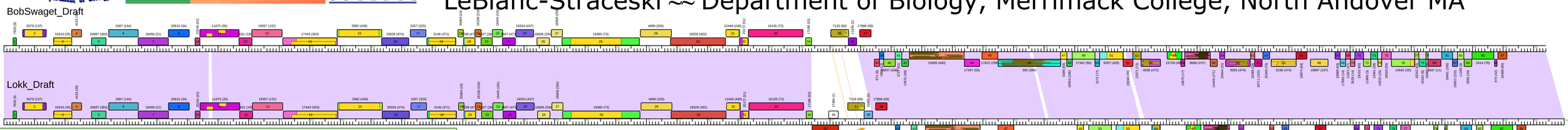
Lokk and BobSwaget: Two Highly Similar Genomes Diverging in the Region Upstream of the parABS Cassette

Alison J. Hanlon, Kathryn A. Wisniewski, Christina M. Bertolez, Janine M. LeBlanc-Straceski ~~ Department of Biology, Merrimack College, North Andover MA



ABSTRACT

Lokk and BobSwaget, both Cluster A2 phage, were isolated from soil samples collected on the same day within 20 meters of each other from the damp, receding banks of the Shawsheen River in Andover MA, after a persistent period of low rainfall. The parABS system of prophage integration and replication was recently identified in ~20% of cluster A2 phage (Dedrick et al. 2016). Lokk and BobSwaget have parA and parB homologues as well as parS-L and parS-R 8 bp repeated sequence motifs immediately 5' and 3' to this gene cluster. Although geographical and temporal proximity does not necessarily correlate with genome sequence homology, Lokk and BobSwaget present a unique circumstance. Their genomes are

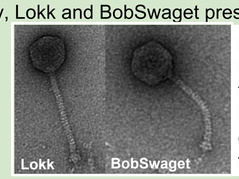


Figure 1. Electron Micrographs of Lokk and BobSwaget. These are the EM images of Lokk (left) and BobSwaget (right). Both are classic examples of siphoviridae (long flexible, noncontractile tail).

98% identical, with a few base changes scattered throughout. The biggest difference in their genomes, however, is in a ~1200 bp region immediately 5' to the parABS cassette. This corresponds to the region identified as encoding a non-coding RNA (Dedrick et al. 2016). This region of the genome is highly divergent in all of the A2 cluster phage with the parABS mechanism of prophage genome integration. It is interesting to note that this region of sequence divergence was the only difference found in two phages that were collected simultaneously, under stressful environmental conditions - conditions that could favor integration into the host genome. Indeed, during this period of drought very few soil samples yielded any phage at all. We speculate that Lokk and BobSwaget might infect similar hosts, or even the same host, perhaps simultaneously, and that this divergent region of the A2 phage genome might specify the integration site on the host genome. Alternatively this region may direct some other mechanism that allows similar phage to occupy different niches during stressful environmental conditions.

Figure 2. Collection Site. At left is the google maps of the collection location and the right is an image of the exact location. GPS coordinates : Lokk: 42.6689 N 71.1463 W. BobSwaget: 42.6694 N 71.1477 W.

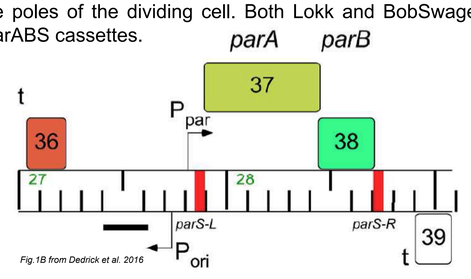
	2015	Observed	Average
Precipitation	July	1.5-2.0 inches	3.0-4.0 inches
	August	2.0-3.0 inches	3.0-4.0 inches

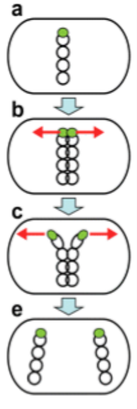
	2015	Observed	Average
Temperature	July	80.5F	78.9F
	August	70.7F	68.8F

Table 1. Weather Conditions Prior to Collection Date. Source: National Weather Service Archive.

INTRODUCTION

Aa an alternative to lysogeny by integration, the parABS system ensures that at each division, both host cells get and keep at least one extra-chromosomal copy of the phage genome. The parABS cassette is located in the genome in relatively the same position as the integration cassette would be. Figure 1B from Dedrick et al. 2016 (below) shows the organization of the parABS cassette in RedRock. Multiple eight base pair repeats (parS), located directly upstream and downstream of the parA and parB genes, can be recognized by the ParB protein. In the diagram at right, the green dots represent the DNA binding protein ParB and red arrows filaments of ParA, an ATPase that shuttles the small circular DNA to opposite poles of the dividing cell. Both Lokk and BobSwaget possess parABS cassettes.





Sengupta et al. 2010.

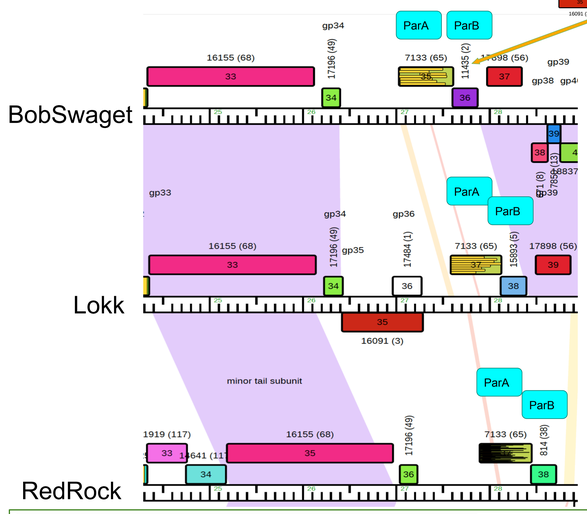
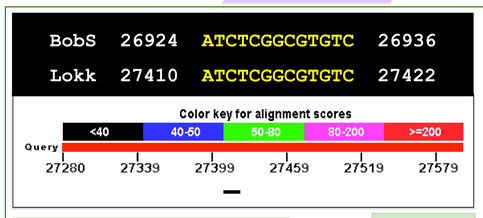


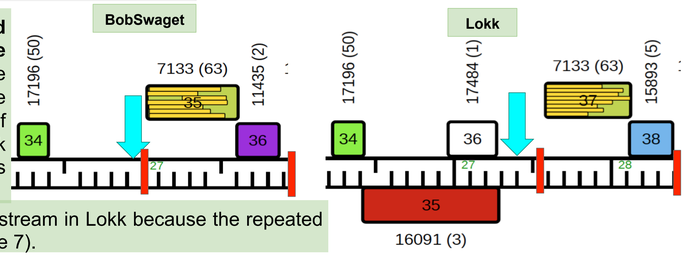
Figure 3. Region of Sequence Divergence Upstream and Within the parABS Cassette. The figure to the left is a phamerator comparison of BobSwaget, Lokk and Red Rock. Notice that ParA for all three phage are in the same protein family. However, each parB from a different pham, reflecting the increased level of diversity in parB.



A conserved 13 bp motif upstream of parS-L was detected in several parABS cassette containing phage including RedRock, Changeling, First, Packman, Myxus, Eidsmore, Catalina, Alma, Luchadore, Echild and CRB1.

Figure 4. A Conserved Sequence Motif within the Region of Divergence. The blue arrows indicate the position of the conserved motif for BobSwaget (left) and Lokk (right). Vertical red bars represent the parS repeats.

The motif is slightly further upstream in Lokk because the repeated sequence is longer (see Figure 7).



ParA Comparisons

Lokk x BobSwaget: 62% Identical, 75% Homologous, 1 Gap

Lokk x RedRock: 56% Identical, 75% Homologous, 7 Gaps

BobSwaget x RedRock: 58% Identical, 71% Homologous, 8 Gaps

Needleman-Wunsch alignment of two sequences

Figure 5. ParA Comparisons. ParA amino acid sequences show a lot of similarities. The orange letters show where the amino acids are the same. The orange + shows where they are similar, and where there is no orange they are not similar at all. Lokk and BobSwaget are 62% identical, Lokk and RedRock are 56% identical and BobSwaget and RedRock are 58% identical.

ParB Comparisons

Lokk x BobSwaget: 48% Identical, 64% Homologous, 2 Gaps

Lokk x RedRock: 33% Identical, 45% Homologous, 4 Gaps

BobSwaget x RedRock: 39% Identical, 50% Homologous, 2 Gaps

Needleman-Wunsch alignment of two sequences

Figure 6. ParB Comparisons. ParB amino acid sequences do not show as many similarities as in the ParA comparison. Lokk and BobSwaget are 48% identical, Lokk and RedRock are 33% identical, and BobSwaget and RedRock are 39% identical.

Acknowledgements

- Phage Hunters of Merrimack College BIO1027 FA2015
- Charlotte Berkes, Mark Birnbaum, Thuy Tran
- Department of Biology, Merrimack College
- Garry Cuneo, Wyss Institute, Boston MA
- Dan Russell + the sequencing team
- Welkin Pope, Debbie Jacobs-Sera

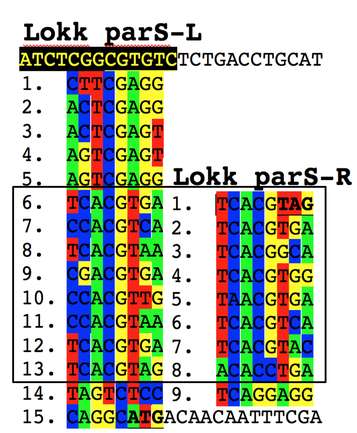
References

Dedrick, Rebekah M., Travis N. Mavrich, Wei L. Ng, Juan C. Cervantes Reyes, Matthew R. Olm, Rachael E. Rush, Deborah Jacobs-Sera, Daniel A. Russell, and Graham F. Hatfull, 2016. Function, expression, specificity, diversity, and incompatibility of actinobacteriophage parABS systems. *Molecular Microbiology* 'Accepted Article', doi: 10.1111/mmi.13414.

Sengupta, Manjistha, Henrik Jorck Nielsen, Brenda Youngren, and Stuart Austin. 2010. P1 Plasmid Segregation: Accurate Redistribution by Dynamic Plasmid Pairing and Separation. *J. Bacteriology*. 192,1175-1183.

Lindås, Ann-Christin and Rolf Bernander 2013. The cell cycle of archaea. *Nature Reviews Microbiology*. 11, 627–638..

National Weather Service



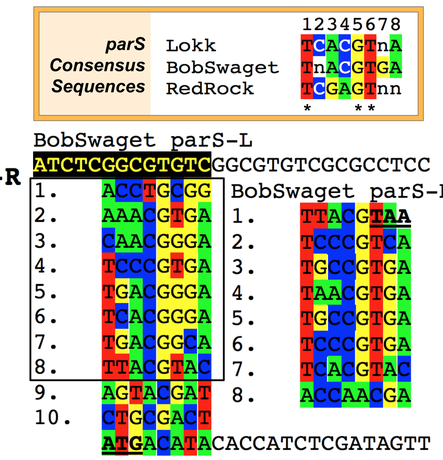
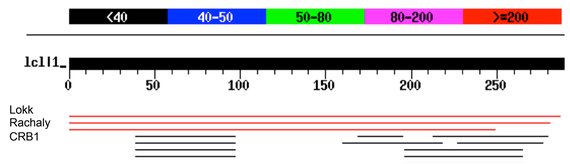


Figure 7. parS-L and parS-R Repeats and Consensus Sequences. The consensus sequence in Lokk (top left) BobSwaget (top right) flanking the parA and parB genes. In Lokk there are an additional 5 repeats in the parS-L with a different consensus. The comparison of the consensus sequences of Lokk, BobSwaget, and RedRock (left). In the comparison the " * " represent a match in all three phage, while the "n" placeholder indicates that there was no consensus in that particular position.



Amino Acid Sequence Homology in the parABS genes

Lokk x	ParA	ParB
Rachaly	88%	48%
CRB1	88%	67%

Figure 8. Is Lokk gp35 a rep protein? Only three phage have this protein: Lokk, Rachaly (from WPI 2015) and CRB1 (Argentina 2014) All three have parABS cassettes. Lock gp35 has 82% homology to a putative rep (replication) protein from mycobacteriophage CRB1, and rep proteins from many species of *Mycobacterium* including 54% homology to *Mycobacterium cosmeticum*. Rep proteins function in rolling circle replication. Is this a functional protein in these three phage? Does it have a promoter in the correct direction? Closest reverse promoter predicted by DNA Master was (-35) at bp 27559 and (-10) at bp -27536, which is 247 bp away from the start codon. The parA proteins have a higher sequence homology than between Lokk and BobSwaget. The top two BLASTp for Lokk hits from the phages.db archive for parA are CRB1 and Rachaly, both at 88% homologous. ParB is much less similar 67 and 48% homologous.

CONCLUSIONS

- BobSwaget and Lokk are A2 *Mycobacterium smegmatis* phage with 98% sequence identity. They were collected at the same site and time after a two month period of low precipitation.
- Their genomes diverge in the region upstream and within a parABS cassette.
- Their ParA proteins are 75% homologous and ParB proteins are 64% homologous.
- Both contain repeated parS consensus sequences that flank the parA/parB cluster.
- These 8 copies of an 8 bp consensus sequence are similar to the parS from RedRock (Dedrick et al. 2016).
- Lokk has an additional 5 copies of a different 8 bp sequence motif at the 5' end of the parS-L.
- Another conserved 13 bp sequence was identified 5' to the parS-L. This was found in several parABS containing phage from both A2 and A9 clusters.
- Gp35 from Lokk has a 54% homology to a rep protein from several species of *Mycobacterium* and is also found in two other parABS phage, Rachaly and CRB1.
- It seems unlikely that this is a functional rep protein because
  - It would be transcribed in the opposite direction from the genes that flank it.
  - There was no promoter identified close to the 5' end of the gene.
  - There are only three phage that carry this gene.
- Further analysis of Lokk and BobSwaget may reveal information on how phage evolve when competing for hosts under stressful environmental conditions.