Modeling Interactions between Transposable Elements and the Plant Epigenetic Response: A Surprising Reliance on Element Retention

Kyria Roessler[1], Alexandros Bousios[2], Esteban Meca[3], and Brandon S. Gaut[1,*]

[1]Department of Ecology and Evolutionary Biology, UC Irvine

[2]School of Life Sciences, University of Sussex, United Kingdom

[3]Departamento de Agronomia, Universidad de Cordoba, Spain

*Corresponding author: E-mail: bgaut@uci.edu.

Accepted: February 20, 2018

Abstract

Transposable elements (TEs) compose the majority of angiosperm DNA. Plants counteract TE activity by silencing them epigenetically. One form of epigenetic silencing requires 21–22 nt small interfering RNAs that act to degrade TE mRNA and may also trigger DNA methylation. DNA methylation is reinforced by a second mechanism, the RNA-dependent DNA methylation (RdDM) pathway. RdDM relies on 24 nt small interfering RNAs and ultimately establishes TEs in a quiescent state. These host factors interact at a systems level, but there have been no system level analyses of their interactions. Here, we define a deterministic model that represents the propagation of active TEs, aspects of the host response and the accumulation of silenced TEs. We describe general properties of the model and also fit it to biological data in order to explore two questions. The first is why two overlapping pathways are maintained, given that both are likely energetically expensive. Under our model, RdDM silenced TEs effectively even when the initiation of silencing was weak. This relationship implies that only a small amount of *RNAi* is needed to initiate TE silencing, but reinforcement by RdDM is necessary to efficiently counter TE propagation. Second, we investigated the reliance of the host response on rates of TE deletion. The model predicted that low levels of deletion lead to few active TEs, suggesting that silencing is most efficient when methylated TEs are retained in the genome, thereby providing one explanation for the large size of plant genomes.

Key words: methylation, silencing, transposable elements, genome size.

Introduction

Angiosperm genomes vary >1,000-fold in size, and this variation correlates strongly with transposable element (TE) content. For plant species with small genomes, like *Arabidopsis thaliana* or *Brachypodium distachyon*, DNA derived from TEs constitute 20–30% of the genome (AGI 2000; IBI 2010). Species with larger genomes have commensurately larger proportions of TE-derived DNA. For example, TE-derived DNA represents >85% of the barley (*Hordeum vulgare*) and maize (*Zea mays* ssp. *mays*) genomes (Wicker et al. 2004; Schnable et al. 2009). When one considers that the average size of a diploid angiosperm genome is similar to that of barley genome, at 6400 Mb, then it is clear that most extant plant DNA is derived from TEs (Tenaillon et al. 2010).

Despite the obvious evolutionary success of TEs, the plant host checks their proliferation. The two entities engage in a continuous arms-race, where TEs seek to proliferate and the host attempts to control them (Lisch and Slotkin 2011). In fact, most—but not all (Li *et al.* 2010)—TEs are epigenetically silenced under normal conditions (Lisch 2009). The plant host exerts this control by suppressing TE activity both before and after transcription. Posttranscriptional modification relies chiefly on *RNAi* that recognizes and degrades TE mRNA produced by RNA polymerase II (*Pol II*). Degradation requires associated factors like *RNA-polymerase 6* (*RDR6*), which converts single-stranded to double-stranded RNA (dsRNA); the *Dicer-like* proteins *DCL2* and *DCL4* that cleave dsRNAs to produce 21 and 22 nucleotide (nt) small interfering RNAs

(siRNAs); and the *Argonaute1* (*AGO1*) protein that guides siRNAs to mRNAs for cleavage (Fultz et al. 2015). Presumably, 21–22 nt siRNAs can prime multiple cycles of mRNA cleavage, but they may have another important function, which is to initiate transcriptional silencing (Nuthikattu et al. 2013; McCue et al. 2015). Hence, 21–22 nt siRNAs can be seen as dual-purpose, because they are involved in post-transcriptional silencing and also because they initiate DNA methylation (Cuerda-Gil and Slotkin 2016).

Transcriptional silencing is achieved through epigenetic modifications like DNA methylation, histone modifications, and shifts in nucleosome positioning (Bernatavichute et al. 2008; Chodavarapu et al. 2010). The first of these, DNA methylation, relies on the RNA-directed DNA methylation (RdDM) pathway. RdDM begins when the plant-specific RNA polymerase *Pol IV* transcribes a TE. The resulting single-stranded RNA is processed into 24 nt siRNAs by *RDR2* and *DCL3*, two homologs that are distinct from those employed in *RNAi*. Ultimately, the 24 nt siRNAs guide protein complexes to homologous DNA sequences that are then targeted for cytosine methylation. Once DNA methylation is established, at least two mechanisms act to maintain it. The first is a positive feedback loop: *Pol IV* and *Pol V*, the RNA polymerases involved in RdDM, preferentially act on methylated DNA (Law et al. 2013; Johnson et al. 2014), thereby reinforcing silencing (Panda and Slotkin 2013). The second is the maintenance of symmetric CG and CHG (where H = A, C, or T) methylation during DNA replication and cell division (Law and Jacobsen 2010). Although the switch from RdDM to maintenance is not well understood (Panda and Slotkin 2013), once a TE is targeted for DNA methylation the host genome employs feedbacks to ensure that the TE reaches and maintains a quiescent state.

Numerous molecular studies have characterized the *RNAi* and RdDM pathways (reviewed in Law and Jacobsen 2010; Fultz et al. 2015; Matzke et al. 2015). These have been complemented by evolutionary studies showing that small RNAs are used for TE defense across both prokaryotes and eukaryotes (Blumenstiel 2011) and that most *RNAi* and RdDM components are present in early land plant lineages (Huang et al. 2015; Ma et al. 2015; Zhang et al. 2015; Tsuzuki et al. 2016). However, several important questions remain about systems-level interactions between TEs and their plant hosts. One major question is why the host relies on two mechanisms—that is, *RNAi* and RdDM—to silence TEs. Presumably both pathways are capable of silencing; they are thus overlapping and potentially redundant. Both require the production of myriad polymerases, methylases and small RNAs and therefore must have some energetic cost (Bousios and Gaut 2016). Why, then, are two pathways maintained? One working hypothesis is that they act synergistically, but this hypothesis has yet to be explored.

A second major question concerns 24 nt siRNAs. As mentioned earlier, 24 nt siRNAs are predominantly produced by the RdDM pathway, which preferentially acts on TEs that have already been targeted for silencing. An important feature of these 24 nt siRNAs is that they can act in *trans* to guide the methylation of TEs that have similar sequence characteristics to the original TE template (Slotkin et al. 2005; Teixeira et al. 2009; Ito et al. 2011; Ye et al. 2012; Fultz et al. 2015). Under this process, 24 nt siRNAs may constitute a kind of "immune memory" that act as a buffer against the possibility of TE activity (Fultz et al. 2015). If true, this implies that the strength of the host epigenetic response is related to the number of similar TEs in the genome that have already been silenced. Yet, no studies have explored the potential codependence between TE copy numbers and the strength of the host response.

Our final systems-level question concerns a separate process that occurs in cells associated with (but not part of) the germline. In cells such as the pollen vegetative nucleus (Slotkin et al. 2009), some TEs are actively demethylated, expressed, and utilized to produce 21–22 nt siRNAs. These siRNAs are then transported to the germline, where they presumably contribute to stable TE silencing across generations (Slotkin et al. 2009; Ibarra et al. 2012; Martínez et al. 2016; Martinez and Köhler 2017). But what is the systems-level benefit of this additional step in the host response, given that there are already at least two overlapping pathways dedicated to silencing TEs and also that symmetric DNA methylation is typically inherited faithfully?

Here, we address these questions by building a model of host: TE interactions based on ordinary differential equations (ODEs). ODE models have been used widely to study biological phenomena that range from population growth (Malthus 1798), to predator–prey interactions (Volterra 1926), to the dynamics of viral infection and reproduction (Perelson 2002). ODE models have also studied the interactions between TEs and the host response (Abrusán and Krambeck 2006), but without a focus on plants and with few details of host response mechanisms. Our model includes proxies for *RNAi*, RdDM, and additional factors like TE propagation and TE deletion. We study properties of the model but also estimate reasonable biological parameters by fitting the model to biological data, specifically from the study of the accumulation of the *Evade* element in an *A. thaliana* inbred line (Mari-Ordonez et al. 2013). Given these parameter estimates, we explore dynamics of the model and address systems-level questions about host: TE interactions. We focus on three sets of questions: 1) Are both pre- and posttranscriptional silencing necessary to control TEs? If not, what advantage is gained by having two mechanisms? 2) Given that methylated TEs may be an important source of immune memory, does TE deletion affect the dynamics of the host response? And, finally, 3) What is the added benefit of a third mechanism for generating 21–22 nt siRNAs in the germline?

Materials and Methods

Equilibria and Stability

Given our ODE model, its equations and its parameters (see Results), we found equilibria by solving for active TEs (aTEs), silenced TEs (sTEs) and siRNAs when all equations were equal to zero. The first, trivial equilibrium point was $aTE_{eq} = sTE_{eq} = siRNA_{eq} = 0$. To derive the stability of this equilbrium, we calculated the Jacobian matrix for the ODEs around that equilibrium, which provided:

$$J_{TE}(0,\ 0,\ 0) = \begin{bmatrix} v*p-d & 0 & 0 \\ 0 & -d & 0 \\ \varepsilon*v & 0 & -1 \end{bmatrix}.$$

The resulting characteristic equation is:

$$\begin{aligned} \det(\ J_{TE}(0,\ 0,\ 0) - \ \lambda \cdot I) = \ & 0 \\ = \lambda^3 + vp\lambda^3 + \lambda(vpd - vp + d^2) & \\ - vpd - \ d^2, & \end{aligned} \tag{1}$$

where the solutions of this equation are the eigenvalues. The equation clearly communicates that stability depends on a complex relationship among *v, p*, and *d* but only on these parameters. The critical eigenvalue (i.e., the one that crosses zero) is in fact ($p\ v - d$), and hence the stability of this equilibrium is controlled by this compound parameter. See supplementary text, Supplementary Material online, for additional details based on a rescaled model.

The second equilibrium point is shown in equations (3) and (4) (see Results) for aTE_{eq} and sTE_{eq}; the corresponding equation for $siRNA_{eq}$ is:

$$siRNA_{eq} = \frac{v}{\frac{r}{\varepsilon*d} - \frac{i}{\left(\frac{d}{v} - p\right).}} \tag{2}$$

We also examined the Jacobian matrix and eigenvalues to study stability for this equilibrium point. Because the stability equation was complex, we analyzed the rescaled model for further insights into the stability of this nontrivial equilibrium (see supplementary text, Supplementary Material online).

Fitted Parameters

We obtained the data from Mari-Ordonez et al. (2013) by loading their figure 3*a* onto WebPlotDigitizer (https://automeris.io/WebPlotDigitizer/, last accessed February 28, 2018). To estimate model parameters that fit the empirical data, we used the sum of least squares method, based on the following formula:

$$\text{sqEr} = \ \sum (E_{CN} - O_{CN})^2 + w \cdot \sum (E_{Exp} - O_{Exp})^2.$$

In this formula, E_{CN} and O_{CN} are the expected and observed copy number, respectively. The expected copy number was defined as the sum of aTEs and sTEs obtained from the model. E_{Exp} and O_{Exp} are the expected and observed values, respectively, for relative expression.

The expected relative expression for generation *n* was obtained from the model by taking the total expression in generation 8, which is equal to *v* multiplied by the aTE copy number at generation 8, and comparing that to the total expression at generation *n*, which is equal to *v* multiplied by the number of aTEs in generation *n*. Note, however, that our measure of relative expression may not correspond perfectly to that from Mari-Ordonez et al. (2013), because the empirical data on relative expression actually compares two genes (*Evade* and *ACT2*) within each generation and also because qRT-PCR can be inaccurate, especially when it is used as a ratio (of *ACT2* vs. *Evade* expression). In the square error (sqER) equation, we assigned *w* a weight of 40 to reflect the magnitude of difference in the empirical data, because copy number reached ~40 and relative expression plateaued at ~1 (fig. 2*A*).

We used a Monte Carlo approach to estimate fitted parameters. In this approach, all seven parameters were initialized with randomly drawn values from a uniform distribution between 0 and 1, except for *v*, which was ranged between 0 and 20. We also imposed the constraint that $p + \varepsilon \leq 1.0$. Given initial parameters, the sqEr was calculated as above. A single parameter was then altered, with a step size between −0.1 and 0.1 for all parameters (except *v* where step size was between −1.0 and 1.0). The sqEr was calculated and the iteration moved forward only if $\text{sqEr}_n > \text{sqEr}_{n+1}$; otherwise a new step size would be calculated. All the parameters (in the following order: *v, l, μ, p, i, r, ε*, and *δ*) were iterated through 100 times with 50 steps for each parameter, until the final fitted parameters were found with the smallest sqEr for each run. The initialization and iteration of all parameters was performed >10,000 times; the lowest sqEr across all 10,000 runs was used to define the fitted parameters. We note, however, that other fitted data sets with low sqEr values produced similar model dynamics (supplementary fig. S6, Supplementary Material online).

Running the ODE Model

The ODE model was run using odeint from the scipy.integrate package (https://docs.scipy.org/doc/scipy/reference/integrate.html, last accessed February 28, 2018) and python (v2.6.6). Figure 1*A* was made with draw.oi (https://www.draw.io/, last accessed February 28, 2018); all other figures were made with R (v. 3.3.2). The heatmaps were made with heatmap2, from the gplots library in R.

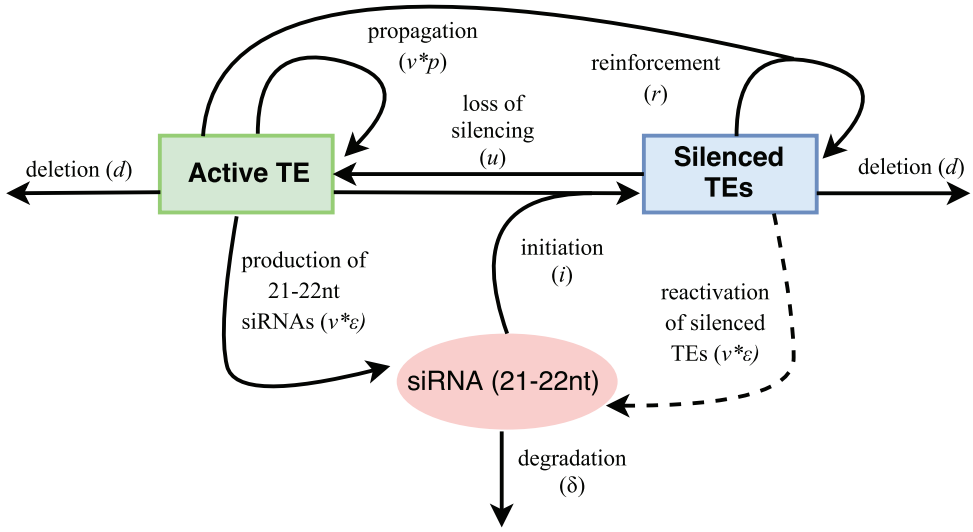

Fig. 1.—A schematic of the model, with details provided in the text. The dashed arrow represents a step specific to cells that contribute to germline material.

Table 1
Summary of Parameters and Their Fitted Estimates

Parameter	Description	Fitted Estimate
V	Amount of *Pol II* mRNA expressed by active TEs	1.630
P	Proportion of mRNA that contributes to transposition	0.340
ϵ	Proportion of mRNA that contributes to 21–22nt siRNA production	0.051
I	The rate at which 21–22nt siRNA initiate methylation	0.062
R	The rate at which 24nt siRNA reinforce methylation	0.025
D	The rate of TE deletion per generation	0.161
U	The rate of methylation loss per generation	0.000
Δ	The rate of degradation of 21–22nt siRNAs per generation	0.999

Results

A Model of TE Propagation and Silencing

Our model assumes that an active TE (aTE) begins as single copy and expresses mRNA at rate v (fig. 1 and table 1). Among the produced mRNA, a proportion p is transposed into new genomic copies of the TE per host generation. Another proportion, ε, of the TE mRNA is processed into 21–22 nt siRNAs. Note that $p+\varepsilon \leq 1.0$ under our model. We assume that the 21–22 nt siRNAs degrade at rate δ and initiate TE silencing at rate i. Initiation encompasses both post-transcriptional silencing (*RNAi*) and the onset of methylation, following previous models (Nuthikattu et al. 2013; McCue et al. 2015). Finally, 24 nt siRNAs reinforce methylation at rate r, representing RdDM. In our model, the amount of 24 nt siRNA is proportional to the number of silenced TEs (sTEs). Furthermore, 24 nt siRNAs are considered to be *trans*-acting and thus may affect numerous TE insertions, including active elements. Overall, active TEs (aTEs) may become sTEs through 21–22 nt siRNAs, 24 nt siRNAs, or by a combination of both (fig. 1).

The model includes two additional parameters. The first is TE deletion from the genome, which occurs at rate d for both aTEs and sTEs. The second is the potential for the loss of silencing from TEs over time (e.g., through the loss of methylation), which we assume can lead to reactivation of TEs at rate u. When $u=0$, maintenance of silencing is perfect, but silencing is not maintained when $u=1$.

The model is represented diagrammatically in figure 1 and consists of three differential equations:

$$\frac{d(\mathrm{aTE})}{dt} = (v \cdot \mathrm{p} - d - i \cdot \mathrm{siRNA} - r \cdot \mathrm{sTE}) \cdot \mathrm{aTE} \; + \; u \cdot \mathrm{sTE},$$

$$\frac{d(\mathrm{sTE})}{dt} = (i \cdot \mathrm{siRNA} + r \cdot \mathrm{sTE}) \cdot \mathrm{aTE} - (d+u) \cdot \mathrm{sTE},$$

$$\frac{d(\mathrm{siRNA})}{dt} = \varepsilon \cdot v \cdot \mathrm{aTE} - \delta \cdot \mathrm{siRNA}.$$

The first equation describes the change in the number of aTEs over time; the second describes the change in the number of sTEs over time, and the third monitors numbers of 21–22 nt siRNAs over time. Although these three equations represent our basic model, figure 1 includes a dashed arrow representing a fourth process, the epigenetic remodeling of TEs in the germline. This process will be incorporated after we first explore the dynamics of the basic model.

We recognize that the model, as presented, is overparameterized in a mathematical sense. For example, the parameter v could be eliminated by redefining ε and p. We retain the parameter definitions provided above throughout the main text, because we believe their meaning to be biologically intuitive. However, in the supplementary text, Supplementary Material online, we also present a mathematical treatment that includes parameter reduction, rescaling, and more extensive derivations of the model's analytical properties. We refer to this supplement throughout the main text, where appropriate.

Model Equilibria

Once a TE has invaded a host it has three possible fates: it may fail to successfully invade and be lost completely; it may establish itself and reach an equilibrium number of copies over time; or it may expand in copy number unabated. An advantage of ODE models is that we can analytically solve the equilibrium points to understand TE invasion behavior and parameter dependence. We analyzed equilibria and the stability of those equilibria. For these analyses we assumed $u = 0$ and $\delta = 1$ for simplicity, but also because it is biologically reasonable to assume both that maintenance of the silenced state is strong ($u = 0$), based on the conservation of symmetric methylation, and that siRNAs degrade rapidly ($\delta = 1$).

We identified two equilibrium points in our system. The first is when there are no TEs and, hence, no 21–22 nt and 24 nt siRNAs in the host. That is, the equilibrium points for the active copies (aTE_{eq}), silenced copies (sTE_{eq}), and siRNA ($siRNA_{eq}$) are equal to zero. Stability around this point provides information as to whether a TE will successfully invade the genome or be lost. We investigated stability (see Materials and Methods; see eq. 1) and found that it does not rely on any of the parameters associated with epigenetic processes—that is, i, ε, or r. Instead, stability relies only on the parameters for TE expression, propagation, and deletion (v, p, and d; see also supplementary text, Supplementary Material online). Although equation (1) is complex, the Jacobian matrix (see Materials and Methods) suggests the intuitive notion that invasion proceeds when expression and propagation ($v \cdot p$) outcompetes deletion (d).

Once a TE has established its presence in the host, it may increase in number until the second, nontrivial equilibrium point (see Materials and Methods). The equilibrium points for aTEs and sTEs are given by:

$$aTE_{eq} = \frac{1}{\frac{r}{d} - \frac{i\cdot\varepsilon}{\left(\frac{d}{v}-p\right),}} \quad (3)$$

$$sTE_{eq} = \frac{\left(\frac{v\cdot p}{d}\right) - 1}{\frac{r}{d} - \frac{i\cdot\varepsilon}{\left(\frac{d}{v}-p\right)}} \quad (4)$$

with $siRNA_{eq}$ given by equation (2) (see Materials and Methods). These two equations illustrate that aTEs and sTEs have similar parameter dependencies. However, equilibrium values of sTEs depend more explicitly on v and p in the numerator than does the equilibrium values of aTEs. This is an interesting observation because v and p are properties of aTEs; it drives home the point that equilibria copy numbers of sTEs relies intricately on the properties of their active counterparts. The denominator of the two equations clearly indicates that increasing r tends to decrease both aTE_{eq} and sTE_{eq}. Finally, the equations also hint at a complex relationship between equilibrium copy numbers and d, because the latter appears twice in the denominator (and once in the nominator for sTEeq). As d increases, these appearances have opposite effects on equilibrium values.

We studied these equilibria using an equivalent rescaled model (see supplementary text, Supplementary Material online). Our analytical results provided additional insights about the behavior of the model and particularly the stability of the nontrivial equilibrium. For example, the nontrivial equilibrium is stable when ($pv - d$) is positive and unstable when this is negative. Conversely, the trivial equilibrium is stable when ($pv - d$) is negative, which implies that both equilibria exchange their stability for ($pv - d$)=0.

Fitting the Model to Biological Data

It can be difficult to identify biologically reasonable parameter values for ODE models. To address this concern, we fitted the model to biological data and then perturbed parameter values separately to explore parameter dependencies and to assess effects on TE copy numbers. We fitted the model to experimental data from the study of Mari-Ordonez et al. (2013), who characterized the expression and transposition of a single-copy of the *Evade* retroelement that had become unmethylated in *A. thaliana met1*-mutant epigenetic recombinant inbred lines. By following two lines to generations 14 and 15, they showed that *Evade* was highly expressed until generation 11 and 7, respectively, after which expression plummeted precipitously, presumably due to host silencing. The number of *Evade* copies increased rapidly while its expression was high, to a maximum of ~40 copies after 11 and 7 generations.

To fit our model to their data, we extracted information about *Evade* copy numbers and relative expression (see Materials and Methods). We focused on one inbred line (*met1)* from their study, because this was the only line for which data were sampled for consecutive generations: in total seven generations (from 8 to 14) since the reactivation of the single *Evade* element. We fitted the model to the *Evade* data

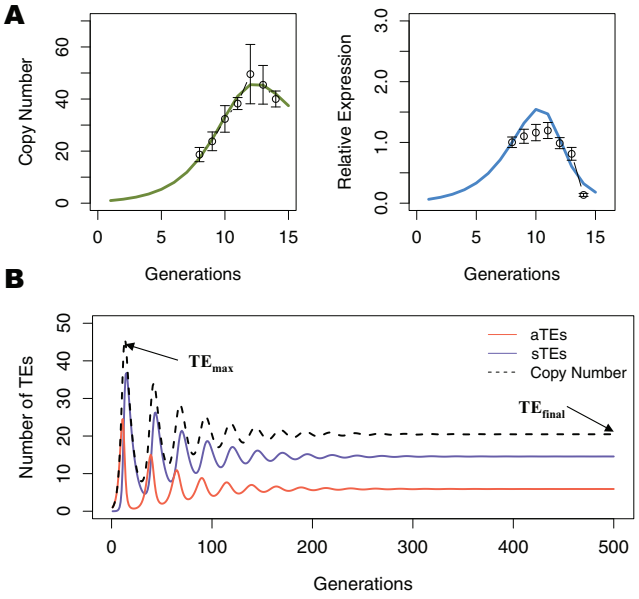

Fig. 2.—(*A*) Model fit to the *Evade* data for total copy number (left) and relative expression (right). The empirical data from the *Evade* study are represented by circles; the whiskers indicate SD. The model results based on the fitted parameters (table 1) are represented by the solid line. (*B*) Long-term behavior of the model, based on the fitted parameters to the *Evade* data. Arrows show TE_{max} and TE_{final}, which are defined in the text. Copy number refers to the summation of aTEs and sTEs.

with a Monte Carlo approach that concurrently considered the total TE copy number (i.e., the combined total of aTEs and sTEs) and TE expression. Our set of fitted parameter values are reported in table 1. These parameter values produced a good fit to the copy number data, and a curve of similar shape to the observed relative expression data over time (fig. 2*A*). (Note that our measure of expression is only a proxy for expression measured experimentally; see Materials and Methods.) We recognize that we have fitted a complex model to relatively simple data and that our fitted parameters may represent one of many potential reasonably fitting parameter sets (but see Materials and Methods). They nonetheless provide a biologically plausible foundation for examining model behavior.

Model Behavior under Fitted Parameters

Given the fitted parameters, we explored host: TE dynamics over 500 generations, monitoring numbers of aTEs, sTEs, and total TE copy number (=aTEs + sTEs) (fig. 2*B*). With these parameter values, the model produces oscillations of all three entities for ~200 generations until it reaches an equilibrium. The oscillations of aTEs and sTEs are somewhat out of phase with one another. We interpret these results as reflecting feedbacks in the epigenetic system. When a TE first invades a host, the combination of expression (fitted value $v = 1.63$; table 1) and propagation ($p = 0.340$) create an initial burst in TE copy number. If TEs were able to grow unabated, there would be an exponential increase at a rate of 0.554 ($=pv$) TEs per generation. However, some transcripts are processed into 21–22 nt siRNAs ($\varepsilon = 0.051$) that silence TEs at rate $i = 0.062$. These 21–22 nt siRNAs degrade quickly for each host generation ($\delta = 0.999$), and therefore any new 21–22 nt siRNAs are not residual, but must be made from active TEs. Once initiation of methylation has begun as part of i, reinforcement quickly takes hold at rate $r = 0.025$. Eventually, the number of sTEs increases and the number of aTEs decreases, so that total expression begins to decline.

As TEs become silenced, they have two fates under our model: they can be deleted from the genome or become active again due to loss of silencing (fig. 1). Since loss of silencing was very low ($u = 4 \times 10^{-6}$) in the fitted parameter set, the main fate of sTEs is to be deleted ($d = 0.16$). As these quiescent TEs are lost, so is the source of reinforcing 24 nt siRNAs. When reinforcement becomes unreliable, the host loses epigenetic control, the subset of remaining aTEs propagate, and the phased cycle begins again. These cycles dissipate in amplitude until equilibria are reached at ~20 total TE copies, with more sTEs (~14) than aTEs (~6) (fig. 2*B*). It is important to note that the equilibrium is not necessarily static; it can be reached when equal numbers of TEs are created versus deleted.

These phased interactions occur with the fitted parameters, but decaying oscillations in copy number also occur regularly with other parameter combinations (see supplementary text and fig. S1, Supplementary Material online). Oscillating TE numbers are not, however, a necessary outcome of the model (see examples below and supplementary text, Supplementary Material online).

Examining Initiation (*i*) and Reinforcement (*r*)

We have shown that the model can have complex, oscillating dynamics based on parameters inferred from biological data. These parameters can be modified independently to explore the importance of various processes. In this section, we assess the effect of perturbing the system by varying either initiation (i) or reinforcement (r), or both, while holding the remaining parameters to the values estimated from the *Evade* data. We first set $i = 0$, and the result was both intuitive and trivial. With $i = 0$ silencing never begins. Hence, the number of aTEs trended upward at an exponential rate, with no resulting sTEs (fig. 3*A*).

The effect of setting r to zero was less straightforward, because i was > 0 and hence silencing was initiated. Without reinforcement, copy numbers no longer oscillated, but instead burst and rapidly reached a maximum for both aTEs and sTEs. These copy numbers remained flat, implying a steady state in which silencing was initiated by 21–22 siRNAs and there was sufficient transposition to counteract TE deletion. Under these parameter values, the steady state of sTEs

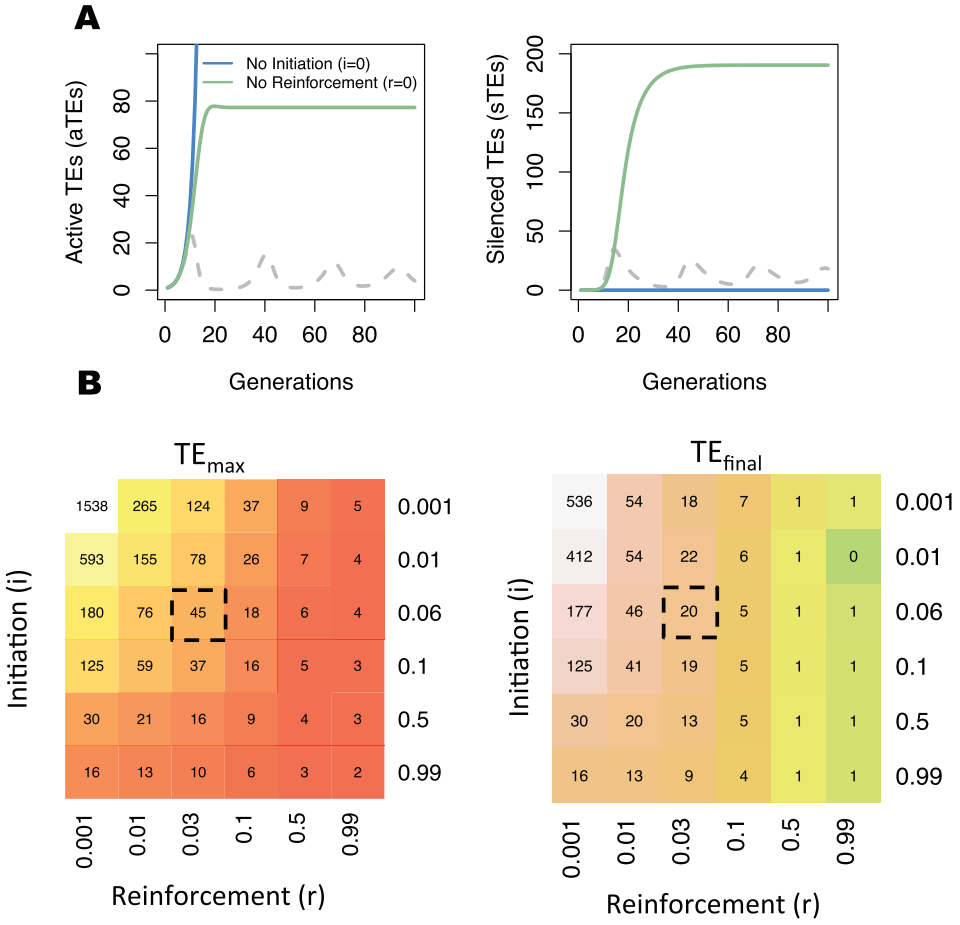

Fig. 3.—Model behavior with the fitted values for all parameters but initiation (*i*) and reinforcement (*r*). (*A*) Graphs illustrate the effect of setting initiation and reinforcement parameters to zero for active TEs (left) and methylated TEs (right). In both graphs, the gray dashed lines represent the number of TEs based on the fitted model parameters to the *Evade* data (see also fig. 2*B*). (*B*) Heat maps showing the TE_{max} (left) and TE_{final} (right) for the total copy number (= aTEs + sTEs) based on varied values of initiation (*y*-axis) and reinforcement (*x*-axis), with copy number displayed in each cell. The dashed cell in each heat map represents the fitted values (table 1).

was higher than that of aTEs (fig. 3*A*), as with the equilibria reached with fitted parameters (fig. 2*B*).

If initiation by 21–22 nt siRNAs is sufficient to reach a steady state and to control TEs, then what is the advantage of reinforcement by 24 nt siRNAs? Equations (3) and (4) show that the steady-state TE copy numbers depend critically on the values of *r* and *i*, since the denominators become very small where *r* and *i* are small. To explore their interdependencies, we varied *i* and *r* across their parameter ranges and assessed total copy numbers (=aTEs + sTEs). To help characterize effects, we focused on two descriptive statistics, TE_{max} and TE_{final} (see fig. 2*B*). TE_{max} is the highest total TE copy number achieved under a set of model parameters, and TE_{final} is the total copy number after 5,000 generations, a point by which total aTE and sTE copy numbers have typically reached a steady state. Our analyses show that when $r \geq 0.5$, any change in *i* had little effect on TE_{max} and TE_{final}, so long as there was at least some initiation (fig. 3*B*). In contrast, when *r* was low (e.g., $r \leq 0.1$), the value of *i* had notable effects on both TE_{max} and TE_{final}. For example, when $r = 0.001$, TE_{max} varied over two orders of magnitude as a function of *i*. Similarly, TE_{final} differed ~33-fold when *i* ranged from 0.001 to 0.99 (fig. 3*B*). This relationship implies that reinforcement can counter TE propagation efficiently, even when initiation of silencing is weak. This observation held true when also adjusting for TE expression (*v*) and deletion (*d*) (supplementary fig. S2, Supplementary Material online).

The Effects of TE Deletion (*d*)

Theoretically, high TE deletion rates should be advantageous for the plant host, because they limit opportunities for transposition and consequent deleterious mutations. However, high amounts of TE deletion could have consequences for immune memory, because quiescent TEs may be a major source of *trans*-acting 24 nt siRNAs (Teixeira et al. 2009; Ito

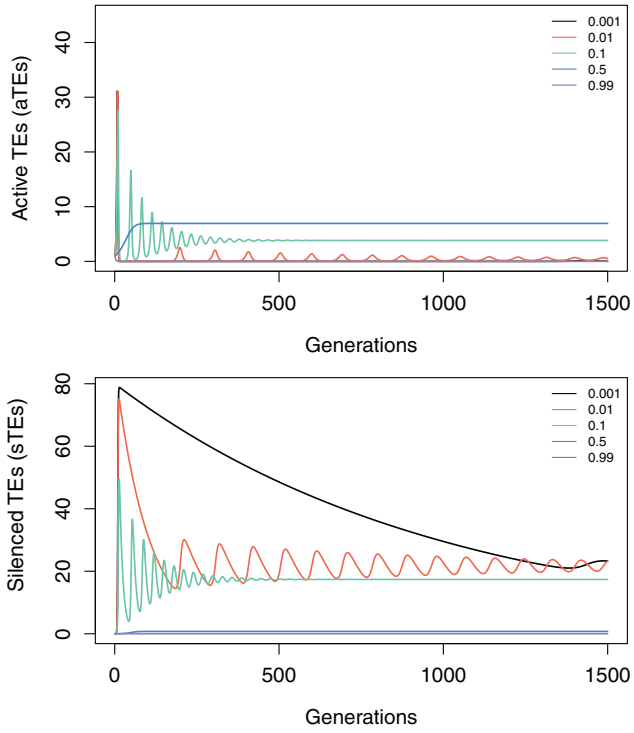

Fig. 4.—Model behavior with the fitted values for all parameters but TE deletion (d), which is varied from 0.001 to 0.99.

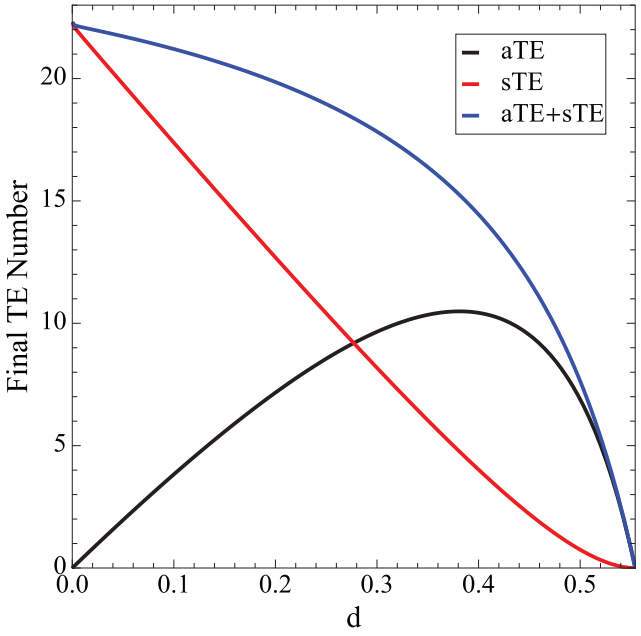

Fig. 5.—The effect of varying rates of TE deletion (d) on the final number of silenced TEs (sTE), active TE (aTE) and total copy number (aTE + sTE). These calculations used the fitted values for all parameters but d. Note also that d begins at an arbitrarily low value of 0.001; when $d = 0$, the number of sTEs diverges.

et al. 2011; Fultz et al. 2015). Hence, high deletion rates may adversely affect the epigenetic response. To illustrate the effect of deletion on TE copy numbers, we varied the deletion parameter d from 0.001 to 0.99 (fig. 4), while holding the remaining parameters to their fitted values (table 1).

The model produced four noteworthy results. First, when TE deletion was very low ($d = 0.001$), aTEs burst quickly to high copy number (~30). After peaking at a total copy number of ~80, all TEs were silenced and the population of sTEs declined slowly over time, reflecting the low rate of deletion (fig. 4). Throughout this process, there were no aTEs after the initial burst. Second, when d increased ($0.01 \leq d < 0.5$), the system generated oscillations in the number of aTE and sTEs. The amplitude, frequency, and equilibrium values (i.e., TE_{final}) varied with d. Note that the running average of sTEs exceeded that of aTEs for these parameter values (fig. 4). Third, when TE deletion was at intermediate levels ($d = 0.5$), aTEs reached a steady state, but there were very few sTEs. Finally, when the rate of TE deletion was very high ($d = 0.99$), all TEs were removed from the genome.

To further illustrate these dependencies on d, we plotted TE_{final} for sTEs, aTEs, and all TEs as a function of d (fig. 5). Overall, these results convey a somewhat counterintuitive idea: if the goal is to have few aTEs, then it is beneficial either to have dramatically high rates of TE deletion (e.g., $d = 0.99$) or to have such low (e.g., 0.01–0.1) deletion rates that a reservoir of sTEs is preserved and contributes to reinforcement of silencing. This supports our observation, based on equilibrium equations (eqs. 3 and 4), that deletion plays a complex role in determining aTE_{eq} and sTE_{eq}.

Additional Parameters

We also varied values of expression (v), propagation (p), and loss of silencing (u), while the remaining parameters were held at their fitted values. The parameter v was arbitrarily ranged between 0 and 5. The chief effect of this range was on the amplitude and periodicity of TE oscillations. Higher expression levels led to more dramatic copy number oscillations (supplementary fig. S3, Supplementary Material online). Importantly, at low parameter values (e.g., $v \leq 0.5$) TEs either did not invade the genome or were maintained at very low copy numbers (<5 total TEs) over the long term. Varying p produced results similar to varying v (supplementary fig. S4, Supplementary Material online). Increasing p did, however, tend to lead to higher TE_{max} and average copy numbers relative to the parameter values we explored for v (supplementary fig. S3, Supplementary Material online). This was presumably because there is a trade off with v; as it increases, so does the production of 21–22 nt siRNAs, which then potentially affect *RNAi*. Propagation (p), on the other hand, contributes only to the proliferation of more TEs. Note that low levels of propagation ($p < 0.25$) resulted

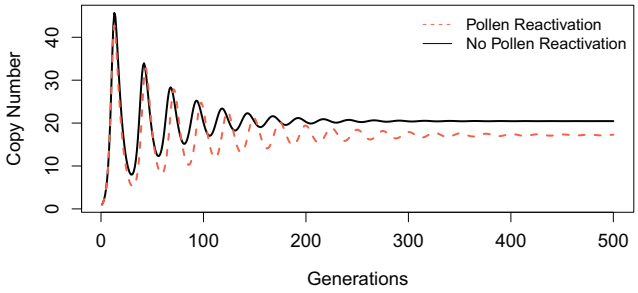

Fig. 6.—TE reactivation in pollen. The black line is based on the model with fitted parameters (no pollen reactivation); the dashed line is using the same parameters but including additional feedback for pollen guard cells (pollen reactivation). Both lines indicate total copy numbers (=aTEs + sTEs). The additional mechanism in pollen guard cells is denoted by the dashed arrow in figure 1.

in no invasion. Hence, TEs cannot invade if expression or propagation is low.

Our model also assumes a process of silencing loss (u), for which the most likely example is methylation loss. Methylation loss is known to be low based on empirical data because symmetric methylation is typically maintained faithfully through cell division (Becker et al. 2011). Indeed, our fitted parameter estimate was $u = 4 \times 10^{-6}$, suggesting that a very low amount of sTEs become aTEs due to, for example, leaky maintenance of symmetrical methylation. Overall, we found that varying the u parameter had little effect on model behavior at parameter values < 0.01 (supplementary fig. S5, Supplementary Material online). This implies that variation in spontaneous demethylation rates is likely to have few effects on the dynamics of host: TE interactions unless u varies by several orders of magnitudes from our fitted estimate.

TE Reactivation Dampens TE Oscillations

Finally, we incorporated an interesting biological observation—that is, the fact that TEs are activated in some reproductive tissues, ostensibly to ensure the transmission of a complement of siRNAs to egg and sperm (Slotkin et al. 2009; Ibarra et al. 2012; Martínez et al. 2016; Martinez and Köhler 2017). TEs are known, for example, to be demethylated and reactivated in the pollen vegetative nucleus, which accompanies the sperm cell, but does not contribute DNA to the fertilized zygote. The reactivated TEs are sources of 21–22 nt sRNAs that are transported to the sperm and presumably target silencing of TEs in the zygote (Slotkin et al. 2009). The net effect of this process is to increase the numbers of 21–22 nt siRNAs in germline cells; these 21–22 nt sRNA originate not only from aTEs but also from sTEs (see below).

We added this mechanism to our model with an equation that increases the number of 21–22 nt siRNAs in the system at a level proportional to the number of sTEs that were demethylated in the companion cells. That is,

$$\frac{d\text{siRNA}}{dt} = \varepsilon \cdot v \cdot (\text{aTE} + \text{sTE}) - \delta \cdot \text{siRNA}.$$

This equation is represented by the dotted arrow in figure 1. We evaluated the effects of this additional process on the system with fitted parameter values. The effects were consistent: it decreased TE_{max}, TE_{final} and the periodicity of copy number oscillations (fig. 6). Thus, this additional process yields notable decrements in TE copy numbers.

Discussion

In this study, we have devised an ODE model to examine the systems dynamics of TE propagation within the context of the epigenetic response of a plant host (fig. 1). Although there are clear limitations to our approach, the model has produced at least four fundamental insights. The first is the prediction of oscillating copy numbers typified by a burst of TE activity, followed by silencing, deletion, and then reactivity. Despite these oscillations, the system often reached equilibrium copy numbers (fig. 2). Second, our model emphasizes the importance of reinforcement by RdDM-like processes, because it buffers potential upstream inefficiencies in the initiation of silencing (fig. 3). Third, we show that these outcomes are linked to the rate of TE deletion. Somewhat nonintuitively, the model predicts that either low or very high levels of deletion lead to more efficient control of the number of aTEs (figs. 4 and 5). Finally, we show that demethylation within germline cells reinforces host defenses by dampening TE bursts and lowering steady-state copy numbers (fig. 6). Below, we first discuss the caveats of our ODE model before placing our insights into the context of plant genome structure and evolution.

Caveats

Every model has limitations, and ours is no exception. One important consideration is that our biological knowledge of the host response is incomplete. For example, the details of the initiation of methylation are not yet clear, because there are at least two competing (but likely nonexclusive) hypotheses as to how the host transitions from *RNAi* to the RdDM response (Mari-Ordonez et al. 2013; Nuthikattu et al. 2013; McCue et al. 2015). Furthermore, some aspects of the host response have not been included in our model, such as recent discoveries that 18–22 nt tRNA fragments (Martinez et al. 2017; Schorn et al. 2017) and some miRNAs (Creasey et al. 2014) may interfere with TE replication and propagation. However, these additional host mechanisms fit relatively easily in our model, because they would likely affect conversion (ε) and initiation (i) (fig. 1). In this sense, our model already

implicitly accounts for some exciting new findings, but other new insights may require model modifications.

Another limitation is that we have studied the invasion of only one TE family. In reality, plant genomes harbor a multitude of TE types that may interact with each other and also vary with respect to the host response. For example, some but not all TE families in *A. thaliana* are recognized by endogenous miRNAs (Creasey et al. 2014), and short, nonautonomous DNA elements are methylated less efficiently than longer, autonomous elements (Hollister and Gaut 2009), perhaps in part due to biases in genomic location (Zemach et al. 2013). Finally, we have used only one data set to fit the model, which followed the invasion of the *Evade* TE for a short period of few host generations (Mari-Ordonez et al. 2013). The reliance on *Evade* reflects the fact that very few studies have monitored the copy number and expression of TEs within a plant genome over time, particularly beginning from recent invasion or reactivation. In short, we recognize the limitations of the empirical data, but they nonetheless allow a glimpse into model behavior under relevant parameter values.

Invasion and Oscillations

How long does it take to silence TEs in vivo? Our understanding of the duration and intensity of TE amplification bursts remains limited (Bousios and Gaut 2016). In order to be silenced, a TE must first invade. Based on our model and analyses of the stability of the first equilibrium point (where aTEs = sTEs = siRNA = 0), invasion depends on expression (v), propagation (p), and deletion (d) but not on downstream properties of the host response, such as conversion of TE transcripts to 21–22 nt siRNAs (ε), initiation (i), and reinforcement (r). Put simply, pv needs to outpace d for a TE to successfully invade the host. We also investigated invasion by modifying v and p from the fitted parameter values (table 1); invasion did not occur when expression or propagation were low ($v < 0.5$, supplementary fig. S3, Supplementary Material online; $p < 0.25$, supplementary fig. S3, Supplementary Material online).

Assuming a TE invades successfully, it has the potential to increase rapidly in copy number. Under our model, we found that copy numbers often oscillated before reaching an equilibrium (e.g., fig. 2). Mathematically, the prevalence of oscillations is related to the value of the expression $(v \cdot \varepsilon \cdot i)/r$ (see supplementary text, Supplementary Material online). Oscillations tend to occur when TE expression is low (v), the proportion of 21–22 nt siRNAs is low (ε), initiation (i) is low or reinforcement (r) is high (see supplementary text, Supplementary Material online).

Under many parameter values explored in this work, the maximum duration of a TE burst lasts for only a few dozen generations before they are temporarily silenced and decrease in copy number (figs. 2 and 4 and supplementary figs. S3–S5, Supplementary Material online). These results likely reflect our reliance on data from a study in which silencing occurred rapidly (Mari-Ordonez et al. 2013), but there is other experimental evidence that host defenses react quickly to silence active TEs within a few host generations (Teixeira et al. 2009; Fultz and Slotkin 2017), perhaps even more quickly than the host response to *Evade*.

It is interesting to note that these experimental studies contradict numerous genome-wide analyses, which suggest that TE families experience massive bursts lasting thousands or even millions of years (Piegu et al. 2006; Schnable et al. 2009; Bousios et al. 2012; Daron et al. 2014). One likely explanation for this incongruence may be the difficulty of resolving the occurrence of multiple rounds of episodic bursts within the expanded timeframes reported by the genome-wide studies. Limited resolution may be due to technical issues related to in silico TE identification, accurate age estimation, and perhaps even heterogeneous rates of TE sequence loss and decay across the genome (Tian et al. 2009). No matter the cause, the apparent gaps between experimental and genome-wide studies deserve further thought and consideration. Longer term experimental studies that monitor TE copy numbers over time and under different stress conditions would certainly be welcome contributions to our empirical understanding of host: TE interactions.

Equilibria

Another question is whether TEs reach long-term equilibria within a genome. In our model, the oscillations often reduce in intensity over time to reach a steady state (figs. 2 and 4 and supplementary figs. S3–S5, Supplementary Material online). In this equilibrium, sTEs are found in higher numbers than aTEs whenever $(pv)/d > 2$ (eqs. 1 and 2 and supplementary text, Supplementary Material online).

Our ODE-based approach regularly predicts two phases of host: TE dynamics: one shaped by oscillating changes in TE numbers, and another characterized by an equilibrium. TE evolution has been modeled extensively with population genetic approaches (Charlesworth and Charlesworth 1983; Charlesworth et al. 1994; Brookfield 2005; Le Rouzic and Deceliere 2005), and the basic models predict that TEs reach steady-state copy numbers after the first TE invasion through either a transposition-selection or transposition-deletion equilibrium. In other words, they do not predict oscillations prior to an equilibrium. In contrast, some studies have expanded their models to include TE sequence evolution or competition between TEs, and these often predict oscillations in TE copy numbers (Le Rouzic and Capy 2006; Le Rouzic, Boutin, et al. 2007; Le Rouzic, Dupas, et al. 2007). For example, Le Rouzic, Boutin et al. (2007) investigated host–parasite interactions between autonomous TEs and their nonautonomous counterparts, and they found oscillations in copy numbers between both entities. Notably, the oscillations continued

indefinitely; an equilibrium was rarely reached unless there were very low mutation rates and few adaptive TE insertions. Le Rouzic, Boutin et al (2007) and also Brookfield (2005) have argued that equilibria are reached under conditions that are probably unrealistic for in vivo TEs. This is because the parameters that affect TE dynamics such as selection, transposition, and deletion are likely to change at faster rates than the time required to reach an equilibrium. Our model does not include autonomous and nonautonomous TEs, nor does it allow perturbations in subsequent generations. Yet, the focus on active and sTEs may mimic some characteristics of host–parasite relationships and may contribute to our observed oscillating dynamics. We must caution, however, that our model is not explicitly evolutionary, because it does not consider fitness or population variation.

The Importance of Overlapping Mechanisms

Why do plants maintain two overlapping and energetically costly pathways (*RNAi* and RdDM) to silence TEs? Here, *i* encompasses posttranscriptional silencing and the initiation of methylation, and *r* represents RdDM (fig. 1). Our results show that only a small amount of *i* is needed to begin silencing of an unrecognized TE, but *r* is necessary to counter propagation efficiently. For example, the host maintains TE copy numbers at low levels even when *i* is inefficient (e.g., $i = 0.001$), so long as *r* reinforces silencing by a value of $r \geq 0.1$ (fig. 3*B*). *RNAi* is clearly not as efficient at limiting TE copy numbers when there is no RdDM, yet it is essential for silencing TEs (fig. 3*A*). Hence, to the extent the model is correct, it implies that plants must have *RNAi* to start the process of silencing, but RdDM vastly enhances host control over TEs. The inclusion of another, apparently overlapping mechanism—that is, the active demethylation of TEs in cells that contribute siRNAs to germline cells—further enhances host silencing (fig. 6).

Our data are consistent with the argument that 24 nt siRNAs are important for buffering TE activity, even though they seem unnecessary because most methylation is maintained independently of RdDM in heterochromatic regions (Zemach et al. 2013). In fact, it was recently shown that these heterochromatic regions also produce 24 nt siRNAs, albeit to a smaller extent (Li et al. 2015). These findings are consistent with the idea that 24 nt siRNAs may act as immune memory (Fultz et al. 2015; Fultz and Slotkin 2017), based on evidence that they may play a key role in suppressing reactivated TEs (Teixeira et al. 2009; Ito et al. 2011; Fultz et al. 2015; Fultz and Slotkin 2017).

The Curious Case of TE Deletion

If 24-nt siRNAs act as a source of immune memory, then the retention of sTEs may be a benefit to the host, because they may be the template for 24 nt siRNA production. This relationship is implied by our analyses of the deletion (*d*) parameter under the Evade model (figs. 4 and 5). If the goal is simply to rid the genome of TEs, the most efficient method is to have a very high d (> 0.5) that removes all aTEs and sTEs. However, deletions are mediated by ectopic recombination and illegitimate recombination (Devos et al. 2002) that may introduce a substantial fitness cost due to the potential for catastrophic mutations (Langley et al. 1988). Assuming that high ectopic recombination carries an unacceptable fitness cost, our model suggests that the next best solution to limit the number of aTEs is to have very low rates of TE deletion ($d \leq 0.01$).

Our argument is that the retention of sTEs may benefit the host by boosting immune memory. In theory, this immune memory provides a defense against the invasion of new TEs that have sequence homology to existing genomic TEs (Fultz and Slotkin 2017) and also against TEs that have escaped silencing and need to be resilenced. Two interesting features of acquired immune memory are that it is energetically expensive but also maintained under frequent cycles of reinfection (Best and Hoyle 2013). Under the parameter values explored with our model, the system usually reaches a steady state in which the copy number of aTEs is >0. To the extent that these dynamics reflect reality, a nonzero equilibrium of aTEs defines a system in which reinfection is not merely frequent but constant. This observation may explain one feature of the selective pressure to maintain RdDM-like mechanisms, even though it seems as if most TEs within plant genomes are effectively silenced. There is also a conjecture that "zombie" TEs are maintained in the genome in order to produce siRNAs that boost immune memory and can trigger the *trans*-silencing of active relatives (Lisch 2009). Indirect in silico evidence for the existence of zombie TEs has been recently uncovered in maize (Bousios et al. 2016).

Finally, if low rates of TE deletion are somehow beneficial to the host response, this process could drive genome size increases over evolutionary time, because each new TE infection or TE reactivation adds copies that are silenced, retained, and not quickly deleted. We also note that this is unlikely to be a run-away process, because there is evidence for selection on genome size (Diez et al. 2013; Bilinski et al. 2017), especially when genome size gets too large (Knight et al. 2005). Nonethless, our model offers a partial explanation for the high TE contents and sizes of plant genomes.

Future Directions

This is the first study to explicitly incorporate features of the plant host response into a quantitative model of host: TE dynamics. We view this model as a foundation for further extensions that will continue to elucidate important features of host: TE interactions. One promising avenue will be to extend our model to include populations, genetic drift, and fitness (Szitenberg et al. 2016), perhaps with a potential for rare beneficial effects (Le Rouzic, Boutin, et al. 2007). Such an

approach is likely to yield more realistic understandings of the evolution of host: TE interactions than are available at present. It will also be illustrative to model multiple TE families, including autonomous and nonautonomous elements, different length and classes of elements, and the possibility of extensive siRNA cross-homologies. Finally, an important future goal will be to mimic reality by introducing stresses and perturbations into the model. One potential example of a perturbation is polyploidy, which is thought to lead to epigenetic repatterning (Matzke et al. 1999), but for which the causes remain a mystery.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We are grateful for feedback from D. Seymour and D. Wodarz. K.R. is supported by the National Institute of Biomedical Imaging and Bioengineering, National Research Service Award EB009418. A.B. is supported by The Royal Society (Award Numbers UF160222 and RGF\R1\180006). This work was also supported by NSF grant DEB-1655808 to B.S.G.

Literature Cited

Abrusán G, Krambeck HJ. 2006. Competition may determine the diversity of transposable elements. Theor Popul Biol. 70:364–375.

AGI, Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815.

Becker C, et al. 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. Nature 480(7376):245–249.

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. 2008. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PLoS One 3(9):e3156.

Best A, Hoyle A. 2013. The evolution of costly acquired immune memory. Ecol Evol. 3(7):2223–2232.

Bilinski P, et al. 2017. Parallel altitudinal clines reveal adaptive evolution of genome size in zea mays. bioRxiv. Available from: https://www.biorxiv.org/content/early/2017/07/13/134528, last accessed February 28, 2018.

Blumenstiel JP. 2011. Evolutionary dynamics of transposable elements in a small RNA world. Trends Genet. 27(1):23–31.

Bousios A, et al. 2012. The turbulent life of Sirevirus retrotransposons and the evolution of the maize genome: more than ten thousand elements tell the story. Plant J. 69(3):475–488.

Bousios A, et al. 2016. A role for palindromic structures in the cis-region of maize Sirevirus LTRs in transposable element evolution and host epigenetic response. Genome Res. 26(2):226–237.

Bousios A, Gaut BS. 2016. Mechanistic and evolutionary questions about epigenetic conflicts between transposable elements and their plant hosts. Curr Opin Plant Biol. 30:123–133.

Brookfield JF. 2005. The ecology of the genome – mobile DNA elements and their hosts. Nat Rev Genet. 6(2):128–136.

Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. Genet Res. 42(01):1–27.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371(6494):215–220.

Chodavarapu RK, Feng S, Bernatavichute YV, et al. 2010. Relationship between nucleosome positioning and DNA methylation. Nature 466(7304):388–392.

Creasey KM, et al. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature 508(7496):411–415.

Cuerda-Gil D, Slotkin RK. 2016. Non-canonical RNA-directed DNA methylation. Nat Plants 2(11):16163.

Daron J, Glover N, Pingault L, et al. 2014. Organization and evolution of transposable elements along the bread wheat chromosome 3B. Genome Biol. 15(12):546.

Devos KM, Brown JK, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res. 12(7):1075–1079.

Diez CM, et al. 2013. Genome size variation in wild and cultivated maize along altitudinal gradients. New Phytol. 199:264–276.

Fultz D, Choudury SG, Slotkin RK. 2015. Silencing of active transposable elements in plants. Curr Opin Plant Biol. 27:67–76.

Fultz D, Slotkin RK. 2017. Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. Plant Cell 29(2):360–376.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 19(8):1419–1428.

Huang Y, et al. 2015. Ancient origin and recent innovations of RNA polymerase IV and V. Mol Biol Evol. 32(7):1788–1799.

Ibarra CA, Feng X, Schoft VK, et al. 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337(6100):1360–1364.

IBI, International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768.

Ito H, et al. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472(7341):115–119.

Johnson LM, Du J, Hale CJ, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature 507(7490):124–128.

Knight CA, Molinari NA, Petrov DA. 2005. The large genome constraint hypothesis: evolution, ecology and phenotype. Ann Bot 95:171–190.

Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. Genet Res. 52(3):223–235.

Law JA, et al. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498(7454):385–389.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 11(3):204–220.

Le Rouzic A, Boutin TS, Capy P. 2007. Long-term evolution of transposable elements. Proc Natl Acad Sci U S A. 104(49):19375–19380.

Le Rouzic A, Capy P. 2006. Population genetics models of competition between transposable element subfamilies. Genetics 174(2):785–793.

Le Rouzic A, Deceliere G. 2005. Models of the population genetics of transposable elements. Genet Res. 85(3):171–181.

Le Rouzic A, Dupas S, Capy P. 2007. Genome ecosystem and transposable elements species. Gene 390(1–2):214–220.

Li H, Freeling M, Lisch D. 2010. Epigenetic reprogramming during vegetative phase change in maize. Proc Natl Acad Sci U S A. 107(51):22184–22189.

Li S, et al. 2015. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. Genome Res. 25(2):235–245.
Lisch D. 2009. Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol. 60:43–66.
Lisch D, Slotkin RK. 2011. Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts. Int Rev Cell Mol Biol. 292:119–152.
Ma L, et al. 2015. Angiosperms are unique among land plant lineages in the occurrence of key genes in the RNA-directed DNA methylation (RdDM) pathway. Genome Biol Evol. 7(9):2648–2662.
Malthus T. 1798. An essay on the principle of population (ed. 1). London: J. Johnson in St Paul's Church-yard.
Mari-Ordonez A, et al. 2013. Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet. 45(9):1029–1039.
Martinez G, Choudury SG, Slotkin RK. 2017. tRNA-derived small RNAs target transposable element transcripts. Nucleic Acids Res. 45(9):5142–5152.
Martinez G, Köhler C. 2017. Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. Curr Opin Plant Biol. 36:22–28.
Martínez G, Panda K, Köhler C, Slotkin RK. 2016. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nat Plants 2:16030.
Matzke MA, Kanno T, Matzke AJ. 2015. RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. Annu Rev Plant Biol. 66:243–267.
Matzke MA, Scheid OM, Matzke AJ. 1999. Rapid structural and epigenetic changes in polyploid and aneuploid genomes. Bioessays 21(9):761–767.
McCue AD, et al. 2015. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. EMBO J. 34(1):20–35.
Nuthikattu S, et al. 2013. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. Plant Physiol. 162(1):116–131.
Panda K, Slotkin RK. 2013. Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. Plant Signal Behav. 8:8 e25206.
Perelson AS. 2002. Modelling viral and immune system dynamics. Nat Rev Immunol. 2(1):28–36.
Piegu B, Guyot R, Picault N, et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. 16(10):1262–1269.
Schnable PS, Ware D, Fulton RS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115.
Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. 2017. LTR-retrotransposon control by tRNA-derived small RNAs. Cell 170(1):61–71.e11.
Slotkin RK, et al. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136(3):461–472.
Slotkin RK, Freeling M, Lisch D. 2005. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nat Genet. 37(6):641–644.
Szitenberg A, et al. 2016. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. Genome Biol Evol. 8(9):2964–2978.
Teixeira FK, Heredia F, Sarazin A, et al. 2009. A role for RNAi in the selective correction of DNA methylation defects. Science 323(5921):1600–1604.
Tenaillon MI, Hollister JD, Gaut BS. 2010. A triptych of the evolution of plant transposable elements. Trends Plant Sci. 15(8):471–478.
Tian Z, et al. 2009. Do genetic recombination and gene density shape the pattern of DNA elimination in rice LTR-retrotransposons? Genome Res. 19(12):2221–2230.
Tsuzuki M, et al. 2016. Profiling and characterization of small RNAs in the liverwort, *Marchantia polymorpha*, belonging to the first diverged land plants. Plant Cell Physiol. 57(2):359–372.
Volterra V. 1926. Pages 409-448 in Chapman RN 1931. New York: McGraw-Hill.
Wicker T, et al. 2004. A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. Plant J. 41(2):184–194.
Ye R, et al. 2012. Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes. Mol Cell 46(6):859–870.
Zemach A, et al. 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153(1):193–205.
Zhang H, Xia R, Meyers BC, Walbot V. 2015. Evolution, functions, and mysteries of plant ARGONAUTE proteins. Curr Opin Plant Biol. 27:84–90.

Associate editor: Emmanuelle Lerat