Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Alexander Martynov

PhD Program: Life Sciences

Title of Thesis: Using mathematical modeling to understand prokaryotic adaptive immunity

Supervisor: Prof. Konstantin Severinov

Chair of PhD defense Jury: Dr. Mikhail Gelfand

***Email**: alexander.martynov@skolkovotech.ru*

Date of Thesis Defense: 25.10.2018

The thesis document includes the following changes in answer to the external review process.

In response to the review of Dr. Gelfand

1. Firstly, it might be interesting to add an evolutionary process of spacer gain and, even more importantly, elimination. At that, it might be interesting to see whether the observed distribution of cassette sizes is consistent with the predictions of the author's mode, with a simple model of spacer gain at one end and loss of spacers by recombination, or a mixed model. Secondly, the author considers an idealized situation of a single, non-changing cell attacked by same or different viruses. Adding a population aspect also might make the model closer to real life.

Indeed, adding evolutionary and population aspects to the model would be a nice extension and continuation of the given model and might be the focus of the future work. However, this is out of the scope of the current work, as we specifically explored the non-evolutionary approach of studying CRISPR-Cas.

In serponse to the review of Dr. Westra,

In response to the major comments

1. *A more critical reflection and - where possible – explicit test would be needed of how robust the model predictions are to the underlying assumptions. In Chapter 3, I am concerned about the assumption of constant time interval between spacer acquisitions. In nature selection may act to increase the rate of spacer acquisition. This would result in longer arrays with more spacers that are functional (i.e. against which the phage has not yet evolved escape), and is likely to be favoured by natural selection in the face of an evolving phage. I feel this aspect is insufficiently explored and/or discussed in the thesis.*

The most general form of the model presented in the equation 3.20 uses the arbitrary values of viral mutation probabilities (which is directly bound to the time of spacer acquisition) for each individual spacer. The further analysis simplifies it to the constant time between spacer acquisitions. The more accurate approach would be to consider this time to be distributed exponentially between each subsequent spacer acquisition. However, according to the central limit theorem, this would mean that for older spacers this time form acquisition is distributed normally with the corresponding mean. This itself leads to the

situation when the spacer number optimum will also be a distribution with the most probable optimum equal to the obtained from our simplifying model. Thus, our model still shows the statistically most probable optimal number of spacers.
From the perspective of further selection that drives the speed of spacer acquisition we consider this outside the scope of this work, yet, the model could be easily redesigned, using the developed framework in order to capture different distributions of spacer uptake.
Similarly, primed adaptation, which is not considered as part of the model in our work, can affect the rate of spacer acquisition. I now mention this in the revised text to address your concern. However, priming is presently only shown for Type I systems, so the results of our model may be (at least) directly relevant to others systems.

2. *In Chapter 4, experiments show that CRISPR ON cells within a colony are heterogeneous with regards to whether or not they retain the target plasmid (only ± 1 in 4000 cells within a colony carries the plasmid). To me, the reasons for this are not intuitive, and more should be done to clarify how this observation is explained by the model. Is the second equilibrium not stable ?*
We believe that this heterogeneity in the population is induced by the action of the CRISPR-Cas system through the kinetic mechanism described in the text. While not part of the model, spatial organization of the colony and/or depletion of antibiotic in the media around it must also be a major contributor ensuring that a major fraction of cells in colonies formed survive despite losing plasmids. I added several new sections to the discussion of Chapter 4, including one addressing this issue.
3. Primer efficiencies for qPCR are outside of the commonly accepted 0.8-1.2 range. Please speficify / discuss how this may influence the results.
I think this is a term confusion. We calculated primer efficiencies according to https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4839616/ where they distinguish primer efficiency/primer multiplication coefficient, which is an average multiplication rate, and amplification efficiency, which is counted in the percentage of optimal multiplication coefficient. That is, a primer efficiency of 2.0 will correspond to the amplification efficiency of 100% or 1.0. I changed the text accordingly to make this clear.

In response to minor comments:
1. *The introduction chapter covers a lot of ground, but in doing so, becomes somewhat unfocussed and lacks synthesis. I would recommend a more concise introduction that provides a clearer vision of the state of the field and the open questions that this thesis aims to address.*
I tried to separate the Literature Review section, i.e., the introduction to the field in general and summary of the current state of the art and the Introduction section, which is very concise and focuses on the aspects of the research question and motivation for my work and a short summary of results.
2. *Page 14 talks about the mechanism of primed spacer acquisition. Apart from the kinetic explanation (i.e. rapid destruction of spacer substrates), selection will also be important (stronger selection in absence of full resistance) – please adjust accordingly.*
We believe that post-acquisition selection is, at least partially, discussed in the other part of the literature review (see Arms-race), and does not fit in this section.
3. *Page 16; this is supported by the fact that there are conserved spacers …. Narrative assumes that conservation is driven by selection (i.e. that this is adaptive). Not necessarily the case, as shown in various models on conservation of trailer end ; could be due to drift or hitch-hiking effects – please adjust.*
While analyzing the existing literature I concluded that selective sweeps are viewed as the major source of trailer-end effect while other reasons such as drift play only a minor role (see Weinberger 2012, Plos Comp). I adjusted the text to reflect this point.
4. Page 20; Section on Abi feels out of context
I changed the accents in the text so it fits more naturally. Indeed, the comparison of CRISPR-Cas and Abi-like systems is a bit far-stretched altogether, to my taste, but it is something that is discussed in the literature.
9. Page 40 – model assumes fixed expression level of CRISPR-Cas system, whereas this would

appear a strong target of natural selection ; please discuss / explore.
Indeed, though the Cas protein expression rate could be under the pressure of selection on the evolutionary timescale, there are no signs that upon activation this expression could be additionally regulated on an individual-level timescale. I added additional clarifications to avoid confusion.
10. Page 41 - Assumption of constant CRISPR array length at odds with experimental evolution studies ; please discuss.
It is important to note, that we do not assume the array of fixed length or fixed array, but a fixed array on the timescale of virus attacks, which supposed to be not that far from the reality given the generally low efficiency of the naïve adaptation. I added several sentences clarifying this point.
12. In Chapter 3 a clearer discussion is needed about spacer retainment vs acquisition rate (related to major comment 1). This point is only briefly mentioned in the discussion, but should be expanded by textual changes or – if possible – further exploration of the model predictions when these assumptions are relaxed.
While in general spacer acquisition and drop-out rate in inapplicable in the framework of the given model, not much actual exploration could be done. Yet, an additional paragraph in the discussion is made on the role of the diversity of the spacer acquisition and retainment rate.

All other points of the review are fully noted and the text is adjusted accordingly.

In the response to the review of Dr. Malsov,

Here is some response to address your comments in the review
1. Regarding the exponential drop in spacer abundancies. Why is it true? Is it the effect of RNA polymerase falling off during transcription or the effect of a finite number of effectors to bind so that the spacers expressed first will occupy most of them. Deserves a discussion (maybe you have it but I missed it)
We indeed propose that this effect could be achieved through termination of transcription by RNA polymerase and could be mediated by hairpin-like structures of CRISPR repeats. Moreover, a drop in the abundance of trailer-end crRNAs is observed experimentally in many systems (see https://www.ncbi.nlm.nih.gov/pubmed/24256237), though, there is not enough experimental data to determine how general is this phenomenon. We have put this analysis in the section 3.4.3 in the discussion to section 3. Competition for limited effectors is unlikely to create a positional bias, for all crRNAs will compete based on the amounts they are produced.
2. Also, some potentially confusing mathematical notation used in Eqs. (3.7) and (3.9) is indicated.
In eq. 3.7 we have not yet introduced the protospacer mutation in the system, thus "matching" should be viewed as "corresponding", and the summation is indeed over all spacer-protospacer pairs. I made the changes to the text to avoid confusion.
3. *Regarding the interaction between the viral burst and CRISPR-Cas system. Is there any reason to assume this? Wouldn't a phage continue to multiply until it reaches the desired burst size?* Indeed it has been shown that viral burst size is a more or less a fixed value and does not depend on, for instance, the multiplicity of infection (see https://www.sciencedirect.com/science/article/pii/S0042682211005630). However, the likely scenario is that the limiting factor here in the amount of energy that the virus can drain from the individual cell, which effectively limits the number of phage particles produced. Given the fact that viral burst size becomes smaller at conditions of CRISPR-Cas targeting has been shown experimentally (see https://www.ncbi.nlm.nih.gov/pubmed/28130424) we assume that indeed the viral DNA seems to undergo some fixed number of replication cycles before the burst.
4. Chi [interference efficiency] is also an important parameter. Why in following pcolor plots you ignore it and focus on beta-delta interplay?
Regarding the presentation of changes in binding efficiency β and not presenting the results for interference efficiency χ. We have chosen to present only one of these, as changes in both parameters contribute to the CRISPR-Cas performance level and changing either one of them leads to similar results. We have chosen to focus on the binding efficiency as we consider it to be more important for both

cost-benefit of CRISPR-Cas in nature and potential applications of CRISPR-Cas as it is related to off-targeting, a serious practical problem.

All other comments and typos were noted and addressed.

In the response to the review of Dr. Bazykin:

1. The dissertation text was worked through and all found imperfections and typos were fixed where found.
2. Thank you for noting the incorrect and missing links and references, while most of them were automatically processed, the ones that did not go out of sight. Fixed in the final version.
3. *(And, to add insult to injury, the usual notation for the mutation rate, \mu, is the opposite to that used here, 1-\mu.) All this makes the text very hard to comprehend.*

The notation of μ was chosen in favor of the equation simplification and, which, unfortunately, came at cost of slightly less intuitive and, probably non-conventional notation.

4. *On a more scientific note, the key assumption of the first model is that the involvement of a spacer in an effector complex declines exponentially with spacer age. The author motivates this by pointing out that the 5' crRNAs tend to be younger; and that the 5' crRNAs are expected to be generally more abundant than the 3' crRNAs (although they provide no data on the shape of this dependence). Although this is outside the scope of this work, it would be interesting to see if the results change a lot when these assumptions are violated. What would change if the decay of crRNA abundance with distance from the promoter is, in fact, not exponential; e.g., if there is a threshold length, up to which all crRNAs are used equiprobably?*

While leader-end spacers indeed are younger due to a mechanism of spacer integration the exponential-like crRNA distribution is indeed a somewhat simplifying assumption. While there is an observed trend that leader-end crRNA are more abundant than trailer-end (see https://www.ncbi.nlm.nih.gov/pubmed/24256237) there is not enough experimental data by now to catch the actual distribution. I added an additional note to the first place where this assumption is discussed to avoid confusion. If we consider some threshold that limits the number of spacers under which the distribution of crRNA is uniform will correspond to the extreme case scenario of the model when crRNA decay coefficient is 1.

5. *In chapter 4, the author suggests that the surviving plasmids have undergone a period of stochastic expansion. As this expansion is "against odds", probability theory tells us that it would have to be fast. Therefore, if the plasmid is fixed in a cell, it has experienced an unusually rapid period of initial expansion. Can this be tested somehow? (Both above questions go beyond the scope of the thesis, but are interesting if these topics are pursued further.)*

Indeed, our results in the 4th chapter imply that the plasmid replication occurs on significantly higher timescale than the cell division and high copy-number plasmids replicate extremely fast. While this generally is considered to be true, there is no accurate data on plasmid replication kinetics within the cell as it seems to be extremely complex experimentally to capture.

In response to the review of Dr. Mironov:

1. The literature review describes in detail the CRISPR-Cas system and the mechanism of its operation. A separate part of the review is devoted to the mathematical models of the CRISPR-Cas system. This part of the review is directly related to the topic of work, but it is written in a very concise form. I would like this part of the review to be more detailed.

Indeed, the current review of the CRISPR-Cas related modeling is very concise, yet covered most of the available modeling works on CRISPR-Cas. I have extended it a bit for the final version of the text.