quad [18]$$

Which is the other central property of the ELT, together with eq.[12] for the expected total population size trajectory.

It is tempting to push the application further to access the variance-covariance of FCs or the variance in fitness. This is possible, but rigorously, it only applies in a true deterministic limit where $\phi(t, \mathbf{z}) \approx M_{\mathbf{m}}(t, \mathbf{z})$ (without the $\langle \cdot \rangle$ for ensemble mean). Indeed, otherwise, at higher orders the central moments involve products of the MGF and thus covariation between MGF terms across stochastic replicates would affect the result. However, we have used an equivalent of this deterministic limit for higher moments in previous work (e.g. MARTIN AND ROQUES 2016) and it does prove fairly accurate in large enough populations. Furthermore the expected higher order raw moments (e.g. $\langle \overline{m^2} \rangle \approx \langle \partial_{\mathbf{z}}^2 M_{\mathbf{m}}(t, \mathbf{z}) \rangle = \partial_{\mathbf{z}}^2 \phi(t, \mathbf{z} \alpha(t))$) can also be directly estimated from replicate empirical trajectories and compared to the theory, without requiring the fully deterministic approximation.

D. Effect of evolutionary forces on the system

1. Demography and selection

Eq.[8] and [9] together provide the expected change in genotype densities, over some small δt under the sole influence of demography (and hence selection), conditional on the current genotype density:

$$\mathbb{E} \left(\frac{\delta N_t(\mathbf{m})}{\delta t} \middle| N_t(\mathbf{m}) \right) = m(t, \mathbf{m}) N_t(\mathbf{m}) = (\alpha(t) \cdot \mathbf{m} + c(t)) N_t(\mathbf{m}) \quad , \quad [19]$$

We note that $\mathbf{m} N_t(\mathbf{m}) e^{\mathbf{z} \cdot \mathbf{m}} = \nabla_{\mathbf{z}} N_t(\mathbf{m}) e^{\mathbf{z} \cdot \mathbf{m}} = \nabla_{\mathbf{z}} g_{t, \mathbf{m}}(\mathbf{z})$. Multiplying eq.[19] by $e^{\mathbf{z} \cdot \mathbf{m}}$ and integrating over genotypes \mathbf{m} , the corresponding change in $\ell_t(\mathbf{z})$, conditional on its current value, is thus:

$$\mathbb{E} \left(\frac{\delta \ell_t(\mathbf{z})}{\delta t} \middle| \ell_t(\mathbf{z}) \right) = \alpha(t) \cdot \nabla_{\mathbf{z}} \ell_t(\mathbf{z}) + c(t) \ell_t(\mathbf{z}) \quad , \quad [20]$$

This is a linear operator acting on $\ell_t(\mathbf{z})$ (as α and c are deterministic functions by assumptions 2 and 3). Thus, deconditioning on $\ell_t(\mathbf{z})$ to get the 'ensemble mean' (eq.[16]) yields a simple linear PDE for the ELT under selection alone:

$$\partial_t \mathcal{L}(t, \mathbf{z}) = \alpha(t) \cdot \nabla_{\mathbf{z}} \mathcal{L}(t, \mathbf{z}) + c(t) \mathcal{L}(t, \mathbf{z}) \quad . \quad [21]$$

In the univariate case ($k = 1$, e.g. simply fitness $\alpha(t) = 1$), this PDE can readily be solved in *Mathematica* ®. The general solution can be found by introducing the auxiliary

function: $Q(t, \mathbf{z}) \equiv \log\left(\mathcal{L}\left(t, \mathbf{z} - \int_0^t \boldsymbol{\alpha}(\tau) d\tau\right)\right) - \int_0^t c(\tau) d\tau$, and noting that eq.[21] implies $\partial_t Q(t, \mathbf{z}) = 0$ so that $Q(t, \mathbf{z}) = Q(0, \mathbf{z}) = \log \mathcal{L}_0(\mathbf{z})$. Inverting then yields the unique solution of eq.[21]:

$$\mathcal{L}(t, \mathbf{z}) = e^{\int_0^t c(\tau) d\tau} \mathcal{L}_0\left(\mathbf{z} + \int_0^t \boldsymbol{\alpha}(\tau) d\tau\right) \quad [22]$$

This gives the general dynamics of the ELT for any starting condition (describing the initial standing variance within and between replicate populations). The expected population size under these eco-evolutionary dynamics is then given by (eq.[12]):

$$\langle \mathcal{N}_t \rangle = \tilde{\mathcal{N}}(t) = \mathcal{L}(t, \mathbf{0}) = e^{\int_0^t c(\tau) d\tau} \mathcal{L}_0\left(\int_0^t \boldsymbol{\alpha}(\tau) d\tau\right) \quad [23]$$

Provided assumptions 1 through 4 hold, there is no assumption on the stochastic effects here: even with the smallest possible starting populations, the prediction for the expected population size (across replicates) should hold.

Under assumption 5 (large population approximation), the trajectory of all mean fitness components is also available given the initial condition:

$$\langle \bar{\mathbf{m}} \rangle = \frac{\nabla_{\mathbf{z}} \mathcal{L}(t, \mathbf{0})}{\mathcal{L}(t, \mathbf{0})} = \frac{\nabla_{\mathbf{z}} \mathcal{L}_0\left(\int_0^t \boldsymbol{\alpha}(\tau) d\tau\right)}{\tilde{\mathcal{N}}(t)} \quad [24]$$

Note that when $\mathbf{m} = m$, $\boldsymbol{\alpha}(t) = 1$, eq.[21] applies to fitness itself in constant conditions, and takes the form: $\partial_t \mathcal{L}(t, z) = \partial_z \mathcal{L}(t, z) + c(t) \mathcal{L}(t, z)$. Selection (the term $\partial_z \mathcal{L}(t, z)$ induced by differences in fitness among genotypes) yields a transport term, from right to left, in z -space. This means that the current shape of the function at large z determines the future shape around $z = 0$. This is because the currently fitter genotypes (fitter edge of the distribution described by \mathcal{L} as $z \rightarrow \infty$) will be those that dominate, in the future, and will thus determine the bulk then (described by \mathcal{L} around $z = 0$). The current bulk (around $z = 0$) will stand among the least fit ($z \rightarrow -\infty$) in the future.

This correspondence between selection effect in fitness space (on genotype densities) and in z - space (on ELT) is illustrated in **Figure 2** below. Here we assume a normal $N(0, \sigma)$ for the initial fitness distribution, a population of size N_0 ($\mathcal{L}_0 = N_0 e^{\sigma^2 z^2 / 2}$), and a constant death rate ($c(t) = -d$). Eq.[22] yields an expanding population ($\tilde{\mathcal{N}}(t) = N_0 e^{\sigma^2 t^2 / 2 - t d}$) with normally distributed fitness $m \sim \text{Normal}(\sigma^2 t, \sigma)$, a gaussian wave traveling at constant speed $\partial_t \bar{m} = \sigma^2$.

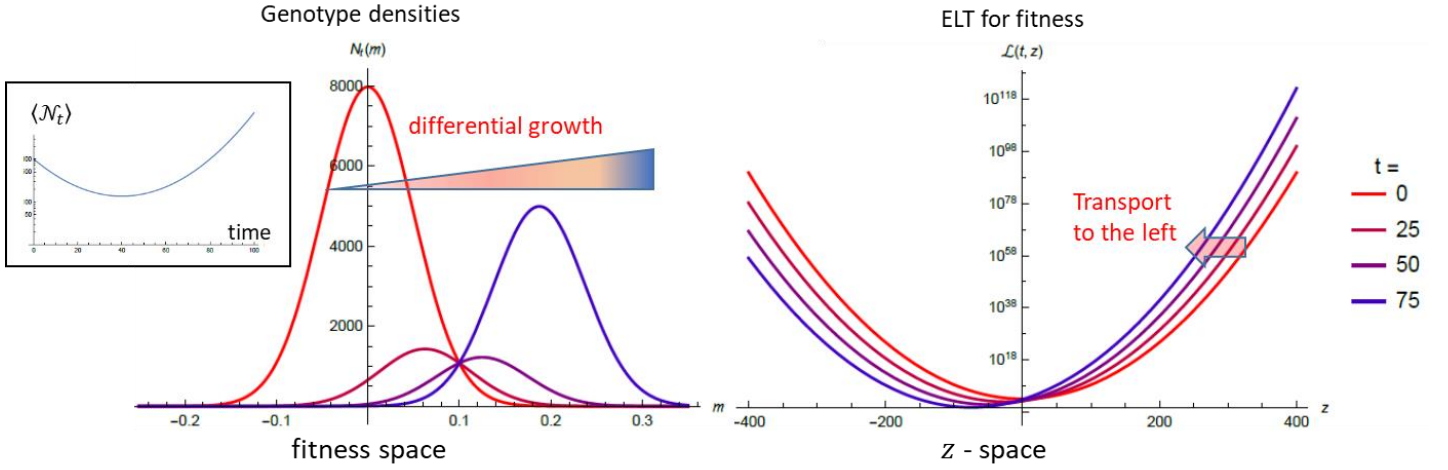


Figure 4: change in density (left panel) and ELT (right panel) over time (see legend). As the fittest classes outcompete the least fit classes, the ELT shifts to the left. Left insert: expected population size dynamics $\langle N_t \rangle$ with decay and rebound when $\bar{m} = \sigma^2 t > d$. Here $N_0 = 10^3, \sigma = 0.05, d = 0.1$.

2. Mutation

We now introduce mutation in the system: mutations arise anywhere in the genome, have a distribution of effects (random variable \mathbf{s}) on all the FCs. This random variable may depend on the background if mutation is epistatic on fitness. The moment generating function (MGF) of the (epistatic) mutation effects on the FCs is denoted $M_s(\mathbf{z}|\mathbf{m})$ for a background with FC value \mathbf{m} . At that stage we require another assumption.

Assumption 6: log-linear background dependent kernel of mutation. In our work with Lionel on generating functions (MARTIN AND ROQUES 2016), this was the key assumption allowing to close the dynamics. If there is epistasis in the effect of mutations on FCs, we require that the resulting depend loglinearly on the parent FCs.

The mutation kernel MGF must be of the form:

$$M_s(\mathbf{z}|\mathbf{m}) \equiv \mathbb{E}(e^{\mathbf{s} \cdot \mathbf{z}}|\mathbf{m}) = M_*(\mathbf{z})e^{\boldsymbol{\omega}(\mathbf{z}) \cdot \mathbf{m}} \quad [25]$$

With some function $M_*(\mathbf{z})$ giving the MGF of the kernel for a parent at $\mathbf{m} = \mathbf{0}$, and $\boldsymbol{\omega}(\mathbf{z})$ a function describing epistasis. As discussed in (MARTIN AND ROQUES 2016) there are in fact a few models in which this is verified, among which the FGM. It is still the case in various extensions of the single fitness model, for example

- When the FCs are fitnesses in different habitats or fitness traits that have distinct optima in terms of the same underlying phenotype (MARTIN AND LENORMAND 2015b)
- When considering the two FCs required to model a moving optimum (with arbitrary movement along a single direction) (ROQUES *et al.* 2020)

- when considering an arbitrarily anisotropic model of mutation on phenotype (HAMEL *et al.* 2020)
- When considering diploids with additive effects of mutations on phenotypes (ongoing collaboration with Mathieu Siol)

Now there are two ways in which the mutation process can be considered, and this time, it does depend on whether we have a discrete time demography (synchronous reproduction) under a continuous time approximation or a truly continuous time demography (birth death process).

Discrete time: mutations occur at some rate U that is constant over time whatever the fitness of individuals. This corresponds to the classic models that have been most studied (and the ones I have studied so far). Individuals with FC \mathbf{m} mutate with probability $U N_t(\mathbf{m}) \Delta t$ over time interval Δt . Upon mutating, an individual of class \mathbf{m} becomes $\mathbf{m} \rightarrow \mathbf{m} + \mathbf{s}$. The resulting effect on the ELT, once averaged over all parents and mutation effects is then

$$\begin{aligned}
\frac{\Delta \mathcal{L}(t, \mathbf{z})}{\Delta t} &= U \langle \mathbb{E}_{\mathbf{m}} (N_t(\mathbf{m}) e^{\mathbf{m} \cdot \mathbf{z}} (\mathbb{E}_{\mathbf{s}} (e^{\mathbf{s} \cdot \mathbf{z}} | \mathbf{m}) - 1)) \rangle \\
&= U \langle \mathbb{E}_{\mathbf{m}} (N_t(\mathbf{m}) e^{\mathbf{m} \cdot \mathbf{z}} (M_*(\mathbf{z}) e^{\boldsymbol{\omega}(\mathbf{z}) \cdot \mathbf{m}} - 1)) \rangle \quad \cdot \quad [26] \\
&= U (M_*(\mathbf{z}) \mathcal{L}(t, \mathbf{z} + \boldsymbol{\omega}(\mathbf{z})) - \mathcal{L}(t, \mathbf{z}))
\end{aligned}$$

The continuous time approximation is then taken by letting $\Delta t \rightarrow 0$ yielding a closed PDE term. So this closes the dynamics of the ELT under mutation, but it introduces a non-local term in the PDE that has made it difficult to solve explicitly, and so far numerically for multiple FCs (not for fitness alone).

BD process with unicellulars: in true continuous time, at least with unicellulars, mutation occurs upon division, and as divisions are not synchronous anymore, mutation rate and birth rate are now connected: $U = b u$ where u is the mutation rate per division. The classic assumption around this is to consider that adaptation occurs on the death rate and that b is stable across individuals and time. But this is really a mathematical convenience: birth rates do evolve and vary across genotypes.

The way I have approximated this so far (unpublished work with my Postdoc Alvaro Mateos Gonzales) is assuming that the distribution of birth rates is narrow around its means, so that we can take a 'mean field' approximation: all individuals are assumed to mutate at rate $U \approx \tilde{U}(t) = \bar{b}(t) u$ where $\bar{b}(t)$ is the current mean birth rate in the population, which is available from the ELT as long as the birth rate is one of the FCs. In the case of a WSSM model with two FCs being the birth and death rate of each genotype, the trajectory of mean birth rate and death rate is can then be expressed as a simple ODE system. Otherwise I don't know so far.

3. [Migration](#)

The flow of individuals from one population to another is linear in densities so it is not difficult to introduce in the model. Let $\mathcal{L}_1(t, \mathbf{z})$ and $\mathcal{L}_2(t, \mathbf{z})$ be the two ELT of the FCs in the two populations connected by migration. The change in density due to migration at rate δ per capita per unit time is $\Delta \mathcal{N}_{t,1} = \delta (\mathcal{N}_{t,2} - \mathcal{N}_{t,1})$ where index 1 and 2 correspond to the population-specific densities. The change in ELT is thus linear:

$$\Delta \mathcal{L}_1(t, \mathbf{z}) = \delta (\mathcal{L}_2(t, \mathbf{z}) - \mathcal{L}_1(t, \mathbf{z}))$$

When considering different habitats for the two populations, the fitness components must be the fitness of a genotype in each habitat. The joint distribution of mutation fitness effects across the two habitats was obtained in my previous paper on FGM with multiple optima (MARTIN AND LENORMAND 2015b): the model is a log-linear background model of mutation so we can close the dynamics with mutation. The problem was implemented and analyzed in Florian Lavigne's first paper (LAVIGNE *et al.* 2020) but the situation simplified to a single habitat problem as we do not have to follow the source (stable). Florian studied the more general problem in the last part of his PhD (see his thesis) but the issue proves much more complicated, because of the coupling between the two ELTs introduced by migration. For the moment we are yet to explore approximations that could overcome this difficulty.

4. [Sexual reproduction](#)

An obvious ultimate extension of my models is to implement sexual reproduction. This may not be so much of a requirement for freely recombining genomes where quantitative genetics and the infinitesimal model provide a quite accurate and explicit modelling framework. However for partially sexuals or highly selfing species it might be of use. For the moment I have not found any easy way to implement the effect of recombination explicitly. However, there is a natural connection between the infinitesimal model and the gaussian kernel of mutation assumed in the FGM. In the case of a narrow phenotypic distribution, we may be able to use a House of Cards approximation, where recombination induces a gaussian change in phenotype around the mean phenotype of the two parents, which is always close to the overall mean phenotype in this situation. Thus I expect that in the SSWM regime, wherever a House of cards approximation is accurate, it be possible to implement sexual reproduction in the same way that mutation is implemented, except that the resulting variance in the phenotypic effects would have to be found explicitly.

E. Approximation methods to obtain explicit solutions

As we have seen, the dynamic of the ELT may be closed into a single PDE but this PDE is non local which makes it difficult to solve. I have so far used one single approach to

get explicit solutions in all previous work on these generating function trajectories. This approximation is valid for weak selection strong mutation (WSSM). I would like in the future to explore a complementary approximation based on the house of cards principle in the opposite regime.

1. [Diffusion approximation \(\$U \gg \lambda\$ \)](#)

The principle here is very simple: we take a leading order of the non-local mutational term in eq.[26] which makes it a linear first order term. This is what we have been using in all our treatments of fitness trajectories in the FGM so far, since (MARTIN AND ROQUES 2016). I adapt it here to the ELT which is a slightly different object from the generating functions we have used so far.

Consider a single fitness component evolving under mutation and selection in a fixed environment: $\mathbf{m} \rightarrow m, \mathbf{z} \rightarrow z, \boldsymbol{\omega}(\mathbf{z}) \rightarrow \omega(z) = -\lambda z^2/(1 + \lambda z), \boldsymbol{\alpha}(t) \rightarrow 1$ and $M_*(z) = (1 + \lambda z)^{-n/2}$ where n is the dimension of phenotypic space and λ is the mutational variance per trait. The fitness component $m \leq 0$ is a fitness distance to the optimum, while some rate $c(t)$ is the growth rate of a perfectly adapted genotype with fitness component $m = 0$.

The PDE for the ELT of the density of m in the population is

$$\partial_t \mathcal{L}(t, z) = \partial_z \mathcal{L}(t, z) + U (M_*(z) \mathcal{L}(t, z + \omega(z)) - \mathcal{L}(t, z)) + c(t)\mathcal{L}(t, z) \quad \cdot \quad [27]$$

Taking a leading order in λ of this equation yields:

$$\partial_t \mathcal{L}(t, z) = (1 - \mu^2 z^2) \partial_z \mathcal{L}(t, z) + (c(t) - z n \mu^2 / 2)\mathcal{L}(t, z) \quad \cdot \quad [28]$$

This linear PDE can be solved explicitly by any calculus software. The choice of $c(t)$ only affects the total size dynamics ($\mathcal{L}(t, 0) = \langle \mathcal{N}_t \rangle$) in the solution, but not the higher derivatives, i.e. the distribution of fitness, as we could expect. It can also be shown that this first order in λ exactly yields a phenotypic distribution that is gaussian at all times {Martin, 2016 #3207}. This provides a proof of the Lande gaussian model broadly speaking, instead of the gaussian being a conjecture.

The advantage of this WSSM approximation is that it always yields a linear 1st order PDE that can be solved explicitly, yielding a full description of the distribution of the FCs at all times. Another advantage is that we can (in the single FC case) derive a threshold value of the parameter for the approximation to be accurate: $U \gg n^2 \lambda / 4$ roughly.

The drawback is its lack of accuracy when things deviate from the WSSM regime. This inaccuracy depends on the context: as long as the population is not too far away from the optimum, and with more dimensions, the approximation tends to perform better, because it depends less on the fitter tail of the phenotypic distribution and because higher dimensions imply more gaussian-like DFs. In particular with a moving

optimum that changes relatively slowly, the approximation seems to be accurate below its anticipated range of validity, at about $U = U_c$ or even below (ROQUES *et al.* 2020).

In a similar vein, this WSSM approximation also allows to derive explicit probabilities of ER when coupled with a stochastic demographic model (inhomogeneous Feller). This is interesting in that it corresponds to a very polymorphic population, a situation that typically makes it difficult to compute probabilities of long-term extinction from classic tools. Here, polymorphism itself makes the evolutionary dynamics more predictable and only modelling the demographic stochasticity is sufficient to get accurate predictions (this is the central finding of Yoann Anciaux's second paper in annex).

2. Accurate approximation for the SSWM regime that I cannot justify

Consider the single fitness component case (same as eq.[27] above), when starting from a clonal population initially. I have observed that the mean fitness component ($\bar{m}(t)$) is well approximated by the solution of the ODE

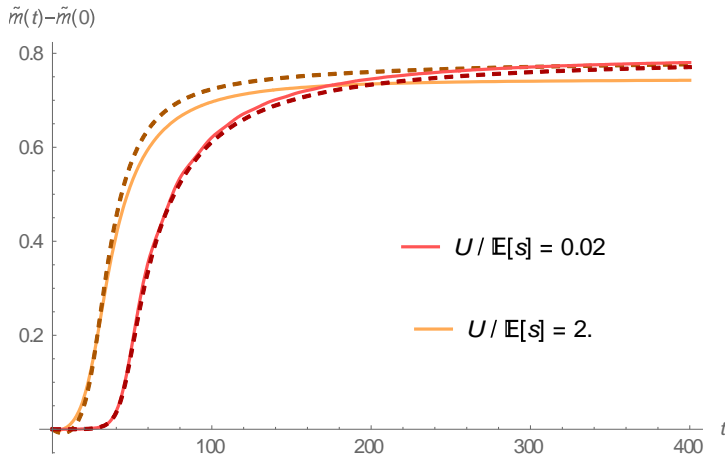
$$\bar{m}'(t) = U \beta^{(1,0)}(t, \bar{m}(t)) \quad [29]$$

where $\beta(z, m) = M_*(z)e^{\omega(z)m}$ and the index gives the derivation order with respect to each variable in $\beta(\dots)$.

Justifying this observation has proven fairly difficult to me. I initially obtained it by making a too optimistic shortcut. I started from the non-epistatic model ($\omega(z) = 0$): $\partial_t \mathcal{L}(t, z) = \partial_z \mathcal{L}(t, z) + \mathcal{L}(t, z)(c(t) + U \beta(z))$ where $\beta(z) = M_*(z) - 1$ and $M_*(z)$ is the MGF associated with some arbitrary (fixed) DFE with finite moments. Under this model {see also \Martin, 2016 #3207}, starting from a clonal population initially ($\mathcal{L}(0, z) = N_0 e^{m_0 z}$), the mean fitness satisfies the ODE: $\bar{m}'(t) = U \beta'(t) = U M_*'(t)$. Based on this, I invoked the following a heuristic. 1) the mean fitness change is fully determined by the current time and the current DFE in the absence of epistasis. 2) in the SSWM regime the effect of standing variance (via epistasis) is negligible so we can consider that epistasis only induces a time dependent mutation kernel, corresponding to the kernel of the current mean fitness. In essence this should be akin to a 'House of Cards' approximation {Turelli, 1984 #1243}. In this case, taking $\beta(z) \equiv \beta(z, \bar{m}(t))$ at time t we would have: $\bar{m}'(t) = U \beta^{(1,0)}(t, \bar{m}(t))$, i.e. eq.[29].

This argument is in fact incorrect, as might be obvious for the reader. However, I did not check my reasoning much because the numerical test of eq.[29] proved very satisfying ! So I believed I had got the argument right. In fact of course, properly taking the House of cards approximation or any time-dependent kernel does not yields eq.[29]. Yet, this simple ODE proves surprisingly accurate (more so than the proper House of Cards it seems), on a wide variety of simulations I checked, as long as U is smaller or of similar order as $\mathbb{E}(s)$ the mean of the DFE (namely for the complementary regime SSWM). I show below an illustration of this accuracy for $U/\mathbb{E}(s) = 2$ (equivalently strong selection and mutation) and $U \mathbb{E}(s) = 0.02$ (SSWM). Here $n = 10$ traits in the

landscape and $\mathbb{E}(s) = 0.02$ per unit time. The plain lines give the mean fitness predicted by the numerical solution of the exact PDE (method of lines) for fitness and the dashed lines give the trajectory from eq.[29].



I have tried quite hard to properly derive this approximation for this HDR but could not manage to do it so far.

For a more complete information here is how I believe the House of Cards (**HC**) approximation (TURELLI 1984) should be implemented in the context of ELT dynamics (or generating function dynamics). Under the HC approximation, the distribution of mutants from the whole population is approximately equal to that expected if all mutants emerged from the mean genotype. This approximation is accurate in an SSWM regime where the distribution of standing genotypes is narrow compared to that of mutants. It has been derived and used so far to describe distributions of phenotypes at mutation – selection balance. However, its principle might be applicable to non-equilibrium distributions of fitness. Here, the idea is to approximate the non-local term in eq.[26] by a HC approximation around the genotype with FC equal to the current mean FC.

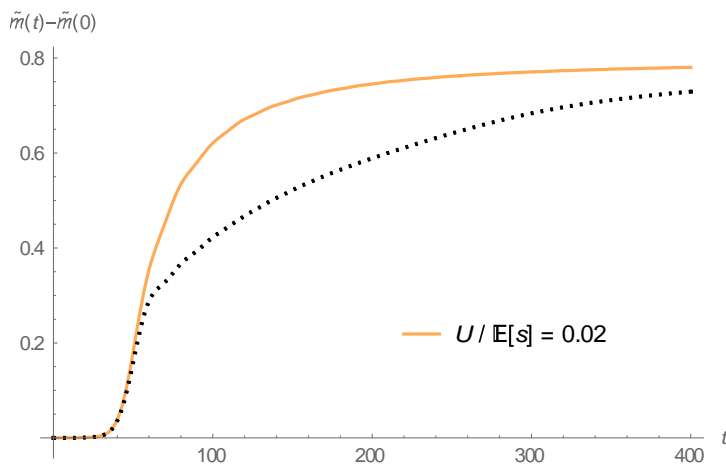
Let $\bar{m}(t)$ be the mean fitness at time t in the population. Under the HC approximation, upon mutating, any background $m \rightarrow \bar{m}(t) + s$. We use a ‘mean field’ approximation where $\langle N_t e^{(z+\omega(z))\bar{m}(t)} \rangle \approx \tilde{N}(t) e^{(z+\omega(z))\bar{m}(t)}$ where we denote $\tilde{m}(t) = \partial_z \mathcal{L}(t, 0) / \mathcal{L}(t, 0) = \langle \bar{m}(t) \rangle$ the ‘ensemble mean’ mean fitness and we recall that $\tilde{N}(t) = \langle N_t \rangle$ is the ensemble mean population size. This ‘mean field’ is exact in the classic deterministic limit of course. The non-local mutational term, which describes the ELT of FCs among mutants, from the whole population, then becomes

$$M_*(z) \mathcal{L}(t, z + \omega(z)) \approx \tilde{N}(t) M_*(z) e^{(z+\omega(z))\tilde{m}(t)} \equiv \tilde{N}(t) \gamma(z, \tilde{m}(t)) \quad [30]$$

$(\gamma(z, m) = M_*(z) e^{(z+\omega(z))m})$. The resulting PDE with selection and mutation becomes

$$\partial_t \mathcal{L}(t, z) = \partial_z \mathcal{L}(t, z) + \tilde{N}(t) U \gamma(z, \tilde{m}(t)) + \rho(t) \mathcal{L}(t, z) \quad [31]$$

where $\rho(t) = c(t) - U$. However, numerical solutions of this PDE yield a fitness trajectory that is much farther from the exact one, compared to eq.[29]. Here is an example in the same SSWM regime as in the figure above (dotted line is the mean fitness trajectory from eq.[30]):



To date this remains a very frustrating puzzle for me. Having a solution that proves accurate and not being able to justify it properly.

The reason I am mentioning this very issue is that the SSWM regime seems to correspond to much more realistic parameter ranges for bacteria (while for some viruses at least the WSSM may be reasonable). Understanding why the simple eq.[29] 'works' would allow to apply it to various extensions where we could only rely on WSSM solutions so far, for explicit results.

I also wonder if the HC approximation could also be used in the context of a backward Kolmogorov equations for multi-type processes. This would serve to derive extinction probabilities with a multitype model with mutation in the FGM. This would provide a correct framework to derive ER probabilities, avoiding the *ad hoc* approximations we have used so far. This is all open and not straightforward to me for the moment.

VII. REFERENCES:

Alexander, H. K., G. Martin, O. Y. Martin and S. Bonhoeffer, 2014 Evolutionary rescue: linking theory for conservation and medicine. *Evol Appl* 7: 1161-1179.

- Alfaro, M., and R. Carles, 2014 Explicit solutions for replicator-mutator equations: extinction vs. acceleration. *SIAM Journal on Applied Mathematics* 74: 1919-1934.
- Anciaux, Y., L. M. Chevin, O. Ronce and G. Martin, 2018 Evolutionary Rescue over a Fitness Landscape. *Genetics* 209: 265-279.
- Anciaux, Y., A. Lambert, O. Ronce, L. Roques and G. Martin, 2019 Population persistence under high mutation rate: From evolutionary rescue to lethal mutagenesis. *Evolution* 73: 1517-1532.
- Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh *et al.*, 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243-U1274.
- Barton, N. H., A. M. Etheridge and A. Véber, 2017 The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* 118: 50-73.
- Bulmer, M. G., 1971 The Effect of Selection on Genetic Variability. *The American Naturalist* 105: 201-211.
- Burger, R., 1991 moments, cumulants, and polygenic dynamics. *Journal of Mathematical Biology* 30: 199-213.
- Chou, H. H., H. C. Chiu, N. F. Delaney, D. Segre and C. J. Marx, 2011 Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science* 332: 1190-1192.
- Depew, D. J., B. H. Weber and E. Mayr, 1996 Darwinism Evolving. System Dynamics and the Genealogy of Natural Selection. *History and Philosophy of the Life Sciences* 18: 135.
- Desai, M. M., and D. S. Fisher, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* 176: 1759-1798.
- Fisher, R. A., 1930 *The genetical theory of natural selection*. Oxford University Press, Oxford.
- Gerrish, P., 2001 The rhythm of microbial adaptation. *Nature* 413: 299-302.
- Gerrish, P. J., and R. E. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* 103: 127-144.
- Gil, M.-E., F. Hamel, G. Martin and L. Roques, 2017 Mathematical Properties of a Class of Integro-differential Models from Population Genetics. *SIAM Journal on Applied Mathematics* 77: 1536-1561.
- Gil, M. E., F. Hamel, G. Martin and L. Roques, 2019 Dynamics of fitness distributions in the presence of a phenotypic optimum: an integro-differential approach. *Nonlinearity* 32: 3485-3522.
- Good, B. H., and M. M. Desai, 2013 Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theoretical Population Biology* 85: 86-102.
- Good, B. H., and M. M. Desai, 2015 The Impact of Macroscopic Epistasis on Long-Term Evolutionary Dynamics. *Genetics* 199: 177-190.

- Good, B. H., I. M. Rouzine, D. J. Balick, O. Hallatschek and M. M. Desai, 2012 Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences* 109: 4950-4955.
- Hallatschek, O., 2011 The noisy edge of traveling waves. *Proceedings of the National Academy of Sciences* 108: 1783-1787.
- Hamel, F., F. Lavigne, G. Martin and L. Roques, 2020 Dynamics of adaptation in an anisotropic phenotype-fitness landscape. *Nonlinear Analysis: Real World Applications* in press.
- Hansen, T. F., 1992 Selection in Asexual Populations - an Extension of the Fundamental Theorem. *Journal of Theoretical Biology* 155: 537-544.
- Harmand, N., R. Gallet, R. Jabbour-Zahab, G. Martin and T. Lenormand, 2017 Fisher's geometrical model and the mutational patterns of antibiotic resistance across dose gradients. *Evolution* 71: 23-37.
- Harmand, N., R. Gallet, G. Martin and T. Lenormand, 2018 Evolution of bacteria specialization along an antibiotic dose gradient. *Evol Lett* 2: 221-232.
- Hazan, R., Y. A. Que, D. Maura and L. G. Rahme, 2012 A method for high throughput determination of viable bacteria cell counts in 96-well plates. *BMC Microbiol* 12: 259.
- Hietpas, R. T., C. Bank, J. D. Jensen and D. N. A. Bolon, 2013 shifting fitness landscapes in response to altered environments. *Evolution* 67: 3512-3522.
- Kendall, D., 1948 On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics* 19: 1-15.
- Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski and T. F. Cooper, 2011 Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science* 332: 1193-1196.
- Kimura, M., 1965 A Stochastic Model Concerning Maintenance of Genetic Variability in Quantitative Characters. *Proceedings of the National Academy of Sciences of the United States of America* 54: 731-736.
- Kishony, R., and S. Leibler, 2003 Environmental stresses can alleviate the average deleterious effects of mutations. *Journal of Biology* 2: 14.
- Kopp, M., and S. Matuszewski, 2014 Rapid evolution of quantitative traits: theoretical perspectives. *Evolutionary Applications* 7: 169-191.
- Kryazhimskiy, S., G. Tkacik and J. B. Plotkin, 2009 The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences of the United States of America* 106: 18638-18643.
- Lande, R., 1976 Natural-Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* 30: 314-334.

- Lande, R., 1980 The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. *Genetics* 94: 203-215.
- Lande, R., 1984 The Genetic Correlation between Characters Maintained by Selection, Linkage and Inbreeding. *Genetical Research* 44: 309-320.
- Lavigne, F., G. Martin, Y. Anciaux, J. Papaix and L. Roques, 2020 When sinks become sources: Adaptive colonization in asexuals. *Evolution* 74: 29-42.
- Lehtinen, J., S. Jarvinen, M. Virta and E. M. Lilius, 2006 Real-time monitoring of antimicrobial activity with the multiparameter microplate assay. *Journal of Microbiological Methods* 66: 381-389.
- Lehtinen, J., M. Virta and E. M. Lilius, 2003 Fluoro-luminometric real-time measurement of bacterial viability and killing. *Journal of Microbiological Methods* 55: 173-186.
- Lenski, R. E., M. R. Rose, S. C. Simpson and S. C. Tadler, 1991 Long-Term Experimental Evolution in *Escherichia-Coli* .1. Adaptation and Divergence During 2,000 Generations. *American Naturalist* 138: 1315-1341.
- Lind, P. A., O. G. Berg and D. I. Andersson, 2010 Mutational Robustness of Ribosomal Protein Genes. *Science* 330: 825-827.
- Lynch, M., and R. Lande, 1993 Evolution and extinction in response to environmental change in *Biotic interactions and global climate change*, edited by P. M. Kareiva, J. G. Kingsolver and R. B. Huey. Sinauer, Sunderland, Mass.
- MacLean, R. C., G. G. Perron and A. Gardner, 2010 Diminishing Returns From Beneficial Mutations and Pervasive Epistasis Shape the Fitness Landscape for Rifampicin Resistance in *Pseudomonas aeruginosa*. *Genetics* 186: 1345-1354.
- Martin, G., 2014 Fisher's geometrical model emerges as a property of large integrated phenotypic networks. *Genetics* 197: 237-255.
- Martin, G., R. Aguilée, J. Ramsayer, O. Kaltz and O. Ronce, 2013 The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments. *Philos Trans R Soc Lond B Biol Sci* 368: 20120088.
- Martin, G., S. F. Elena and T. Lenormand, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat Genet* 39: 555-560.
- Martin, G., and A. Lambert, 2015 A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theor Popul Biol* 101: 40-46.
- Martin, G., and L. Lenormand, 2015a The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution in press*: 1433-1447.
- Martin, G., and T. Lenormand, 2006a The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* 60: 2413-2427.

- Martin, G., and T. Lenormand, 2006b A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60: 893-907.
- Martin, G., and T. Lenormand, 2015b The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution* 69: 1433-1447.
- Martin, G., and L. Roques, 2016 The Nonstationary Dynamics of Fitness Distributions: Asexual Model with Epistasis and Standing Variation. *Genetics* 204: 1541-1558.
- McCandlish, D. M., 2011 Visualizing fitness landscapes. *Evolution; international journal of organic evolution* 65: 1544-1558.
- McCandlish, D. M., and A. Stoltzfus, 2014 MODELING EVOLUTION USING THE PROBABILITY OF FIXATION: HISTORY AND IMPLICATIONS. *Quarterly Review of Biology* 89: 225-252.
- Osmond, M. M., S. P. Otto and G. Martin, 2020 Genetic Paths to Evolutionary Rescue and the Distribution of Fitness Effects Along Them. *Genetics* 214: 493-510.
- Perfeito, L., A. Sousa, T. Bataillon and I. Gordo, 2014 Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* 68: 150-162.
- Rattray, M., and J. L. Shapiro, 2001 Cumulant dynamics of a population under multiplicative selection, mutation, and drift. *Theoretical Population Biology* 60: 17-31.
- Regoes, R. R., C. Wiuff, R. M. Zappala, K. N. Garner, F. Baquero *et al.*, 2004 Pharmacodynamic functions: a multiparameter approach to the design of antibiotic treatment regimens. *Antimicrob Agents Chemother* 48: 3670-3676.
- Remold, S. K., and R. E. Lenski, 2004 Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nature Genetics* 36: 423-426.
- Roques, L., J. Garnier and G. Martin, 2017 Beneficial mutation-selection dynamics in finite asexual populations: a free boundary approach. *Sci Rep* 7: 17838.
- Roques, L., F. Patout, O. Bonnefon and G. Martin, 2020 Adaptation in General Temporally Changing Environments. *SIAM Journal on Applied Mathematics* 80: 2420-2447.
- Schoustra, S., S. Hwang, J. Krug and J. A. G. M. d. Visser, 2016 Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proceedings of the Royal Society B: Biological Sciences* 283: 20161376.
- Sousa, A., S. Magalhaes and I. Gordo, 2011 Cost of Antibiotic Resistance and the Geometry of Adaptation. *Molecular Biology and Evolution* 29: 1417-1428.
- Tenaillon, O., 2014 The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics* 45: 179-201.
- Tenhami, M., K. Hakkila and M. Karp, 2001 Measurement of effects of antibiotics in bioluminescent *Staphylococcus aureus* RN4220. *Antimicrobial Agents and Chemotherapy* 45: 3456-3461.

- Trindade, S., A. Sousa and I. Gordo, 2012 antibiotic resistance and stress in the light of fisher's model. *Evolution* 66: 3815-3824.
- Tsimring, L. S., H. Levine and D. A. Kessler, 1996 RNA virus evolution via a fitness-space model. *Physical Review Letters* 76: 4440-4443.
- Turelli, M., 1984 Heritable Genetic-Variation Via Mutation Selection Balance - Lerch Zeta Meets the Abdominal Bristle. *Theoretical Population Biology* 25: 138-193.
- Turelli, M., and N. H. Barton, 1994 Genetic and Statistical-Analyses of Strong Selection on Polygenic Traits - What, Me Normal. *Genetics* 138: 913-941.
- Vasi, F. K., and R. E. Lenski, 1999 Ecological strategies and fitness tradeoffs in *Escherichia coli* mutants adapted to prolonged starvation. *Journal of Genetics* 78: 43-49.
- Via, S., and R. Lande, 1985 Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* 39: 505-522.
- Wiser, M. J., N. Ribeck and R. E. Lenski, 2013 Long-Term Dynamics of Adaptation in Asexual Populations. *Science* 342: 1364-1367.
- Witz, G., E. van Nimwegen and T. Julou, 2019 Initiation of chromosome replication controls both division and replication cycles in *E. coli* through a double-adder mechanism. *eLife* 8: e48063.
- Wright, S., 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*: 356-366.