

General Introduction

Life, as such does not exist. This was Szent-Gyorgy's (1972) perspective about what is life; *The question, in itself its wrong*, he wrote. The view nowadays, is that we cannot actually define life, since there is no physical or intrinsic component that makes it happen, and rather is conceived that life itself is a process (Mayr, 1982). Lwoff (1965) stated *life is a state of the organisms*, which at first read does not say too much, since organisms are those capable of being alive. However it sets the question in such a way that we can have a working definition of life, that is, listing the properties that we associate with 'living entities' (Maynard-Smith and Szathmáry, 2000, p. 3). These properties are (Maynard-Smith and Szathmáry, 2000): (i) multiplication, (ii) variation, and (iii) heredity (MVH). This definition, as more specific ones, suffers from Sagan's "fundamental handicap of biologists" (Sagan, 1973; Emmeche, 1998), that is that we define life only on basis of the organisms that we know, and that are subject to the same "laws" of evolution, ecology, physiology, genetics, chemistry, etc. Ernst Mayr's (1982) list of properties that define life, picture it as an evolutionary and dynamic process, rather than as an intrinsic property of organisms. Notice that the notion of variability and growth are fundamental elements, as well as that the several definitions of life apply to populations, not to individuals (although of course we can extend the definition that an individual is alive if it descends or belongs to a population that has those properties; Maynard-Smith and Szathmáry, 2000).

These definitions, in particular the above mentioned ones, are a reflection of our conception on how the process of life is. Multiplication (population growth) and variability in populations -the two main subject of this thesis- are fundamental in evolutionary biology, and is no coincidence that they constitute -at least in part- the definition of life. Thus somehow (and this is an existentialist argument) the questions that we ask in evo-

lutionary biology, contribute to build an approximation to this unanswerable question. The abstractions of the vital processes that are used to model biological evolution (in this thesis, mathematically), help to slightly widen our handicapped notions.

For example, what we conceive as a replicator (or for these matters, a reproducer), need not to be as physical as we regard it in an organism, or its hereditary material need not to be genes, nor their composition be of DNA (e.g. memes). The evolutionary models and the modern evolutionary synthesis assume certain properties and explain in terms of processes that do not depend on these biological details. Thus the handicap becomes less pronounced (although it is still there).

Part of this flexibility on our re-interpretation of the evolutionary processes out of our own formulation, is because we are modernly regarding the evolving systems as transducers of information (from a generation to another). Hence the limits of our abstract interpretations can be widened without tormenting ourselves with mechanistic details of a high level of complexity, which is to some extent and from this evolutionary perspective, unnecessary. Essentially, in studying evolutionary processes, we are abstracting what life is by addressing the question *how can this evolve?* in both instances: of a particular biological aspect (e.g. a particular mechanism, a particular trait, etc.), or of a system as a whole (without neglecting the environment, of course). In answering this question, (how can this evolve?) we are forced to invoke MVH (plus other things).

Motivation for this research

Putting aside the philosophical aspects, there are of course more specific question or subjects that we (me and co-authors) will deal with. There are also practical needs to understand how populations adapt to new environments. We will dissect

MVH into separate processes. This is still another simplification, perhaps not the most natural, but the most historically parsimonious.

On the one hand, the techniques for studying population dynamics have improved, and in consequence have improved their applications. We do not consider all populations as logistic and exponential any more. But along developments in non-linear dynamics, and stochastic processes, as well as advances in classical and Bayesian statistical tools for estimation and forecasting, we have had a more comprehensive view of population growth.

On the other hand, a great deal of the approaches to problems in genetics, specially in the *age of molecular biology* came from the Neutral Theory of Molecular Evolution (Kimura, 1985). Much of the ongoing process of constructing the evolutionary synthesis was halted by the arrival of the neutral theory. Along with the central dogma of molecular biology, neutrality ruled the view of world. In the last two decades, with access to so much genetic data, along with the development and application of bioinformatic tools and the study of epigenetic factors, we have found that the molecular world is not neutral, and that there is no linear relation between the genetics and the phenotypes.

So its time to get back to the basics. We need to understand the dynamics of populations in a better -perhaps more fundamental- approach. Not necessarily to construct more complicated models (although why not!), but to understand what the simple models have given and why, and how this relates to the genetic structures in a population. These are not separated problems. We just view them like that. The study of the roles of selection in natural and artificial populations has also been of great importance in the last years, since have given further insight in the processes of evolution.

Nowadays biologists are more prepared for integrative and fundamental questions, since the use of mathematical and computational tools have (fortunately) permeated to even the most practical aspects of biology, and even triggered the creation of new mathematical tools.

This thesis aims to complement the theories of population dynamics and quantitative genetics. Seeking a synthetic approach to each of these two subjects, where there is plenty of research to discover new facts but not so much to bring them together.

The first subject: Population dynamics (PD) is among the oldest subjects in theoretical and mathematical biology, dating back to Malthus (1798); Gompertz (1825), and Verhulst (1838), and originated in the study of demography. We don't concisely know how to 'derive' population dynamics from first principles. But nowadays it conforms to one of the central elements of the theory of evolution, both as a subject to study and as a tool to study other subjects.

As Gilpin and Ayala (1973) put it: "*Biology is at a Keplerian stage*"; this was more than three decades ago, and much advance came until today. We could say that nowadays it is rather at a Newtonian stage⁴: there are many ways populations grow, or "growth laws"; some of them have mechanistic explanations, some of them only phenomenological justifications. But the applications of these dynamics is widespread in biology. However there is not a consistent theoretical background that leads to the understanding why populations grow in particular ways.

This is of course not coming as a surprise. If we give it a thought, there are so many factors that determine growth that it is hard to think where to start. For the sole purpose of exer-

⁴Although I. Pen suggests is rather a Laplacian stage.

cising I made a mind map of the factors that I would consider to directly affect growth. The result, was a long list of factors which besides growth, affect each other in a fully connected network fashion. Not too good for a start. But it reflects that distinct disciplines have considered the action and effect of biotic and abiotic factors on growth, so there are several pieces of the puzzle. But how growth has been tuned by the undergoing evolutionary process, is not a question with a trivial answer. There are two classes of hindrances in the search for this answer. From a perspective, and to the main concern in this thesis, the way in which we model population growth is to some extent arbitrary. This has resulted in having a battery of models (Henle et al., 2004a) that are 'adapted' to their use in particular problems. This is interesting, in that it might reveal that the diversity of growth strategies is big. Often these models are chosen by distinct criteria, and do not necessarily reflect biological factors that might be of relevance. Furthermore, it is common to find distinct models that result in very similar –or identical– growth patterns, but which have radically different biological implications. But there is no criterion that suffices for a choice of biological significance. Even if we were to apply statistical methods for model selection, a set of hypotheses is likely to be biased by mathematical easiness. Between lines, I am assuming a reductionist position. The problem might well be how to better 'explain' a growth pattern with minimal set of parameters. But I am referring first, how to understand the factors determining growth, and second, how evolution can shape these factors to result in an evolutionarily stable growth strategy.

There is plenty of work on the evolution of population growth, and how populations adapt to particular conditions. I identified five main trends in the study of the evolution of growth strategies. I am considering models that do not take into account

competition (e.g. Lotka-Volterra types are excluded), and these may or may not comprise age structure. I will briefly go over them, and bear in mind that this classification is arbitrary and the classes are not exclusive. First, following the unexpected results that the simple discrete logistic equation shows complex (chaotic) behaviours (May, 1976), there was a rush to study this new phenomenon, and its consequences. To the big regret of many, studies revealed that in most cases, evolution would tend to tune the dynamics in such a way that they result in a stable equilibrium (Doebeli and Koella, 1995; Ebenman et al., 1996; Schliekelman and Ellner, 2001), although some special conditions would allow the chaotic dynamics (Ferreire and Gatto, 1993; Gatto, 1993; Doebeli and Koella, 1995; Gonik et al., 2005). Second, the influence of stochastic factors on population growth has been a stereotypic model with vast applications (Tuljapurkar, 1990; Lande et al., 2003). In this context, it has also been studied how distinct strategies would evolve to cope with these fluctuating realms (Tuljapurkar and Orzack, 1980; Tuljapurkar, 1982; Orzack and Tuljapurkar, 1989; Yoshimura and Jansen, 1996). Third, the most biologically comprehensive approach is of optimizing growth rates that are determined by life-history traits. To begin with, the proper fitness measure has been debated (Murray, 1997; Metz et al., 1992; Rueffler et al., 2006), and it seems that optimizing the Malthusian parameter (exponential growth rate) is the most consistent option. To follow, the dynamics are coupled to the evolutionary benefits of individuals that maximize their fitness (Metz et al., 1992; Mylius and Diekmann, 1995; Coulson et al., 2006; Pelletier et al., 2007). This approach is very versatile, and allows to modeling of specific situations for which biological details are included with easiness (Orzack and Tuljapurkar, 1989; Charnov, 1993; Shertzer and Ellner, 2002; White et al., 2006). Fourth, ecological variables and the spatial structure also de-

termine growth rates of populations (Lion and van Baalen, 2008). Fifth, there are genetic determinants to growth (Travis and Greenwood, 1990; Hastings and Harrison, 1994; Doebeli, 1996b; Doebeli and de Jong, 1999) which may be assumed to act directly on growth rates, or indirectly on any other life-history traits.

Most, if not all of these approaches employ specific growth models, parameters of which are tuned by evolution. A classical example, is the logistic model, and the notions of r and K selection. This analysis presumes that growth is logistic, and the evolutionary reasoning sets the details on the values of these parameters (MacArthur and Wilson, 1967; Pianka, 1970). But any alternative strategy that would result in growth dynamics different from logistic, is of course disregarded. A synthetic approach, is missing (though see Metz et al., 1992; Meszéna et al., 1992; Page and Nowak, 2002). Thus we need a way to understand the growth of populations from a wider view. That is to understand what do the different growth patterns have in common. There are two common assumptions, which are (i) growth at low densities is approximately exponential, and (ii) growth at high densities is limited. From the mathematical side, this is equivalent as considering the first two terms in a series approximation of *the* growth law. So naïvely we could adopt the third term, and so on. Surely we can do better than that. Indeed a notable advance was initially achieved by Ayala et al. (1973); Gilpin and Ayala (1973); and Gilpin et al. (1976). They introduced a model that accounted for the internal competition of a population. This has been later applied to study global patterns of growth (Sibly et al., 2005). Although not yet in the evolutionary context, this application of a generalized model reveals another dimension of evolutionary possibilities. How much can we extend this, without invoking more parameters or artificial models? This is the research subject of the first part of this thesis.

The second subject: the theory of quantitative genetics (QG) aims to explain the evolution of quantitative traits and characters based solely on measurable quantities. That is, without making reference to variables that we can not measure, like allele frequencies, genetic effects, or number of loci. The theory of population genetics (PG), on the other hand, studies the evolution of the frequency of alleles, and of the genetic effects of allelic substitutions over phenotypic traits. Still, the relation between both approaches is obscure. Whether it is possible or not to bridge both, we still don't know, but quantitative genetics relies on the ansatz or conjecture that it is.

Mechanistic approaches from population genetics have, in a sense, failed to achieve this bridge. Only some approximations have been fruitful for specific situations. Still it looks from the experimental and empirical perspective that it is indeed possible. Thus the subject persists.

Why do we need at all the integration of both sub-disciplines of genetics? In *The Origin of Species* Darwin (1859) recognized that the mechanisms of natural and artificial selection were essentially the same. Artificial selection is applied on phenotypic traits, and in a sense ignores what is behind it, in the genetic composition (whose nature was at the time unknown to Darwin). This points out the primary role of understanding the nature of selection. Even when mathematical tools started to be applied to compute the response to selection, the approach was entirely phenotypic (Pearson, 1896), and to certain degree, it has remained like that. It was not until Fisher (1918) when the relation between the Mendelian nature of phenotypic traits under selection was addressed.

But Kimura's (1985) *Neutral Theory of Molecular evolution* went totally in the opposite direction. It required molecular data, and assigned the major evolutionary cause to point mutations and random drift, rather than to natural selection. As

the saying goes, the truth is the intersection of two independent lies, hence given that these three factors, selection, mutation, and drift (SMD) are potential causes for the evolution of virtually any trait, we are thus interested in their relative importance to generate and maintain the diversity that nowadays exists in populations.

It is actually not possible that phenotypic traits change without consequent evolution of their genetic composition. But as breeders have shown, it is possible to predict –to a certain degree– the average values of the offspring’s traits of breeding individuals selected for a trait. Thus it is clear that there are immediate applications. Yet, this will only work for some generations, thus for purposes of understanding and explaining biological diversity, these predictive capabilities are not enough. This is because the predictability of the trait values in a population depends on the amount of variation that is available at breeding time. And this variation is changing. How is this change? *That is the question.*

Predicting this change of the genetic variance is not a trivial matter, since it depends on several biological factors, many of which are not measurable quantitatively. PG plays a role here, indicating the factors that influence the evolution of the genetic variance. The down side, is that it necessarily depends on these non-observable elements (Barton and Turelli, 1987).

A somewhat more realistic situation is the various traits co-evolve. Lande (1979) studied this scenario, showing how the genetic co-variances (\mathcal{C} -matrix) are influenced by pleiotropy (Lande, 1980). In general, linkage, epistasis, (Turelli and Barton, 1994), environment, sexual selection (Barton and Turelli, 1991; Turelli and Barton, 2004), and other genetic and ecological causes (Arnold et al., 2008) will affect the co-variant structure of any trait.

It is possible to study this situation in a bottom-up fashion,

that is considering the evolution of allelic effects, and how it leads to a change in the traits. But it is not easy to identify these elements from quantitative data alone, in order to give a fulfilling explanation of evolutionary response from a QG approach alone. Thus we are in need of a way to relate the non-observable factors to the observable variables. Then it might be possible to predict the long term evolution of quantitative variables with accuracy.

Although this remains an important subject, which we will address in the second part of the thesis, we might still wonder why the question of the integration of PG and QG is important, given the advances in molecular biology techniques, which allow us to screen the genetics effects over any kind of trait.

First, of course is the fact that these empirical analysis have certain limitations. Recent theoretical studies (Sella and Hirsh, 2005) have shown that even if an equilibrium between SMD is maintained, the rates of molecular substitution are equal. This is a result which was reserved to, and interpreted as, neutrality. The implications are not yet studied. But clearly we might be missing something. The easiness with which sequence data is analyzed under a neutral model assumption (Li, 1997) might prove misleading, compared to the view where selection is considered.

Second, the identification of quantitative trait loci (QTL) is of major relevance to quantitative genetics. It gives an idea of the amount of loci that might be contributing to the quantitative trait and its variation, as well as their effect. However, QTLs have resolution to discern only those loci of major effect. Furthermore, the technique is able to identify only two alleles. Thus we still are uncertain of the number of loci which are actually contributing, and the overall effect of the contribution of those loci with smaller effect over the quantitative variables. This is critical, since many models assume a contin-

uum of alleles and/or infinite number of loci (Kimura, 1965a; Bulmer, 1972; Lande, 1975; Kingman, 1978; Bulmer, 1980, all reviewed by Turelli, 1984). In these cases, the distribution of the trait will be essentially Gaussian (Turelli and Barton, 1994). Although this situation in practice is that with few di-allelic loci (say five to eight, as will be shown in chapter 5) normality is already a good approximation. In addition, the estimation of the QTL effects are a statistical matter. Xu (2003) has shown that the sampled population size might substantially bias the estimation of the effects (the *Beavis effect* Beavis, 1998), and its likely that many QTLs that have been reported suffer from this oversight. We know that the distribution of genetic effects is highly skewed, as stated above, with many QTLs of small effect and few with large effect. The problem of underestimating the variation contributed by the many loci with small effect, is that these will actually compensate the variation that is rapidly lost by selection on those loci of high effect (Barton and Turelli, 1989; Barton and Keightley, 2002), so forecasting of evolutionary response is dependent on these underestimated genetic elements.

Third, the architecture of the trait, that is how the allelic effects affect simultaneously and non-additively different traits, plays a crucial role in the response to selection (Orr, 2000; Cheverud, 2006; Wagner et al., 2008). For simplicity, most works assume that the contribution of the genetic effects over a trait can be decomposed into the additive and non-additive factors. The former just adds the effects of all the genes contributing to the trait, whereas the latter comprises the interacting factors among all these (or other) genes. With this division in mind, it is possible to make appropriate design to identify QTLs not only for the traits, but also for their interaction (Lynch and Walsh, 1998; Cheverud, 2000). Yet the nature of these interactions is uncertain (Hansen, 2006). Theoreticians typically

assume the influence over the traits comes out of pairwise epistasis (Kondrashov and Turelli, 1992; Turelli and Barton, 1994; Carter et al., 2005). In any case, the issue is that the complex essence of evolving traits, or alternatively, the complex background on which additive or non-additive traits evolve, will by all instances affect the change of genetic variation (Gavrilets and de Jong, 1993; Goodnight, 1995; de Brito et al., 2005), and the knowledge on how this variation will change, is by no means obvious even when knowing pleiotropic and epistatic QTLs.

Fourth, from the genetic point of view it is ambiguous to gauge the evolutionary causes of quantitative characters. Not only for the reasons above, but also because the action of selection at the level of phenotypes might be of a distinct nature than selection at a given locus (for example, selection for a specific amino acid in a protein, Yang and Swanson, 2002; Yang et al., 2005; promoter regions, Haygood et al., 2007; Kawabe et al., 2007; or any other specific molecular unit Rand, 2001; Hoede et al., 2006; Haddrill et al., 2008; Kim and Wiehe, 2008). If selection is acting over these genetic elements, or if selfish genes are inducing genetic conflict (Werren et al., 1988; Hatcher, 2000), at the same time as they are affecting traits under selection, then the net effects of selection over each of these two units will be hard to discern on a particular locus. This kind of effects, consonant with the theory of multilevel selection (Okasha, 2006), have not been studied for the evolution of polygenic traits. Nevertheless, it will not be long before this happens.

To conclude, personally I think that this issue should be interpreted in the opposite direction. That is, how can this genetic information help us to refine the synthesis of PG and QG. Combining these two fields in one thinking seems to give much more than what each field give on its own. The availability of genetic and molecular data helps to refine the quantitative

studies, since it allows a clear view of the assumptions that we can actually make in achieving a successful quantitative theory that can support the empirical facts.

It is recognizable that these are two classical subjects of the 'modern evolutionary synthesis' which are still debated and require completion, and it is desirable that their foundations are solid, so they can support the study of complex interactions in the micro and macro scales. subjects to which in the meantime we are moving further.

Guide to this thesis

The thesis is divided into two parts and a synthesis. Each part contains some chapters with the original results. These are published or (almost) submitted for publication in peer-reviewed journals. They are followed by perspectives chapters that includes research that has not yet been published for distinct reasons, and speculations about the future prospects of the results. I will end with a synthesis, which builds the 'big-picture' of my results for both the specific subjects, as well as the integrative view of them.

Part I. Population dynamics As discussed above, in PD we find a variety of models that typically describe various phenomenologies, yet a full integration of these models is absent. We pursue such an integrative theory, for a class of growth patterns which are common in the literature of PD, and that describe non-interacting populations. This consolidation is achieved by following the dynamics not only of population size, but also of the per-capita rate. Surprisingly, at least for the class of density dependent patterns that are studied in Chapter 1, the per-capita

growth rate is independent of the population size and carrying capacity, yet describes patterns of density dependence. The carrying capacity is determined by the initial growth rate of the population, thus it is rather a consequence of a population's trait, rather than a purely environmental property. (Needless to say, the initial rate is by no means independent from the environment!).

In Chapter 2, environmental stochastic effects on the growth rate are considered. A first prediction is that even those growth patterns that would lead to a population explosion are controlled by the stochastic effects. Infinite-sized population are thus avoided without the need to invoke, but not excluding the presence of, carrying capacity. But there is a second prediction, much more quantitative. The patterns of growth will result in a logistic form, irrespective of the deterministic properties. The equilibrium will be attained at random population densities. Even if the deterministic population is logistic, its carrying capacity will not be a predictor of the equilibrium density.

Part II. Population Genetics PG describes the mechanisms of QG, but it is not entirely clear whether the phenomenological framework of QG can be derived from PG. The subject is full with fuzzy results, with few (or limited) final statements. In Chapter 3 this question is undertaken. This is achieved by considering not the quantitative variables themselves but expectations of these. The distribution used to compute the expectations is obtained through the maximization of entropy (ME), restricted to the quantitative observable variables. This distribution coincides with the exact solution to the diffusion approximation in equilibrium, and approximates very well the distribution when evolution is taking place. (This method is equivalent, or analogous to the coupling of statistical mechanics between microscopic and macroscopic variables). The main applications

in this chapter is to polygenic traits of arbitrary number of diallelic loci with distinct effects, but the applications to other situations, like stabilizing selection and other schemes inducing epistatic effects is also addressed.

But polygenic traits tend to be correlated among each other. In Chapter 4 the results of the previous one are extended to include pleiotropic effects. Under these circumstances, the evolutionary dynamics involve the genetic co-variances matrix, \mathcal{G} , which contains in every entry the genetic covariance between any pair of co-evolving traits. We can thus, as above, calculate the expectancies of the genetic co-variances. In this chapter the analyses are not restricted to theoretical constructs, but are employed to analyse previously published data on *Rana temporaria*, for which differences of \mathcal{G} have been quantified across two different locations on four covariant traits. We contrast the results to the scenarios where we assume that directional or stabilizing selection is maintaining the observed genetic variability.

These specific results from the ME method, are accompanied by an interesting conceptual system. These ideas are explored in Chapter 5. The evolution of expected values are of a different nature and we must think of them in different terms than of the quantitative (which are stochastic) variables themselves. These are of course not unrelated. We discuss and apply the concepts above in relation to fitness landscapes, and how to retrieve information of n genetic variables (allele frequencies) using only m quantitative variables, where $n \gg m$. Another result concerns an extension of Fisher's Fundamental Theorem of Natural Selection in the expected values of mean fitness, which will always increase, under the effects of selection and drift. This is an extension that appears only at a statistical level. We also discuss the possibility of employing the ME framework to test distinct hypotheses of the mutation-selection-drift circumstances that

determine the structure of the G matrix. In particular we apply it to contrast the scenarios where directional and stabilizing selection are determining observed empirical \mathcal{G} 's

In Chapter 6, I come back to correlated evolution, but from the genetic side. The ME estimations show that the \mathcal{G} matrix is very much constrained by pleiotropic effects. Although selection will induce change in the mean trait, the changes in the \mathcal{G} matrix are delayed to latter stages, perhaps hundreds of generations (for typically low mutation rates), remaining practically constant until then. The effects of apparent stabilizing selection and of the amount of pleiotropic loci over \mathcal{G} 's evolution are also studied.

As a last pivotal example, the extension of the theory to stabilizing selection (SS) is presented in detail (chapter 7). The situation is more complicated in that SS is inherently non-linear over the trait under selection. Thus introducing mathematical complications that are limiting. We investigate ways on how to overcome these technical difficulties. The case of directional selection is revisited, with allelic effects with dominance, which can be viewed as stabilizing (or disrupting) selection over each locus.

In the appendices it is shown that maximizing entropy with constant fitness is equivalent to maximize fitness at constant entropy. A discussion about the analogy between statistical mechanics in physics and the ME methods herein presented, is addressed. And finally a battery of the most important formulas for the quantitative variables, in all the above situations is provided.

