Characterizing the dynamical accumulation of nuclear DNA in the sperm cells of *Lycium barbarum* L.

Hua Deng,[1] Pierre Nouvellet,[2] David Waxman[1]

[1]Centre for Computational Systems Biology, Fudan University, Shanghai, China; [2]Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, UK

Abstract

When sperm cells of the plant *Lycium barbarum* L. (*L. barbarum*) form in a style they begin to synthesize nuclear DNA (nDNA), which monotonically increases over time. To characterize the dynamics of nDNA accumulation, we present two new dynamical/statistical models. We applied these models to the accumulation of the nDNA content of sperm cells in *L. barbarum* between 16 to 32 hours after pollination in a style. A statistical analysis of experimental data, involving Markov chain Monte Carlo methods, allowed estimation of parameters of the models. We conclude that the model with no variation in the rate of nDNA accumulation adequately summarizes the data. This is the first work where the dynamics of nDNA accumulation has been quantitatively modeled and analyzed.

Introduction

Plant embryology includes studies of the nuclear DNA (nDNA) of male and female gametes.[1] Most of the previous research on this subject, for example, in the plants *Arabidopsis* and *Nicotiana tabacum* investigated the nDNA *content* of the gametes.[2,3] By contrast, in the present work, we present mathematical modeling and analysis of the *dynamics* of nDNA accumulation. In particular, we analyzed data collected on the plant *Lycium barbarum* and studied the dynamics of nDNA accumulation in sperm cells. We introduced and employed two statistical models for this purpose. We used these models to determine the intrinsic rate of accumulation of nDNA content in *L. barbarum*. This plant is in the same family as *Nicotiana tabacum*, namely *Solanaceae*. The plant *L. barbarum* is economically important because it is a fruit producer.

It also has applications in Traditional Chinese Medicine.[4-6] During the initial stages of fertilization, a pollen grain is transferred to a plant's stigma by the process of pollination. The plant *L. barbarum* falls under the class of plants that produce bicellular pollen grains (*i.e.*, containing one generative cell and one vegetative cell). Under appropriate conditions, a pollen grain germinates a pollen tube on a stigma, in which case the pollen tube elongates, and grows into the transmitting tissue in the style. For pollen of a bicellular type, a generative cell (which is linked to a vegetative cell) divides to form two sperm cells, some time after pollination, during pollen tube elongation. In previous experiments on *L. barbarum*,[7] the styles were examined every 4 hours from the time of pollination and sperm cells were observed (under a microscope) only at 16 or more hours after pollination. Hence it is plausible that individual sperm cells formed between 12 and 16 hours after pollination. During the process of sperm cell development, good quality pollen tubes (*i.e.*, those which are potentially successful at achieving fertilization), contain sperm cells in the vicinity of the tip, and these are transported along, with the growth of the pollen tube.[7-9] Sperm cells, once formed initiate nDNA content synthesis (their nuclei begin to accumulate DNA). Experimental techniques allow the measurement of the nDNA content of sperm cells at different times after their formation, at different stages of development. The analysis carried in the present work was used to *quantitatively* test a new dynamical hypothesis concerning DNA accumulation of the sperm cells (see below). A qualitative analysis of the data has been reported elsewhere,[10] and relevant details of the experiments are given in this previous work; the primary focus of the present work is modeling and data analysis.

Hypothesis

The work presented here is based on the hypothesis that accumulation of nDNA content in plant sperm cells occurs *linearly* over time. This hypothesis is original to this paper and has not, to the best of our knowledge, been made elsewhere. This linear behavior is assumed to apply for the experiments on *L. barbarum* from the time of first observation of sperm cells, namely 16 hours after pollination, to the longest time that sperm cells were observed after pollination (32 hours). In principle, if measurements had been made over longer times than the longest times adopted (32 hours), then it is possible that some sort of saturation effect in the nDNA content of sperm cells could set in, prior to zygote formation, with the nDNA content/time curve leveling off at long times. We saw no evidence of this in the data collected up to 32 hours after pollination (Table 1), hence models with a linear

increase appear adequate, but a natural extension of this work could involve nDNA increase according to a saturating function of time. We thus introduced two models with linear accumulation of nDNA content over time. The virtue of the models is that they have only a few parameters, which can be estimated from the experimental data. Any non-linear model will generally involve more parameters than a linear model, and make greater demands on the data. However, extensions of the models are possible.

We assume the nDNA content of sperm cells have variability from *two* different sources. The first is an aspect of the fluorescence technique employed, which leads to errors in the measured nDNA content at any time. In principle, there is a second source of variability: the nDNA content of the sperm cells arises from intrinsic variation in the rate of accumulation of nDNA of different sperm cells. That is, different sperm cells may have different rates of DNA accumulation.

Materials and Methods

nDNA content determination

We determined the nDNA content of sperm cells using microscopy and fluorescence techniques. With a Leica DMR fluorescence microscope, we observed and photomicrographed sperm cells. The software associated with the microscope (Simple PCI) was used in the

analysis of the data. All parameters of the software were set in advance and the only parameter requiring adjustment was the exposure time. To analyze the data, we selected a cell in a photomicrograph, and by manually circling the cell, the software calculated the fluorescence value. The fluorescence value was not influenced by the exposure time, when below the saturation level, because the measured fluorescence was relative to the background fluorescence, and the difference, to good accuracy, is independent of the exposure time.

Statistical analysis

We used MCMC with the Metropolis-Hastings algorithm to sample the likelihood.[11] See the Models Section for a derivation of the likelihood functions.

Models

We introduce two dynamical models that incorporate the hypothesis that accumulation of nDNA content in the plant sperm cells occur *linearly* with time. We assume that at t hours after pollination, the nDNA content of plant sperm cells given by

$$D = A + Rt + \varepsilon \qquad \text{Eq. (1)}$$

where:

D is the nDNA content (measured in units of C, where, by definition, a diploid cell has an nDNA content of 2C) of a sperm cell at time t; A is a constant; R is the rate of accumulation of

nDNA content of a sperm cell; ε is the error in the measured nDNA content of a sperm cell that is introduced by the fluorescence technique; we assume ε is an intrinsic property of the fluorescence technique.

We model the *fluorescence error*, ε, as a random variable that varies from sperm cell to sperm cell. We make the simplest assumptions that the fluorescence errors are independent for different sperm cells and unbiased - hence ε has an expected value of zero.

In the experiments, the nDNA content of sperm cells are measured at given times, with the measurement destroying the sperm cells, so they cannot be re-measured, hence different measurements are on different sperm cells.

Two models, within the framework of Eq. (1) suggest themselves.

Model 0

All sperm cells have an *identical* rate of DNA accumulation. Thus the parameter R in Eq. (1) is a constant that can be estimated from the data.

Model 1

A slightly more sophisticated model assumes that different sperm cells have different rates of DNA accumulation, so a randomly picked sperm cell will have a value of its rate of nDNA content accumulation, R, that is drawn from a continuous distribution. We model R as a *normal* random variable that for different

sperm cells are statistically independent and identically distributed. We also assume statistically independent fluorescence errors (ε). Because we model R as a normal random variable, we have a characterization of the rate of nDNA accumulation in terms of the median value of R and its variance, which we write as μ_R and σ^2_R, respectively.

Statistical analysis of the models

We took the nDNA content of sperm cell i (with $i =1, 2, 3, \ldots$) at time t, which we denote by $D_{i,t}$, as $D_{i,t} = A + R_i t + \varepsilon$. In this formula, A is common to all sperm cells, R_i is the rate of nDNA content accumulation of the i'th sperm cell at time t, and ε is the corresponding random error arising from the fluorescence measurement technique.

We next summarize the statistical analysis; fuller details are given in the Methods.

Model 0

In the first model we assumed no random effect in nDNA accumulation rate, and the model is summarized by $D_{i,t} = A + \mu_R t + \varepsilon$ where that fluorescence errors (ε) follow a normal distribution with mean zero and variance σ^2_ε, while the R_i's take only a single value, which we write as μ_R. The distribution of nDNA content is thus a normal distribution with mean $A + \mu_R t$ and variance σ^2_ε. This corresponds to a classical linear regression and we estimate the parameters using a MCMC procedure with the Metropolis-Hastings algo-

Table 1. Nuclear DNA content of sperm cells at different times.

Description	No. measurements	RFU±SD	DNA level/C
Fluorescence value of bare slide (background value)	20	1.4±0.7	
Somatic cell in style (control)	240	131±19	2
Sperm cells after 16h of pollination in style	10	91±10	1.4
Sperm cells after 20h of pollination in style	28	112±11	1.72
Sperm cells after 24h of pollination in style	38	112±9	1.72
Sperm cells after 28h of pollination in style	55	120±12	1.83
Sperm cells after 32h of pollination in style	11	122±15	1.87

RFU, relative fluorescence units; SD, standard deviation; C, nuclear DNA content a diploid cells has, by definition, an nDNA content of 2C. This table describes the nDNA content of sperm cells at different times after pollination in the style. The table summarizes how many sperm cell samples were tested. As an approximate guide, 100 relative fluorescence units (RFU) represent 1C of DNA. The calculation of DNA level of a sperm cell is given by: DNA level = [(RFU of a sperm cell − RFU of background (paraffin) − RFU of a slide) / (RFU of a somatic cell − RFU of background (paraffin) − RFU of a slide)] ×2C, where the background fluorescence level is calculated for an area of paraffin identical to the area of the image of the sperm cell's nucleus.

Table 2. Summary of statistical results for the two models.

	Model parameters, means (95%CI)				DIC (effective No. parameters)
	A	μ_R	σ^2_R	σ^2_ε	
Model 0 (R constant)	1.16 (0.97-1.32)	0.024 (0.017-0.031)		0.032 (0.026-0.041)	−84.7 (2.3)
Model 1 (R random effect)	1.14 (0.96-1.30)	0.025 (0.018-0.032)	1.40×10^{-5} (0.08×10^{-5}-4.12×10^{-5})	0.024 (0.009-0.036)	−83.8 (2.9)

This table gives all of the parameters estimated for Model 0 and Model 1 from the data. The median values of the parameter A are 1.16 and 1.14 respectively. These values are both close to one another (<2% different), suggesting that this parameter is not particularly sensitive to variability of in the nDNA rates. We note that the estimated constant rate of nDNA content accumulation of Model 0 is 0.024/hour and the median value of Model 1, of 0.025/hour, are close to one another, suggesting, again, insensitivity to the assumption of variability of different sperm cells. The Deviance Information Criterion supports this; Model 0 is marginally superior, according to this criterion. Thus it appears completely sufficient to adopt Model 0 for the analysis and interpretation.

rithm,[11] since this method can also accommodate Model 1 (which allows variation in the DNA accumulation rate). We shall compare the fit of both models using the Deviance Information Criterion (DIC)[12] - where a smaller value of DIC suggests a better fit.

Model 1

In a second, slightly more sophisticated model, we assumed that the rate of nDNA accumulation exhibits variation, from sperm cell to sperm cell, and hence the nDNA accumulation rate of a sperm cell is a random variable. We make the simplest assumption, that the R_i's follow a normal distribution with mean μ_R and variance σ^2_R. Each measurement of the nDNA content of a sperm cell exhibits randomness because of the R_i variation, and also because the fluorescence errors (the ε) follow a normal distribution with mean zero and variance σ^2_ε. The distribution of the measured nDNA content at time t is thus a normal distribution with mean $A + \mu_R t$ and variance $\sigma^2_\varepsilon + \sigma^2_R t^2$. The analysis is more complicated since different numbers of sperm cells were collected (and hence measured) at different times after pollination.

We give detailed results for Model 1, which includes variation in the rate of nDNA accumulation. Under the assumptions of this model, the parameters are: a constant, A, which is common to all sperm cells (see Eq. (1); the median rate of nDNA accumulation of sperm cells, μ_R; the variance of the rate of nDNA accumulation of sperm cells, σ^2_R ; the variance of the error in a sperm cell's nDNA from the fluorescence technique, σ^2_ε. We reported the median of the posterior distribution, as this quantity is known to be more stable than the mean to changes in sample size, and is more representative of a central tendency (especially for distributions which are asymmetric). Given the data D, the log likelihood is given by

$$\log\left[\mathrm{Prob}(D|A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R)\right] = -\frac{1}{2}\sum_{i,t}\log(2\pi(\sigma^2_\varepsilon+\sigma^2_R t^2))$$
$$-\frac{1}{2}\sum_{i,t}\frac{(D_{i,t}-A-\mu_R t)^2}{2\pi(\sigma^2_\varepsilon+\sigma^2_R t^2)}$$

Eq. (2)

According Bayes' theorem

$$\mathrm{Prob}(A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R|D) = \mathrm{Prob}(D|A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R)$$
$$\mathrm{Prob}(A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R)/\mathrm{Prob}(D)$$
$$\propto \mathrm{Prob}(D|A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R)\,\mathrm{Prob}(A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R).$$

Eq. (3)

Using a flat prior, corresponding to no previous information on the parameters, we have

$$\mathrm{Prob}(A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R|D)$$
$$\propto \mathrm{Prob}(D|A,\mu_R,\sigma^2_\varepsilon,\sigma^2_R)$$

Eq. (4)

The results of the statistical analysis of the models, based on Eqs. (2), (3) and (4), are summarized in Table 2.

Results

The nDNA content of sperm cells of *L. barbarum,* estimated from experimental data, was given in terms of relative fluorescence units. These were converted to C value units (where the nDNA content of diploid cells have, by definition, an nDNA content of $2C$). Vegetative nuclei should, optimally, be treated as the control in this experiment because they are connected to the sperm cells and have an nDNA content of $2C$. However, during the experiments, vegetative cells were not seen, and somatic cells (which are diploid) were adopted as controls. We determined that these cells were all in the same stage of development, by testing their fluorescence level (we chose a window of fluorescence levels of width 20 RFU around the mean value, as an indication of the stage of the development). It follows that in the experiments, the relative fluorescence of style somatic cells was used as the standard by which the sperm cells were compared and hence calibrated. Subtracting the background fluorescence of both the cytoplasm and the embedding medium allowed the nDNA content of the sperm cells to be estimated.

In Figure 1 we illustrate the data, in the form of frequency histograms, at different times after pollination. This illustrates the variability of the measured value of the nDNA content of a sperm cell.

The results of the statistical analysis of this data are summarized in Table 2.

For both Model 0 and Model 1, we obtained the joint posterior distribution of the parameters (Eq. (3), and its analogue for Model 0) by using MCMC methods.[11] The MCMC procedure was iterated 10^5 times, and we tuned the proposal variances to achieve good mixing of the different posterior distributions (*i.e.*, around 20% acceptance probability). Reported param-
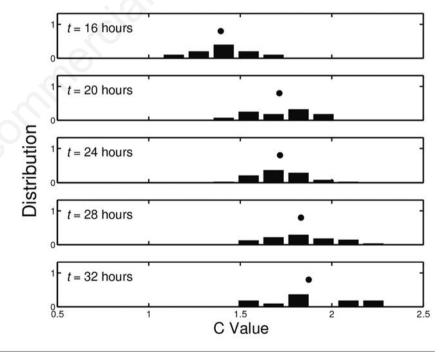

Figure 1. Distribution of nuclear DNA content of sperm cells at different times after pollination. In this figure, we show the empirical distribution of the nDNA content of sperm cells at different times after pollination. For example, the final histogram, labeled *t*=32 hours, is the empirical distribution of the sperm cell's nDNA content, when measured in C units, at 32 hours after pollination. The black dots in the figure represent the mean value of the nDNA content of a sperm cell, at a given time, as calculated from the data. The figure shows that the distribution of the nDNA content of sperm cells, at different times after pollination, has an increasing trend. Note that the measured nDNA content of some sperm cells exceeds the 2C which is the maximum level. This is assumed to be a consequence of the fluorescence technique, which effectively adds a random component to the actual nDNA content, and thereby extends the measured range.

eters and distributions were computed once convergence to stable distributions was achieved. Given the computational simplicity of the two models, we used a fixed burn-in period of 10^4 iterations for both models. Visual inspection of the likelihood traces indicate that this burn-in period is very conservative; for both models, convergence was achieved well before this number of iterations. Figure 2 shows the posterior distributions of the statistical analysis for Model 1.

To show differences of the two models, in Figure 3 we have plotted the *best straight lines* through the data, according to the parameters in Table 2. Additionally, we have averaged the data values collected at each time after pollination, and also plotted these in Figure 3. At different times after pollination, different numbers of sperm cells were collected, and the resulting data make different contributions to the parameters in Table 1. These differences were incorporated into the statistical analyses carried out.

Discussion and Conclusions

Measurements of the nDNA content of sperm cells in plants, have, previously, had the objective of determining the nDNA content.. In the present work we have analyzed the experimentally measured nDNA content of sperm cells of a plant at different times after pollination and focused on the dynamical aspect of nDNA synthesis. We have used the experimental data to investigate the rate at which nDNA content synthesis (or nDNA accumulation) occurs in sperm cells. We carried out a statistical analysis where two different dynamical models were fitted to the data. In both models, we made the dynamical assumption that for a given sperm cell, the accumulation of nDNA content occurs *linearly* with time. However, in our first model (Model 0) we assumed the rate of accumulation of nDNA content of different sperm cells was identical. We thus assumed that there was no variation in the rate of nDNA accumulation of different sperm cells. In Model 1 we allowed the possibility that there was variability in the rate of nDNA accumulation of different sperm cells.

Eq. (1) describes the nDNA content of the plant sperm cells and has been fitted to data covering 16 to 32 hours after pollination. It is not meaningful to apply this equation for times where the sperm cells do not have an independent existence. We do not have a direct measure of this time, but the microscopy observations suggest that no sperm cells exist up to 12 hours after pollination, but only later than this time. However, if we nonetheless apply Eq. (1) from the time of pollination (time 0) onwards, not just from 16 hours after pollination, then the parameter A would have the plausible interpretation as the nDNA content of *half* of a generative cell (which is the direct precursor of a sperm cell). This would suggest a value of A that is close to unity, since all generative cells have an nDNA content of $2C$.[3] Remarkably, the statistical analysis leads, in both models, to values of A within 16% of unity. This makes it plausible that the structure which develops into a sperm cell, from the time of pollination, accumulates nDNA near *linearly* over time, but we have no direct evidence of this. Let us now consider the factors influenc-
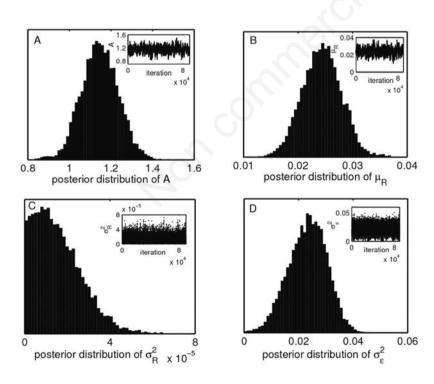

Figure 2. Posterior distributions of the statistical analysis. In this figure we give posterior distributions for Model 1 of: the initial nDNA content of a sperm cell, A, (Panel A); the rate of nDNA accumulation, μ_R, (Panel B); the variance in the rate of nDNA accumulation, σ^2_R, (Panel C); the variance in measurement error, σ^2_ε (Panel D). In each of the panels, we also plot, in inset, the parameter trajectories after the burn-in period, confirming that convergence and good mixing was achieved.
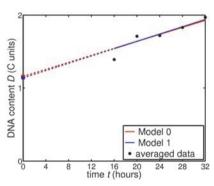

Figure 3. Plot of the best straight lines from the models. In this figure, we have plotted the two *best straight lines* that can be derived from the two models, namely $D=A+\mu_R t$. These lines cover the range 16 hours to 32 hours after pollination and are indicated by solid colored lines. They illustrate the small level of difference of the two models. The black dots mark the average of the measured values of the nDNA content of the sperm cells, at different times after pollination. In addition, we have plotted an extrapolation of the fitted lines to the range of times 0 to 16 hours after pollination (indicated by colored dashed lines). According to observations from microscopy, sperm cells do not have an independent existence during most of the extrapolated time interval, and hence the lines cannot directly refer to the nDNA content of sperm cells. However, the intercepts of the lines (indicated by colored dots) occur at the values 1.16C and 1.14C which are both close to 1C, i.e., close to the haploid C content of one half a generative cell, which is a precursor of a sperm cell.

[International Journal of Plant Biology 2015; 6:5996]

ing the statistical results. We note that generally, the value of the nDNA content of sperm cells after pollination is likely to be influenced by the validity of two assumptions. The first assumption is that the measured nDNA errors, that arose from the fluorescence technique, were *unbiased*. That is, we assumed the distribution of the fluorescence errors were symmetrically distributed around zero.

The second assumption is that the value of the nDNA content of sperm cells, for a range of times after pollination, increases *linearly* with time. This assumption is the simplest that can be made, and is likely to break down at sufficiently long times, and lead to a *saturation effect*, since once an nDNA content of $2C$ has been achieved in a sperm cell, no further change in the sperm cell will be observed, prior to zygote formation.

The statistical analysis of both models (see the Statistical Analysis of the models section) allows their comparison, using the Deviance Information Criterion, which is known to combine goodness of fit with a penalty associated with the number of parameters contained in a model.[12] It appears that on the basis of the experimental data, there is no benefit in adopting the more complex model (Model 1) to summarize the data. Certainly, the variance in the rate of nDNA accumulation, for Model 1, is small: from Table 2, we have $\sigma^2_R = 1.40 \times 10^{-5}$. For different data sets, however, Model 1 may

be a more appropriate and more useful description.

References

1. Hu Sh Y. Reproductive biology of angiosperms. Beijing: Higher Education Press; 2005.
2. Friedman WE. Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development 1999;126:1065-75.
3. Tian HQ, Yuan T, Russell SD. Relationship between double fertilization and the cell cycle in male and female gametes of tobacco. Sex Plant Reprod 2005;17:243-52.
4. Tian M, Wang M. Studies on extraction, isolation, and composition of Lycium barbarum polysaccharides. J Trad Chin Herb Drugs 2006;31:1603-7.
5. Zheng GQ. The Studies on the Relationship between the structure, development and sugar accumulation in fruits of Lycium barbarum L. Degree Dissertation. Northwestern University of China, China; 2011.
6. Tang WC, Eisenbrand G. Lycium barbarum L. and L. chinensis Mill. In: Tang WD, Eisenbrand G, eds. Chinese drugs of plant origin: chemistry, pharmacology, and use in traditional and modern medicine. 1st ed. Berlin: Springer; 1992. pp 633-638.
7. Deng H, Wang YY, Zhu XY, Tian HQ. Generative and sperm cell isolation in Bauhinia blakeana (Fabaceae). Ann Bot Fennici 2012;49:1-6.
8. Ge LL, Gou XP, Yuan T, et al. Migration of sperm cells during pollen tube elongation in Arabidopsis thaliana: behavior during transport, maturation and upon dissociation of male germ unit associations. Planta 2011;233:325-32.
9. Qiu YL, Yang YH, Zhang SQ, Tian HQ. Isolation of two populations of sperm cells from the pollen tube of Tobacco. Acta Botanica Sinica 2004;46:719-23.
10. Deng H, Song YX, Qin K, Tian HQ. DNA content and cell cycle changes of male and female gametes of Lycium barbarum L. Plant Physiol J 2012;48:869-73
11. Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. Am Stat Assoc 1995;49:327-35.
12. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc, B 2002;64:583-639.