



HAL
open science

Hierarchical bayesian modelling for saccharomyces cerevisiae population dynamics

Aymé Spor, Christine Dillmann, Shaoxiao Wang, Dominique de Vienne,
Delphine Sicard, Éric Parent

► **To cite this version:**

Aymé Spor, Christine Dillmann, Shaoxiao Wang, Dominique de Vienne, Delphine Sicard, et al.. Hierarchical bayesian modelling for saccharomyces cerevisiae population dynamics. International Journal of Food Microbiology, 2010, 142 (1-2), pp.25-35. 10.1016/j.ijfoodmicro.2010.05.012 . hal-01197603

HAL Id: hal-01197603

<https://hal.science/hal-01197603>

Submitted on 31 May 2020

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.



Hierarchical Bayesian Modelling for *Saccharomyces cerevisiae* population dynamics

Aymé Spor^{a,*}, Christine Dillmann^a, Shaoxiao Wang^{a,b}, Dominique de Vienne^a,
Delphine Sicard^a, Eric Parent^c

^a Univ Paris-Sud, UMR 0320/UMR 8120 Génétique Végétale, F-91190 Gif-sur-Yvette, France

^b Department of Biochemistry and Molecular Biology, Louisiana State University Health Sciences Center, Shreveport, LA 71130, USA

^c UMR MIA 518, INRA/AgroParisTech, ENGREF, 19 avenue du Maine, F-75015 Paris, France

ARTICLE INFO

Article history:

Received 10 March 2009

Received in revised form 6 April 2010

Accepted 14 May 2010

Keywords:

Hierarchical Bayesian Modelling

Saccharomyces cerevisiae

Population dynamics

ABSTRACT

Hierarchical Bayesian Modelling is powerful however under-used to model and evaluate the risks associated with the development of pathogens in food industry, to predict exotic invasions, species extinctions and development of emerging diseases, or to assess chemical risks. Modelling population dynamics of *Saccharomyces cerevisiae* considering its biodiversity and other sources of variability is crucial for selecting strains meeting industrial needs. Using this approach, we studied the population dynamics of *S. cerevisiae*, the domesticated yeast, widely encountered in food industry, notably in brewery, winery, bakery and distillery. We relied on a logistic equation to estimate the key variables of population growth, but we took also into account factors able to affect them, namely environmental effects, genetic diversity and measurement errors. Our probabilistic approach allowed us: (i) to model the dynamical behaviour of strains in a given condition under some uncertainty, (ii) to measure environmental effects and (iii) to evaluate genetic variability of the growth key variables.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Biological data, whatever the field of research, are mostly dynamical or spatial, *i.e.* they are function of time and/or of spatial coordinates. The challenge of the biologist is to explain the variations of traits by the variation of explanatory variables. Whenever the observations are collected at different time or space points from the same biological sample, they become dependent because they are jointly related through time or space. Statistical analysis consists in modelling (depicting the various sources of variation) and inference (estimating the parameters of the model). Historically, statistical analysis has been developed from a “frequentist” point of view: the parameters are considered to have a fixed value, and estimates of this value are searched *via* various statistical procedures of inference (moment adjustment, maximum likelihood estimates, etc.). Most statistical toolboxes that are available to biologists are designed according to the frequentist approach. However, they are generally restricted to the analysis of linear models, *i.e.* to the cases where the response is linear through time and/or space. When dealing with non-linear processes, the problem becomes much more complex, and requires sophisticated statistical tools which are usually not mastered by the biologists. As a result, they have to use non-optimal methods

for properly analyzing or even simply detecting differences between two curves, for instance population growth curves.

The Bayesian approach is another way for analyzing biological data. Given uncertainty on parameter values, a so-called prior probability distribution is assigned to the parameter in a modelling step, taking possibly into account previous knowledge on the parameter. Bayesian inference can be interpreted as formulating a probabilistic judgment about the unknowns of the model given the observed data (updating the prior into a posterior distribution). Because Bayesians traditionally put more emphasis in the modelling process, the Bayesian statistical framework provides an easy way of thinking about biological problems. Unlike the frequentist estimation techniques, dealing with complex models (non-linearity, dependence) does not bring much additional difficulties to the Bayesian inferential algorithms.

Hierarchical Bayesian Modelling (HBM) is a probabilistic, adaptable and efficient framework for modelling dynamical processes by taking into account multiple sources of variation. This type of model is not restricted to specific problems and can be generically applied to a vast extent of dynamical and spatial systems. Hierarchical statistical modelling has the potential to match high dimension problems through conditional decomposition into a series of probabilistically linked simpler substructures (Clark and Gelfand, 2006). Hierarchical statistical models are made of three “layers” (Wikle, 2003). First, an experimental data level specifies the distribution of the observables at hand given the parameters and the underlying processes. Second, a latent process level depicts the various hidden biological mechanisms that make sense of the data. For example in this article, the latent

* Corresponding author.

E-mail address: ayspor@gmail.com (A. Spor).

process level describes the population growth process with a logistic model. Third, a parameter level identifies the fixed quantities that would be sufficient, were they known, to mimic the behaviour of the system and to produce new data statistically similar to the ones already collected. Sources of variation on the parameters can be added: some authors (Bernier et al., 2000; Nauta, 2000, 2002) make the distinction between “variability” (i.e. uncertainty by essence that cannot be reduced by additional information) and “uncertainty” (or uncertainty by ignorance that should decrease as the sample size increases). In this paper, variability will describe changes with respect to biotic or abiotic variation, while uncertainty accounts for measurement errors and model imperfections (Shorten et al., 2004).

In predictive microbiology, HBM has been particularly used for risk assessment and food shelf-life estimation. It is convenient to predict pathogenic bacterial behaviour in case of contamination, because it makes it possible to quantify separately the effects of environmental factors (temperature, pH and available resources), genetic variation and measurement uncertainty. Some population dynamic model using Bayesian inference are now available and describe for example the effect of temperature on growth of *Listeria monocytogenes*, *Salmonella*, *Escherichia coli*, *Clostridium perfringens* and *Bacillus cereus* on pork meats, milk, seafoods or egg products (Delignette-Muller et al., 2006; Membre et al., 2005; Pouillot et al., 2003) or the development of a growing population of bacterial cells from an inoculum of dormant spores (Barker et al., 2005).

In ecology, this probabilistic framework is increasingly used to examine population dynamics because it can take easily into account multiple sources of stochasticity (such as space, time and individual heterogeneities), while in standard statistical models, only process errors are routinely included (Calder et al., 2003; Clark, 2003). HBM has much to offer, including more precise parameter estimation (Calder et al., 2003), and it becomes more and more used to predict exotic invasions, extinction risk or development of emerging diseases. For example, HBM has been successfully implemented to model the invasive Eurasian Collared-Dove dynamics (Hooten et al., 2007), to estimate species richness and spatial occupancy (Kéry and Royle, 2008), the various failures in a Dynamic Energy Budget mechanism for ecotoxicological Daphnid data (Billoir et al., 2008) or to predict the relative abundance of House Finches over the eastern United States (Wikle, 2003).

The yeast *Saccharomyces cerevisiae*, a common biological model in genetics, genomics and physiology, has been exploited since Neolithic period to produce fermented beverages and bread dough. Because of the consumers' reluctance about genetically modified organisms, it seems unrealistic to improve strains by genetic engineering. Another strategy is to exploit present natural biodiversity of yeast, which requires characterizing strains, searching for suitable physiological traits for industrial purposes, and planning genetic resource management because it is not possible to give the same maintenance effort to all strains.

Bakers need to develop strains with hyper-osmolarity resistance, brewers strains with high fermentation rates and short lag phases, and oenologists strains tolerant to ethanol for completing fermentation (Boekhout and Robert, 2003). These different properties have been shown to be related to population dynamic characteristics and to interact with the environment. The population dynamics of *S. cerevisiae* depends both on the genetic background of the strains and on environmental factors such as temperature (Beltran et al., 2002) or glucose content of the medium (Spor et al., 2008). The latter study demonstrated a strong impact of the food-processing use of strains on population dynamic key variables (Spor et al., 2008). Similarly, Domizio et al., 2007 described a close relationship between wine attributes and *Saccharomyces spp* population dynamics. Thus, predicting population growth and modelling genetic and non-genetic variation would help for yeast genetic resource management and for selecting industrial starter strains.

We used HBM to describe *S. cerevisiae* population dynamics. The experimental data consisted in population size counts over time for 12 *S. cerevisiae* strains grown in three culture media. The latent process relied on a logistic equation depending on three population parameters, which divides the population growth into two phases, an exponential growth from an initial population of size N_0 with an intrinsic growth rate r , followed by a decrease of the population growth which leads to a stationary phase, characterized by a maximum population size K , also called carrying capacity in ecology. The latent process model described differences in these key variables with respect to both environmental effects (glucose content in the culture medium) and genetic variation between strains. Finally, the uncertainty related to measurement errors was described.

2. Materials and methods

2.1. Principle of the Bayesian inference

Bayesian inference, or model learning, is the process of updating prior beliefs about unknowns by probabilistic machinery based upon the relationships in the model and the observations recorded about the situation.

By contrast with the ‘classical’ approach, which begins with a hypothesis test that proposes a specific value for an unknown parameter θ , Bayesian inference proposes a prior distribution $p(\theta)$ for this parameter which represents the beliefs originally encoded in the model. Data x_1, x_2, \dots, x_n are collected and the likelihood $f(x_1, x_2, \dots, x_n | \theta)$ is calculated given the parameter values (as in the frequentist case).

Then the probabilities of all the other variables that are connected to the variable representing the new data are updated. Bayes's theorem is used to calculate the posterior distribution $g(\theta | x_1, x_2, \dots, x_n)$. After inference, the updated probabilities reflect the new levels of belief (or probabilities of) all possible outcomes encoded in the model.

2.2. Data

The experimental data used to develop this model have been published in (Spor et al., 2008). Strain origin, culture medium composition and population size measurements are detailed in the Material and method section of Spor et al. (2008). To sum it up, 12 strains stemming from three industrial origins (winery, brewery and bakery) were grown in three media differing by their glucose concentration (0.25%, 1% and 15%). Every two hours samples were taken, diluted and plated to estimate population size. Three biological replicates were performed for each medium-by-strain combination, each time starting with a new inoculum. The population size was expressed in CFU/mL (Colony Forming Units). The experimental data are also called “observations” in the Bayesian setting.

2.3. Model

Our aim was to construct a population dynamic model capable to correctly predict the population size $N_{s,m,t}$ of strain s in medium m over time t . Fig. 1 illustrates the corresponding Directed Acyclic Graph that points out the conditional dependence between nodes. In this framework, parameters and observations could either be considered as logical or stochastic nodes of the model. Logical nodes correspond to nodes that are deterministic functions of other nodes, and stochastic nodes correspond to nodes that are described by probabilistic laws. The description of the nodes is given in Table 1.

2.4. Description of the latent process

We assumed that *S. cerevisiae* population growth follows a logistic equation. This equation is classically used in ecology to model microbial as well as animal population dynamics, and is central in

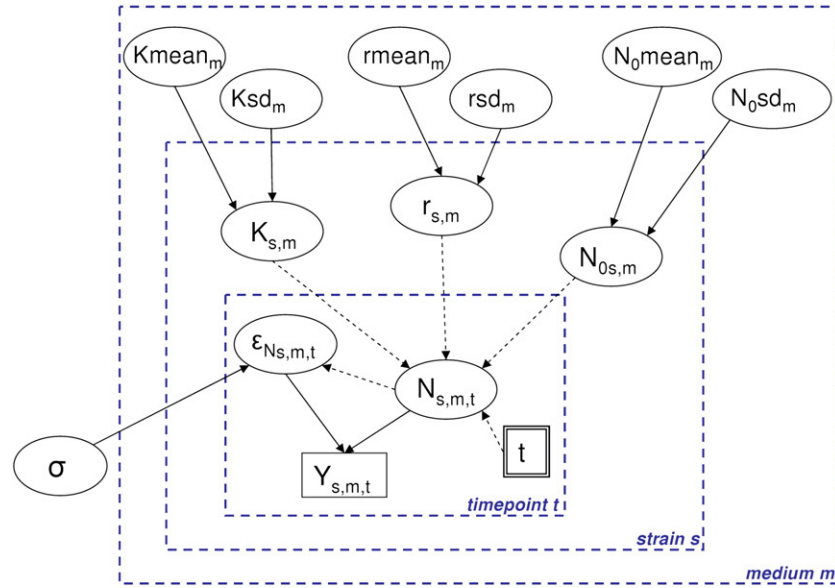


Fig. 1. Directed Acyclic Graph of the model (DAG). Data ($Y_{s,m,t}$) are denoted by rectangles; covariates by double rectangles (t) and latent variables by ellipses. Solid arrows correspond to stochastic dependences between nodes while broken arrows indicate logical link. Dotted blue rectangles illustrated the embedded levels of the modelling (timepoint, strain and medium).

the mathematical definition of the famous ‘ r ’ and ‘ K ’ strategies in ecology (MacArthur and Wilson, 1967). Two types of logistic equation could be considered: with or without lag-phase. Because fresh medium was inoculated after an overnight pre-culture, we used the logistic model without lag-phase. Thus population size followed:

$$N([K, r, N_0]_{s,m}, t) = \frac{K_{s,m} N_{0s,m} e^{r_{s,m} t}}{K_{s,m} + N_{0s,m} (e^{r_{s,m} t} - 1)} \quad (1)$$

where $N([K, r, N_0]_{s,m}, t)$, the population size at time t , depends on the variables K , r and N_0 of strain s in the medium m . K is the carrying capacity (maximum population size) expressed in CFU/mL, N_0 is the initial population size also expressed in CFU/mL and r is the intrinsic growth rate (equivalent to the maximum rate of increase of the population, in min^{-1}).

2.5. Sources of variability on parameters

Our aim was to estimate posterior distributions of the latent key variables K , r and N_0 for each strain in each glucose condition. In the literature of system analysis, there are commonly named population dynamic parameters, which turns to be a rather inappropriate term in a statistical modelling framework since, contrary to statistical parameters, they vary as latent (i.e. unobserved) random variables depending on factors of explanations or grouping of data. Modelling the variability consists in defining how the environment, as well as the genetic variation between strains, would affect population dynamic key variables. In this context, variations can be described

by normal distributions $N(\mu, sd)$ defined by two parameters, the mean μ and the standard deviation sd . A fixed effect would be an effect that changes the mean μ of a latent variable K , r or N_0 , while a random effect would change their standard deviation sd . This introduces additional correlations between individuals of the same group, i.e. individuals of the same strain. The degree of resemblance will be tuned by the standard deviation sd of the latent variable.

In our case, two sources of variation could affect population dynamics: the glucose concentration in the medium and genetic differences between strains. As each culture condition may affect yeast population dynamics in a specific manner, the medium effect was considered to be fixed and was described by a mean value for each parameter $Kmean_m$, $rmean_m$ and N_0mean_m in each glucose condition m . Note that in the case of N_0mean_m , there is no causal relationship between the glucose content of the medium and this parameter. However, the experiments in the 15% glucose condition were performed by a different experimenter from those performed in the 1% and 0.25% glucose conditions. The variation of the N_0mean_m parameter represents therefore the variation of the inoculum conditionally to the experimenter. The mean values of the population dynamic latent variables K , r or N_0 were assumed to be the same for all strains in a given glucose condition. The differences between strains were considered as a genetic random effect, statistically described by the standard deviations Ksd_m , rsd_m , N_0sd_m of the normal distributions.

Mathematically, for each strain s in each medium m , we chose to draw the latent key variables (K , r and N_0) in independent normal distributions with parameters $Kmean_m$, $rmean_m$ and N_0mean_m as expected values. The other parameters, the standard deviations Ksd_m , rsd_m and N_0sd_m , rule the range of variation for the variables around their mean in each glucose condition m :

$$\begin{aligned} K_{s,m} &\sim N(Kmean_m, Ksd_m), \\ r_{s,m} &\sim N(rmean_m, rsd_m), \\ N_{0s,m} &\sim N(N_0mean_m, N_0sd_m). \end{aligned}$$

In other words, there is a random effect $\eta_{Ks,m} = K_{s,m} - Kmean_m$, corresponding to the different behaviours of two strains s and s' in a given glucose condition m ($\text{cov}(\eta_{Ks,m}, \eta_{Ks',m}) = 0$), while there is correlation between data stemming from the same strain ($\text{cov}(\eta_{Ks,m},$

Table 1
Description of the links between nodes.

Node	Type	Definition ^a
$Y_{s,m,t}$	Stochastic	$N(N_{s,m,t}, \epsilon_{Ns,m,t})$
$N_{s,m,t}$	Logical	Eq. (1)
$K_{s,m}$	Stochastic	$N(Kmean_m, Ksd_m)$
$r_{s,m}$	Stochastic	$N(rmean_m, rsd_m)$
$N_{0s,m}$	Stochastic	$N(N_0mean_m, N_0sd_m)$
$\epsilon_{Ns,m,t}$	Stochastic	$N(0, \sigma \times N_{s,m,t})$

^a $N(a,b)$, normal distribution with expected value a and standard deviation b .

$\eta_{K_{s,m}} = (Ksd_m)^2$ when $s = s'$). This covariation gives the dependence structure of the model.

Note that we explicitly allowed for genotype-by-environment interactions because the standard deviation of the latent variables depended on the environment m .

2.6. Description of the uncertainty related to measurement errors

If we consider a strain s in a culture condition m at a time t , the observation $Y_{s,m,t}$ writes:

$$Y_{s,m,t} = N_{s,m,t} + \varepsilon_{N_{s,m,t}}$$

where $\varepsilon_{N_{s,m,t}}$ corresponds to the residual error around the theoretical law of $N_{s,m,t}$. The three replicates for each medium-by-strain combination were pooled, so that η encompasses both technical and biological variations. This model was chosen because the variations between the three replicates of the same strain in the same medium were very low. In such model, the variability between the replicates is taken into account, but we neglect the dependencies between data points belonging to the same replicate.

As the inspection of the observations revealed that the experimental error was increasing with the population size, we chose to draw the error in a normal distribution, centered at 0, with a standard deviation equal to $\sigma \times N$:

$$\varepsilon_{N_{s,m,t}} \sim N(0, \sigma \times N_{s,m,t}),$$

where σ is the residual standard deviation multiplier of the model.

2.7. Prior distributions

The prior distributions of the means $Kmean_m$, $rmean_m$ and N_0mean_m have been drawn in normal distributions (Table 2). The prior distribution for $Kmean$ has been chosen as wide and as flat as possible because the culture media covered a wide range of glucose conditions, and the carrying capacity $Kmean$ should reflect the nutrient content of the medium. From the literature, the prior distribution for $rmean$ has been chosen with a mean value fixed at 0.01 min^{-1} and a relatively large standard deviation (Wloch et al., 2001). Finally the prior distributions for N_0mean have been fixed at 1, with a standard deviation allowing reaching 5 because from 1 to 5×10^6 cells have been inoculated in a fresh culture at the beginning of the experiments.

The prior distributions of the standard deviations of the model (Ksd , rsd , N_0sd and σ) were chosen to favour large values. The underlying assumptions were (i) there is variation between strains in a given environment (Ksd , rsd , N_0sd) and (ii) the measurement error is large (σ). WinBUGS (the software used to perform the Bayesian

inference, described in next paragraph) works with precision parameters ($Kprec$, $rprec$, N_0prec and σ^{-2}) which are the reciprocal of the square of the standard deviations. Precision parameters were drawn in $G(10^{-3}, 10^{-3})$, where $G(a, b)$ is a Gamma distribution of shape parameter a and scale parameter b (Table 2). Setting $a = b = 10^{-3}$ is a common Bayesian practice for picking non informative precision priors.

2.8. Bayesian inference

Bayesian inferences of parameter values were performed using WinBUGS software (© MRC Biostatistics Unit (Spiegelhalter et al., 2003)). After an adaptation phase (also called “burn-in” phase (Gilks et al., 1996)) of 4000 iterations, the convergence of the Monte Carlo Markov Chain (MCMC) algorithm was checked by visual inspection of the good mixing of three independent chains starting at three different initial values for each parameter. Inferences were made on the following 15000 iterations after the “burn-in” phase.

2.9. Empirical posterior distributions

Altogether, our model comprises 19 parameters: $Kmean$, $rmean$, N_0mean , Ksd , rsd , N_0sd for each of the 3 culture media and σ . The model also comprises 36 latent variables ($K_{s,m}$, $r_{s,m}$ and $N_{0s,m}$). Note that the biological and technical variability due to replicated datapoints is reflected by the posterior distribution of the latent variables. The posterior Monte Carlo samples have been directly used to evaluate the statistics related to the parameters and the latent dynamic population variables (posterior means, standard deviations, medians and 95% credibility intervals). Joint posterior distributions of parameters and latent variables were studied using the function “pairs” under R software. The precision parameters obtained from WinBUGS have been transformed in standard deviations, Ksd , rsd , N_0sd and σ , to have the same unit for the variability and for the mean of the population dynamic key variables.

3. Results

A Bayesian approach was used for estimating population dynamic key variables in yeast, relying on a modelling framework in which the population size N depends not only on the parameters of a logistic function (K the carrying capacity, r the intrinsic growth rate and N_0 the initial population size), but also on fixed and random factors related to environmental variation, to genetic differences between lines and to measurement error (see the modelling scheme in Fig. 1).

Table 2
Prior distributions used for the parameters.

Parameter	Distribution ^a	$P_{0.025}$ ^b	Median	$P_{0.975}$ ^c
$Kmean_m$	$N(70 \times 10^6, 70.7 \times 10^6)$	-69.8×10^6	50×10^6	207.62×10^6
$rmean_m$	$N(0.01, 5.7 \times 10^{-3})$	-0.1×10^{-3}	0.01	0.02
N_0mean_m	$N(1 \times 10^6, 2.24 \times 10^6)$	-3.33×10^6	1	5.284×10^6
Ksd_m	$Kprec_m^d \sim G(0.001, 0.001)$	0	7.92×10^{-305}	3.04×10^{-8}
rsd_m	$rprec_m^d \sim G(0.001, 0.001)$	0	7.92×10^{-305}	3.04×10^{-8}
N_0sd_m	$N_0prec_m^d \sim G(0.001, 0.001)$	0	7.92×10^{-305}	3.04×10^{-8}
σ	$\sigma^{-2d} \sim G(0.001, 0.001)$	0	7.92×10^{-305}	3.04×10^{-8}

^a $N(a, b)$, normal distribution with expected value a and standard deviation b ; $G(a, b)$, Gamma distribution with shape parameter a and scale parameter b . Note that for $Kmean$, $rmean$ and N_0mean , we draw only in the positive part of normal distributions because these parameters can only be positive in our conditions. Note that the actual means of normal distribution truncated for positive values for $Kmean$, $rmean$ and N_0mean are respectively: 89.18×10^6 , 0.01 and 2.26×10^6 .

^b 2.5th percentile.

^c 97.5th percentile.

^d WinBUGS deals with precision parameters, i.e. the reciprocal of the square of the standard deviation.

3.1. Efficiency of HBM

This modelling scheme was efficient for studying the *S. cerevisiae* population dynamics. For each strain/medium, growth modelling allowed us to estimate the key variables K , r and N_0 , and to predict the resulting growth curves as shown in Fig. 2. As expected, the interval $N_{s,m} \pm 2 \sigma \times N$ (σ is the residual standard deviation of the model) included the majority of the experimental data points, indicating that both the model used and the way we described experimental error seemed to be relevant to describe the *S. cerevisiae* population dynamics. Striking genotype-by-environment interactions could be observed for the carrying capacity, since the strain with the highest K value in 15% glucose (Fig. 2A) has the smallest value in 1% glucose (Fig. 2C).

The comparisons between prior and posterior distributions of K mean, r mean, K sd and r sd are shown in Fig. 3. The prior distribution of K mean was very flat and uninformative, whereas the three posterior distributions (one for each medium) were very narrow, with distinct means, even for K mean_{1%} and K mean_{0.25%} (Fig. 3A). For the r mean distributions, differences between prior and posterior distributions were less, probably because the prior distribution for r mean was chosen from relevant literature. Note that choosing a uniform uninformative prior distribution gave the same posterior distributions. Posterior distributions were more tightened than the prior, and were distinct between media even if they overlap in a large part. Posterior distributions of K sd parameters were all Gamma like distributions despite their quite different shapes (Fig. 3C). Finally, posterior distributions of r sd in the three different media merged and were quite different from the prior one, which indicates a similar

genetic variability of the intrinsic growth rate in the three different culture conditions.

Empirical posterior distributions are shown in Table 3, and illustrated in Fig. 4 for the 15% glucose conditions. The distributions of K mean, r mean and N_0 mean were roughly symmetric, except for N_0 mean_{0.25%}, whereas posterior distributions of standard deviation parameters (K sd, r sd and N_0 sd) were slightly skewed to the right.

3.2. Environmental and genetic effects on population dynamics

The environment and the genetic differences between strains had a strong effect on population dynamics. Descriptive statistics of empirical posterior distributions are given in Table 3. K mean mean values increased when glucose increased in the medium (K mean_{0.25%} = 35.57×10^6 , K mean_{1%} = 42.33×10^6 and K mean_{15%} = 96.02×10^6), and r mean mean values decreased when the environment was richer (r mean_{0.25%} = 1.13×10^{-2} , r mean_{1%} = 8.65×10^{-3} and r mean_{15%} = 6.86×10^{-3}). The differences between the N_0 mean mean values reflect experimental variations in the cell density at the beginning of the kinetics: in the 15% glucose medium, more cells were inoculated (N_0 mean_{15%} = 2.71×10^6) than in the two other media (N_0 mean_{1%} = 0.47×10^6 and N_0 mean_{0.25%} = 0.34×10^6).

The standard deviations K sd and r sd directly reflect the genetic variability of population dynamic latent variables K and r among our collection of strains in a given medium. Descriptive statistics for standard deviation parameters are given in Table 3. The variability of the carrying capacity was about 2 times higher in the 15% glucose medium (K sd_{15%} = 21.46×10^6) than in the 1% glucose medium (K sd_{1%} = 12.76×10^6), and about 3 times higher than in the 0.25%

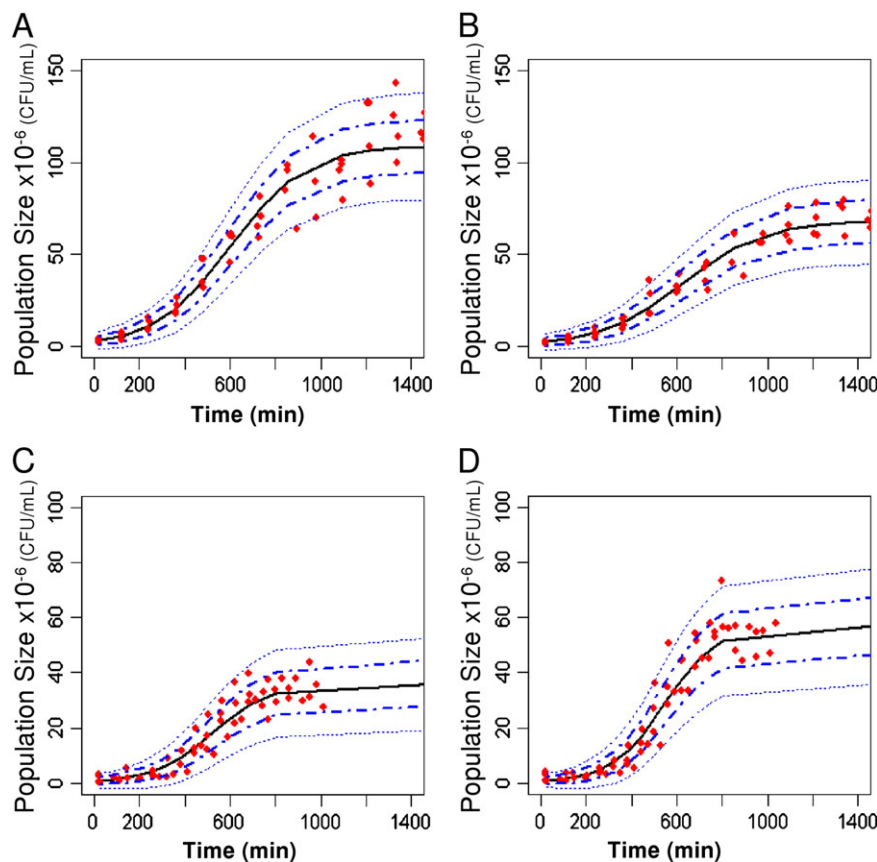


Fig. 2. Modelling the population dynamics of two strains grown in two culture media. Red solid diamonds represent experimental data. Black curves represent the evolution of the modelled population size $N_{s,m}$ over time from K , r and N_0 estimates. Blue dot dashed and blue dotted curves represent respectively $N_{s,m} \pm (\sigma \times N)$ and $N_{s,m} \pm (2\sigma \times N)$ where σ represent the residual standard deviation of the model (or the uncertainty related to the measurement). A and C represents respectively a given strain in the 15% and 1% glucose media, while B and D represents another strain grown respectively in the 15% and 1% glucose media.

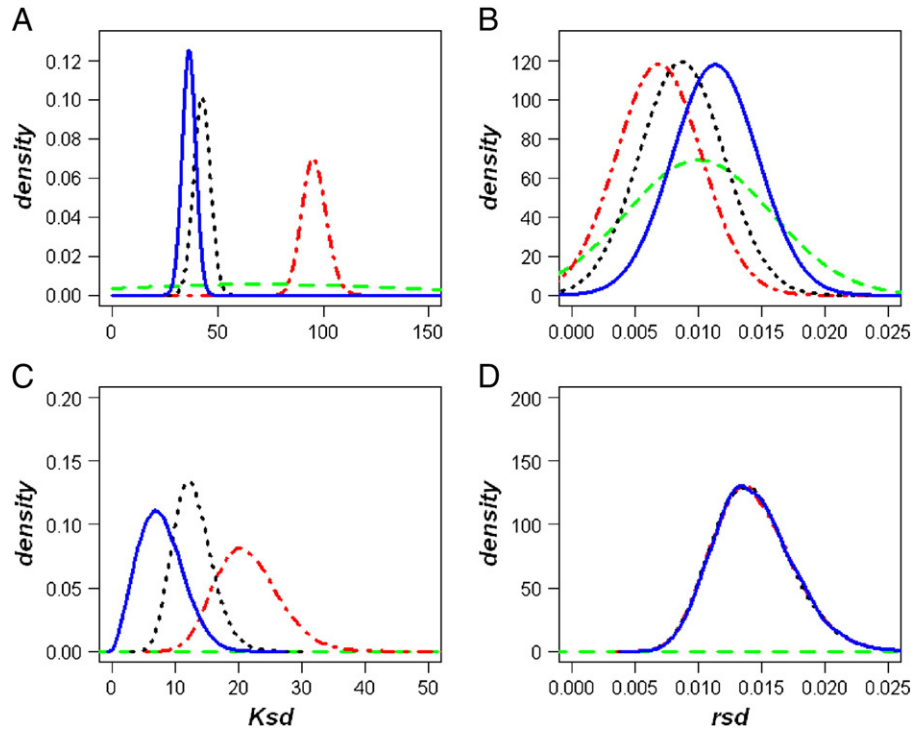


Fig. 3. Comparison of prior and posterior distributions of population dynamic parameters K_{mean} (A), r_{mean} (B), K_{sd} (C) and r_{sd} (D). Red dot dashed, black dotted and blue curves represent respectively posterior distributions in 15%, 1% and 0.25% media. Green dashed curves represent prior distributions.

glucose medium ($K_{sd_{0.25\%}} = 7.85 \times 10^6$) indicating that genetic variability between strains depended on the medium. In other words, we found genotype \times environment interactions for K . On the opposite, the genetic variability of the intrinsic growth rate (r_{sd}) was high but robust with regard to environmental changes ($r_{sd} \sim 1.44 \times 10^2$ in the three media). The variability of the initial population size was more than 10 times higher in the 15% glucose medium than in the two other media (1.08×10^6 for $N_{0sd_{15\%}}$ vs. 0.14×10^6 and 6.45×10^4 respectively for $N_{0sd_{1\%}}$ and $N_{0sd_{0.25\%}}$).

Finally, the residual standard deviation multiplier of the model, σ , has been estimated to 1.398. Thus the real residual standard deviation around the theoretical law of population dynamic $N_{s,m,t}$ is drawn in a

normal distribution centered on 0 with a standard deviation equal to $1.39 \times N_{s,m,t}$.

3.3. Joint distributions and prediction

Over all strains in a given medium, no correlation was observed when studying the joint posterior distributions of the parameters of the model as illustrated by the correlation coefficients and the smooth lines in Fig. 4. Parameter joint distributions are given for illustration in the 15% glucose medium, but the lack of correlation is also valid in the two other media. On the other hand, for a given strain s in a given condition m , key latent variables K and r were negatively correlated as illustrated in Fig. 5. The variance of the joint posterior distribution of the latent variables $K_{s,m}$ and $r_{s,m}$ stems from microenvironmental variations and reflects the variability between two replicates of the same strain in the same environment. In a prediction point of view, these results lead to different ways of drawing latent variables. When correlations are detected, it becomes necessary to draw jointly a K value and an r value in their empirical joint posterior distribution. With our data set, taking into account joint distributions becomes particularly important to model different replicates of a given strain s in a given condition m .

A major interest of Bayesian modelling is its predictive capacity. From our modelling achievement, it becomes possible to predict the typical behaviour of any strain grown in our different glucose conditions, as illustrated in Fig. 6. We first simulated population dynamics in 15% and 1% glucose media by drawing 30 values of K and r in the K_{mean} and r_{mean} empirical joint posterior distributions illustrated in Fig. 4. The 30 growth curves obtained represent the mean typical behaviour in each medium under the assumption that all strains behave like a hypothetical average one (Fig. 6A and B) and reflect the effect of the culture medium. Then, to demonstrate the effect of genetic variability on population dynamics, we drew 30 K and r values respectively in $N(K_{mean}, K_{sd})$ and in $N(r_{mean}, r_{sd})$ with respect to the joint posterior distributions of these parameters. As illustrated in Fig. 6, taking into account the genetic variability between

Table 3
Descriptive statistics of empirical posterior distributions of parameters.

Parameter	Mean	S.D.	$P_{0.025}^a$	Median	$P_{0.975}^b$
$K_{mean_{15\%}}$	96.02×10^6	6.76×10^6	82.35×10^6	96.01×10^6	109.5×10^6
$K_{mean_{1\%}}$	42.33×10^6	4.06×10^6	34.29×10^6	42.28×10^6	50.56×10^6
$K_{mean_{0.25\%}}$	35.57×10^6	3.06×10^6	29.71×10^6	35.48×10^6	41.93×10^6
$r_{mean_{15\%}}$	6.86×10^{-3}	3.36×10^{-3}	4.55×10^{-4}	6.81×10^{-3}	1.37×10^{-2}
$r_{mean_{1\%}}$	8.65×10^{-3}	3.34×10^{-3}	2.21×10^{-3}	8.63×10^{-3}	1.52×10^{-2}
$r_{mean_{0.25\%}}$	1.13×10^{-2}	3.39×10^{-3}	4.59×10^{-3}	1.13×10^{-2}	1.79×10^{-2}
$N_{0mean_{15\%}}$	2.71×10^6	0.35×10^6	2.02×10^6	2.70×10^6	3.41×10^6
$N_{0mean_{1\%}}$	0.47×10^6	0.10×10^6	0.37×10^6	0.45×10^6	0.75×10^6
$N_{0mean_{0.25\%}}$	0.34×10^6	0.09×10^6	0.19×10^6	0.33×10^6	0.56×10^6
$K_{sd_{15\%}}$	21.46×10^6	5.31×10^6	13.79×10^6	20.57×10^6	34.07×10^6
$K_{sd_{1\%}}$	12.76×10^6	3.23×10^6	8.09×10^6	12.21×10^6	20.42×10^6
$K_{sd_{0.25\%}}$	7.85×10^6	3.37×10^6	0.53×10^6	7.79×10^6	14.84×10^6
$r_{sd_{15\%}}$	1.44×10^{-2}	3.33×10^{-3}	9.53×10^{-3}	1.37×10^{-2}	2.24×10^{-2}
$r_{sd_{1\%}}$	1.43×10^{-2}	3.26×10^{-3}	9.5×10^{-3}	1.37×10^{-2}	2.22×10^{-2}
$r_{sd_{0.25\%}}$	1.44×10^{-2}	3.28×10^{-3}	9.53×10^{-3}	1.38×10^{-2}	2.22×10^{-2}
$N_{0sd_{15\%}}$	1.08×10^6	0.32×10^6	0.60×10^6	1.03×10^6	1.85×10^6
$N_{0sd_{1\%}}$	0.14×10^6	0.18×10^6	2.98×10^4	8.75×10^4	0.78×10^6
$N_{0sd_{0.25\%}}$	6.45×10^4	3.94×10^4	2.05×10^4	5.41×10^4	0.17×10^6
σ	1.39	2.63×10^{-2}	1.34	1.39	1.44

^a 2.5th percentile.

^b 97.5th percentile.

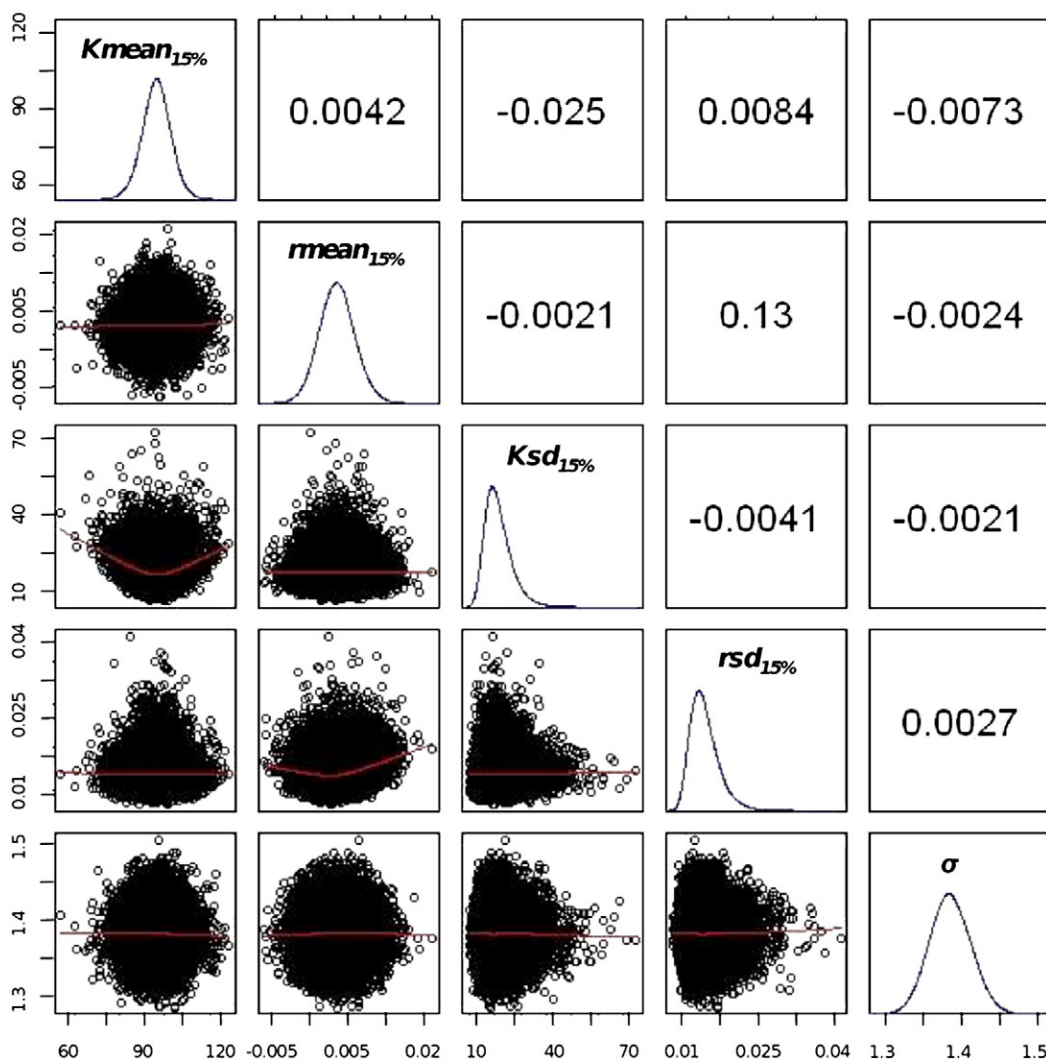


Fig. 4. Empirical joint posterior distributions of parameters $K_{mean_{15\%}}$, $r_{mean_{15\%}}$, $K_{sd_{15\%}}$, $r_{sd_{15\%}}$ and σ , corresponding adjusted distributions and correlation coefficients. Empirical joint posterior distributions are shown on the lower panel of the figure (under the diagonal). The diagonal contains adjusted posterior densities of parameters $K_{mean_{15\%}}$, $r_{mean_{15\%}}$, $K_{sd_{15\%}}$, $r_{sd_{15\%}}$ and σ . Correlation coefficients between parameters are given in the upper panel of the figure (over the diagonal).

strains led to more variable behaviours than predicted by the sole effect of the culture medium on population dynamic key variables.

4. Discussion

We developed a probabilistic model to describe population dynamics of 12 different strains grown in three culture media differing by their glucose content. To our knowledge, none of the previous studies relied on a Bayesian framework to model *S. cerevisiae* population dynamics.

4.1. Differences between “frequentist” and Bayesian approaches

Data presented here had previously been analyzed by fitting a population dynamic model using a frequentist approach. Population key variables were then recovered and an explicative secondary model (ANOVA) had been used to determine if the forcing factors (in our case, medium variation and strain variability) did have a significant effect on the estimated parameter values. With this method, uncertainty caused by the lack of fit of the population dynamic model to the biological data was not taken into account. Yet we had just estimates, and less confidence should be granted to values obtained with fewer points. In addition, non-linear relationships may prevent from a rapid and unbiased convergence of such estimates (the

speed of convergence depends on the number of data points necessary to reach a given estimate, and the bias corresponds to the differences between estimates obtained from different experiments). This strategy can lead to rough approximations, and even to wrong conclusions in some extreme cases. However, for this particular study, the two approaches produced similar estimates.

Within the frequentist context, a better strategy could have been to reconstruct an all-inclusive analysis of variance based on a global non-linear procedure (Mc Culloch et al., 2008; Molenberghs and Verbeke, 2006; Muller and Stewart, 2006). There are now powerful EM algorithms (McLachlan and Krishnan, 1996) for the inference of a broad range of non-linear models (e.g. multi level models, mean-dispersion models, longitudinal models with individual evolutions ruled by differential equations, etc). However, in this kind of models, tests for significant effects of controlling factors need to be adapted and developed specifically for each type of non-linear model. These methods may be powerful in terms of precision of parameter estimators, but are reserved to high skilled statisticians since very few user-friendly software’s are available. For instance, SAS ‘Nlmix procedures’, Monolix and R ‘Nlme’ routines cannot presently deal with a statistical model like the one we have used. Moreover, from a practical point of view, classical frequentist framework encounters obstacles for the treatment of missing data, which either should be deleted or replaced by approximated values. In Bayesian settings,

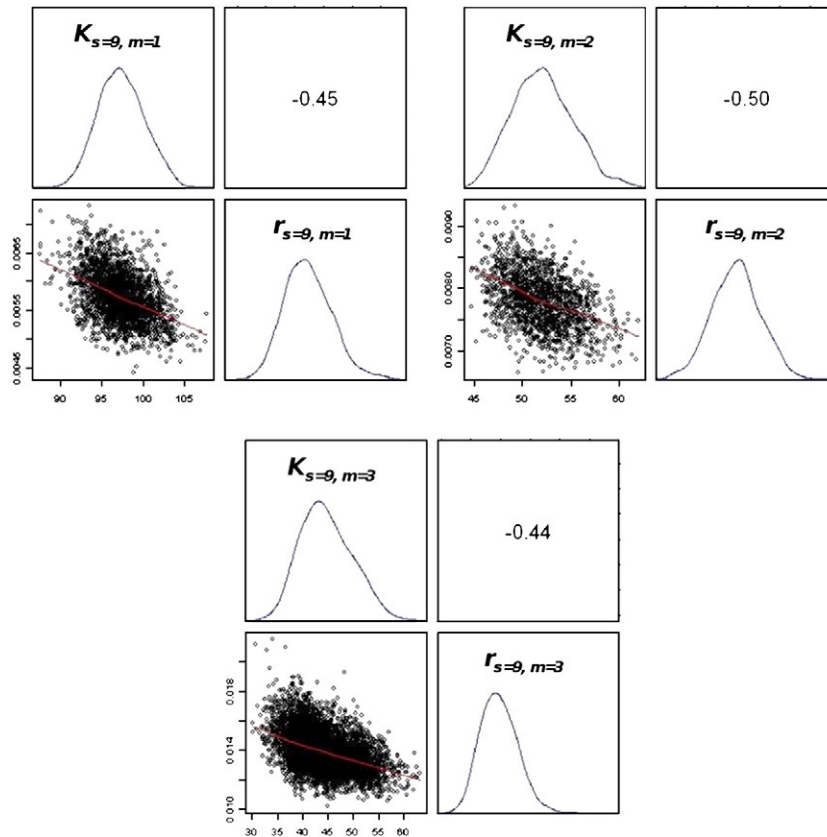


Fig. 5. Empirical joint posterior distributions of latent key variables K and r , corresponding adjusted distributions and correlation coefficients for one strain in the three environmental conditions. Empirical joint posterior distributions are shown on the lower panel of each sub-figure (under the diagonal). The diagonal contains adjusted posterior densities of variables K and r . Correlations coefficient are given in the upper panel of each sub-figure (over the diagonal).

missing data are just considered as latent variables estimated after consideration of the data. No particular treatment of missing values is then needed for the inference.

A fundamental advantage of the Bayesian approach is the possibility to combine various sources of information to formulate hypotheses on parameters of interest, and thus to define their prior distributions. In classical frequentist framework, hypotheses on

parameter distributions are not related to expert knowledge or to previous experiments, while in Bayesian framework, posterior distributions of a first study could be recycled as prior distributions for a second one. The Bayesian approach can be viewed as statistical learning machinery that progressively updates the state of knowledge about a specific phenomenon by processing data from a particular field of research.

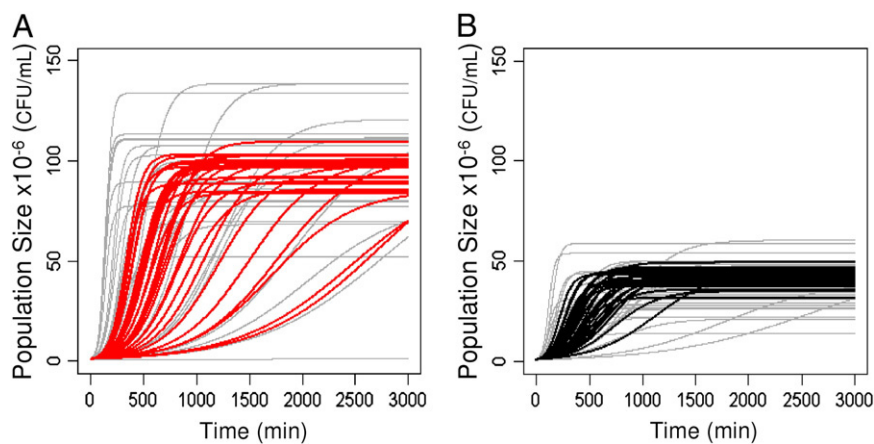


Fig. 6. Modelling the typical population dynamics of *S. cerevisiae* strains in two different environments. Population dynamics was modelled according to probability laws obtained from empirical posterior distribution adjustments. A. 30 growth curves (red lines) were simulated in 15% glucose medium, with the same initial density ($N_0 = N_0 \text{mean} = 10^6$ cell/mL), and $K_{\text{mean}_{15\%}}$ and $r_{\text{mean}_{15\%}}$ drawn in the empirical joint posterior distributions. The 30 grey lines represent the additional information brought after incorporating genetic variability (genetic differences among strains) on population dynamic key variables. For each grey line, K and r are respectively drawn in $N(K_{\text{mean}_{15\%}}, Ksd_{15\%})$ and $N(r_{\text{mean}_{15\%}}, rsd_{15\%})$ taking into account correlations related to empirical joint posterior distributions. B. Same simulation conditions as in A, except that the glucose concentration was 1%.

4.2. Bayesian modelling, a thinking framework for biologists

In statistical terms, we developed a hierarchical non-linear mixed-effects model with heteroscedastic variance. The conception of this model was quite simple and followed the natural way of thinking of biologists. The first step was to imagine the phenomenological process into action. Microbial populations are known to grow in 3 phases: lag, exponential and stationary. The classical mathematical description of this dynamics is a logistic equation summarized by 3 population dynamic variables: K , r and N_0 . This is the deterministic part of the model. Then, variability could be added to the process by drawing these latent variables in probabilistic distributions. The way these variables are drawn reflects directly how biologists understand the process and its behavioural similarities. We could have considered the effect of glucose concentration on population dynamic key variables as a random effect, but we chose to model it as a fixed effect because we had only three culture media not homogeneously distributed (15%, 1% and 0.25%) and not representative of a typical environmental effect. Statistically, this means there is an average value for each parameter in each glucose condition (K_{mean} , r_{mean} and N_{0mean}) and that these three experimental situations are considered as independent. Then we introduced a random effect due to genetic variability, described by the standard deviation of the law from which the population latent variables are drawn (K_{sd} , r_{sd} and N_{0sd}). The last step was to incorporate uncertainty in measurements. Since measurement errors increased with population size, which was due to the higher number of dilutions required to plate the same number of cells, we simply chose to draw the residual variance in $N(0, \sigma \times N_{s,m,t})$, i.e. the standard deviation of the error was proportional to the population size.

Our approach borrowed many concepts from the seminal Delignette-Muller et al. (2006)'s paper regarding Bayesian modelling of growth curves for risk assessment. Specifically, we employed: *i*) the well-known logistic equation used as a primary growth model, *ii*) using their words to separate uncertainty and variability of a model, we tried to account for the main sources of variability (random effects stemming from genetic differences between strains) and uncertainty (essentially measurement errors and only partial knowledge about parameters) and *iii*) the same powerful MCMC techniques were used to perform posterior distribution of the unknowns. However, major differences in the modelling assumptions were made in our article, since main sources of variability and uncertainty were different. In Delignette-Muller et al. 2006, the researchers had to consider growth parameters as a function of changing conditions over a large range, since in addition to their own 61 curves, 35 others were taken from 10 publications. As all our data have been obtained specifically for this study under controlled conditions, we didn't need such a secondary growth model. Similarly, due to the specific features of our strains data, we didn't need to introduce a lag time in the model. In that sense, our model is simpler. Nevertheless, we had to relax the assumption of a homogeneous measurement error variance described in Delignette-Muller et al. (2006) over each growth curve. Thus, we proposed here a more general way to deal with measurement uncertainties. Because we had the same strains replicated in different conditions, our model allowed us to take explicitly genotype \times environment interactions into account, which was not the case in the Delignette-Muller et al. (2006). Concerning the applications, Delignette-Muller et al. (2006) used parameter estimations that were performed separately on one bacterial species and on the total microbial flora to predict the results of the competition between the bacterial species of interest and the total flora. Because we were interested in predicting the environmental conditions that favour yeast growth, we used our model to predict the range of variation for growth in different environments.

Thus the construction of the model was quite intuitive. Moreover, we would like to emphasize that Bayesian approaches are very

powerful and accessible to model any dynamical or spatial processes, such as the impact of any environmental abiotic (temperature, oxidative stress, poisoning) or biotic factors, the competition between two or more micro-organism species or strains, etc. A free dedicated software, WinBUGS (© MRC Biostatistics Unit, (Spiegelhalter et al., 2003)), is available and a library has been developed under R (library "BRugs" http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/2.9/BRugs_0.5-3.zip, the R Development Core Team) giving to the scientist the opportunity to develop his own models.

4.3. Additional improvements

Another possible approach could have been functional data analysis (Ramsay and Silverman, 2005) which aims at explaining the variations of a "response" function of time by the information of other "explanatory" variables, such as, in our case, the culture media or the strains. The analysis consists in converting the data into a system of basic functions which are combined linearly to define the actual functions. Then, a collection of statistical techniques is applied to the parameters of the basic functions. Statistical packages are available for R and Matlab. This approach is really efficient for answering questions like "How do the curves vary from one condition to another?" or "Are the differences between the curves significant?". However, because parameter estimation is performed on basic mathematical functions, and not on the actual response curves, they may lack of biological meaning. In the case of population dynamic analysis, the key variables of the latent process, K , r and N_0 , have a direct biological meaning.

Various types of equations could have been used to model *S. cerevisiae* population dynamics. We chose to use a logistic equation, because it is parsimonious (three variables having a biological meaning are sufficient) and possesses an analytical solution which properly describes *S. cerevisiae* population dynamics. A more sophisticated population dynamic model based on ordinary differential equations, as described in (Billoir et al., 2008), could have been used. In this type of modelling, the specific add-on module PKBUGS developed by Lunn can be used, as in pharmacokinetics, whenever the ordinary differential equation models with unknown coefficients cannot be analytically solved (Lunn et al., 1999).

Competing models could have been designed and compared via their posterior probabilities (e.g. with Bayes factors; (Kass and Raftery, 1995)) or according to their predictive ability on a test sample (predictive posterior checks; (Gelman et al., 2003)). The hypothesis of conditional independence between media could be relaxed. For instance a continual covariation of the parameters with the environments could be added. A random strain origin effect independent from glucose concentration could have been declared, and variance estimates could have been compared to the estimates of the present model. In addition, a more complex model would assume prior correlations between population dynamic features with a 3-dimensional random effect instead of three independent random effects for each of these. This would help taking into account biological adjustments in the life-history strategy of strains such as an increased value of r to compensate for a below average K . Other refinements may affect the measurement error structure: at very low additional computation costs, a two-parameter proportional error $\varepsilon_{N_{s,m,t}} \propto N(0, \sigma \times N_{s,m,t}^\gamma)$ could be studied in the model (γ is the second parameter).

It is also relatively straightforward to develop the same type of model for more than two factors providing that enough data are collected to convey sufficient information to update the additional parameter priors to be involved into the analysis. However when many factors interplay, we can no longer trust the common microbiological good sense to focus on a single model and many competing models can be designed. When the number of possible models gets large (and this number grows very rapidly with the number of factors), the Bayesian model selection problem may

become tricky and refined MCMC strategies (clearly out of the scope of this paper) should be used to select the appropriate model (reversible MCMC jump as in Green, 1994 or stochastic search as in chapter 4 of Marin and Robert, 2007).

4.4. A measure of G^*E interactions

Genotype-by-environment interactions reflect the way a genotype behaves across different environments. In classical quantitative genetics, an ANOVA is performed on the trait considered, and the G^*E effect is inferred when (i) the ranking of the genotypes changes according to the medium or (ii) the inter-individual variance varies in the different environmental conditions (Lynch and Walsh, 1997).

Bayesian inference provides us with both the empirical posterior distributions of growth key variables ($K_{s,m}$, $r_{s,m}$, $N_{0s,m}$) of each strain in each glucose condition, and the standard deviation (Ksd_m , rsd_m , N_0sd_m) of those variables in each glucose condition, which reflects strain variability in each condition. Then comparisons of the ranking of the average growth key variables of the strains could be performed (for example rank correlation of parameter values in two different environments), as well as comparisons of posterior distributions of standard deviations of parameters among culture conditions (using for example Kolmogorov–Smirnov tests).

In our case, the means of the empirical posterior distributions of Ksd are very different in the three culture media while the means of rsd stay constant. The between-strain genetic differences for carrying capacity are increased in rich media compared to poor ones, while genetic variation of the intrinsic growth rate stays robust towards environmental changes. This is a good example of standing genetic variation expressed only under certain environmental conditions (Ksd) and canalization (rsd) (Waddington, 1942). HBM would be powerful for providing better predictions of the genetic diversity of population dynamic key variables in relation to environmental variation.

Last but not least, another fundamental aspect of the Bayesian approach is the possibility to study the joint posterior distributions of latent variables and parameters of interest. Studying the joint posterior distributions of $Kmean$ and $rmean$ in each glucose condition should inform us about the possible existence of a genetic trade-off between carrying capacity and intrinsic growth rate in *S. cerevisiae* populations. We did not detect it in our data. However we showed that modelling the population growth of a given strain s in a given condition m requires to take into account the conditional dependence between key latent variables K and r . In other words, the different possible ways a given strain can grow in a given condition (because of the microenvironmental variations, within population variations and stochasticity between technical replicates) are constrained by a trade-off between K and r (Novak et al., 2006).

4.5. Further prospects

The next step would be to extend this model and to classify strains according to their population dynamic behaviour. By defining a fixed effect of the industrial origin of strains (as we did for the glucose effect), and modelling population growth in typical conditions that maximize the differences between strains coming from different industrial origins, we could determine the adjusted distributions of population dynamic key variables. Then, when testing a novel natural strain in this environment, it would be possible to assign a measure of “goodness” of this strain in a specific industrial condition by classifying it according to its population dynamics. Another possible application could be the choice of typical growing conditions in order to maximize the differences between industrial origins. A medium that seems appropriate to maximize differences between strains coming from different industrial origins would be a medium in which all strains are able to grow (not too stressful) but in which strains

could exhibit quantitative variation for population dynamic parameters. For example, in our experiment, the 15% glucose medium typically maximizes the variance for K between strains (Fig. 6). However, from a statistical perspective, it is rather a challenge to plan in practice an optimal design, which would maximize beforehand the expected differences between strains coming from different industrial origins. In the Bayesian setting (Müller, 1999), a utility function has to be elicited from the end-users and burdensome high dimensional maximization and integration have to be performed (Amzal et al., 2006).

Acknowledgements

The authors would like to thank Yves Rousselle, Adrienne Ressayre and Thibault Nidelet for critical comments and programming assistance. We are also grateful to reviewers for their critical comments. This work was supported by a PhD fellowship of the Ministère de l'Enseignement Supérieur et de la Recherche to Aymé Spor, and indirectly benefited from a grant by the French Agence Nationale de la Recherche (ANR Project ADAPTALEVURE no. NT05-4_45721).

References

- Amzal, B., Bois, F., Parent, E., Robert, C.P., 2006. Bayesian optimal design via interacting particle systems. *Journal of the American Statistical Association* 101 (474), 773–785.
- Barker, G.C., Malakar, P.K., Peck, M.W., 2005. Germination and growth from spores: variability and uncertainty in the assessment of food borne hazards. *International Journal of Food Microbiology* 100, 67–76.
- Beltran, G., Torija, M.J., Novo, M., Ferrer, N., Poblet, M., Guillamon, J.M., Rozes, N., Mas, A., 2002. Analysis of yeast populations during alcoholic fermentation: a six year follow-up study. *Systematic and Applied Microbiology* 25, 287–293.
- Bernier, J., Parent, E., Boreux, J.J., 2000. *Statistique pour l'environnement* Lavoisier, Paris.
- Billoir, E., Delignette-Muller, M.L., Péry, A.R., Charles, S., 2008. A Bayesian approach to analyzing ecotoxicological data. *Environmental Science & Technology* 42 (23), 8978–8984.
- Boekhout, T., Robert, V., 2003. *Yeasts in Food*. Woodhead Publishing Limited, Cambridge, England.
- Calder, C., Lavine, M., Müller, P., Clark, J.S., 2003. Incorporating multiple sources of stochasticity into dynamic population models. *Ecology* 84, 1395–1402.
- Clark, J.S., 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology* 84, 1370–1381.
- Clark, J.S., Gelfand, A.E., 2006. *Hierarchical Modeling for the Environmental Sciences: Statistical Methods and Applications*. Oxford University Press Inc., New York.
- Delignette-Muller, M.L., Cornu, M., Pouillot, R., Denis, J.B., 2006. Use of Bayesian modeling in risk assessment: application to growth of *Listeria monocytogenes* and food flora in cold-smoked salmon. *International Journal of Food Microbiology* 106, 195–208.
- Domizio, P., Lencioni, L., Ciani, M., Di Blasi, S., Pontremolesi, C., Sabatelli, M.P., 2007. Spontaneous and inoculated yeast populations dynamics and their effect on organoleptic characters of Vinsanto wine under different process conditions. *International Journal of Food Microbiology* 115, 281–289.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. *Bayesian Data Analysis*. Chapman & Hall/CRC.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Green, P.J., 1994. Reversible jump MCMC computation and Bayesian model determination. Tech. Report. University of Bristol.
- Hooten, M.B., Wikle, C.K., Dorazio, R.M., Royle, J.A., 2007. Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics* 63, 558–567.
- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Kéry, M., Royle, J.A., 2008. Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology* 45, 589–598.
- Lunn, D.J., Wakefield, J., Thomas, A., Best, N., Spiegelhalter, D., 1999. *PKBugs User Guide*. Dept. Epidemiology and Public Health, Imperial College School of Medicine, London.
- Lynch, M., Walsh, B., 1997. *Genetics and Analysis of Quantitative*. Traits Sinauer associates, Inc., Sunderland, Massachusetts.
- MacArthur, R., Wilson, E.O., 1967. *The Theory of Island Biogeography*. Princeton University press.
- Marin, J.M., Robert, C.P., 2007. *Bayesian Core: a Practical Approach to Computational Bayesian Statistics*. Springer, New York.
- McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. *Generalized, Linear and Mixed Models*. Wiley-Interscience.
- McLachlan, G.J., Krishnan, T., 1996. *The EM Algorithm and Extensions*. Wiley-Interscience.
- Membre, J.M., Leporq, B., Vialette, M., Mettler, E., Perrier, L., Thuault, D., Zwietering, M., 2005. Temperature effect on bacterial growth rate: quantitative microbiology approach including cardinal values and variability estimates to perform growth simulations on/in food. *International Journal of Food Microbiology* 100, 179–186.

- Molenberghs, G., Verbeke, G., 2006. Models for Discrete Longitudinal Data. Springer, Berlin/Heidelberg.
- Müller, P., 1999. Simulation-based optimal design. Bayesian Statistics 6, 459–474.
- Muller, K.E., Stewart, P.W., 2006. Linear Model Theory: Univariate, Multivariate, and Mixed Models. Wiley-Interscience.
- Nauta, M.J., 2000. Separation of uncertainty and variability in quantitative microbial risk assessment models. International Journal of Food Microbiology 57, 9–18.
- Nauta, M.J., 2002. Modeling bacterial growth in quantitative microbiological risk assessment: is it possible? International Journal of Food Microbiology 73, 297–304.
- Novak, M., Pfeiffer, T., Lenski, R.E., Sauer, U., Bonhoeffer, S., 2006. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. The American Naturalist 168, 242–251.
- Pouillot, R., Albert, I., Cornu, M., Denis, J.B., 2003. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to *Listeria monocytogenes*. International Journal of Food Microbiology 81, 87–104.
- Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis. Springer, New York.
- Shorten, P.R., Membre, J.M., Pleasants, A.B., Kubaczka, M., Soboleva, T.K., 2004. Partitioning of the variance in the growth parameters of *Erwinia carotovora* on vegetable products. International Journal of Food Microbiology 93, 195–208.
- Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS Version 1.4 User Manual MRC Biostatistics Unit, Cambridge, UK.
- Spor, A., Wang, S., Dillmann, C., de Vienne, D., Sicard, D., 2008. “Ant” and “grasshopper” life-history strategies in *Saccharomyces cerevisiae*. PLoS ONE 3, e1579.
- Waddington, C.H., 1942. Canalization of development and the inheritance of acquired characters. Nature 150, 563–565.
- Wikle, C.K., 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84, 1382–1394.
- Wloch, D.M., Szafraniec, K., Borts, R.H., Korona, R., 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. Genetics 159 (2), 441–452.