BINF-6999 Research Proposal:

Conserved domain-based annotation of site-specific tyrosine recombinases of mobile genetic elements and data mining for the discovery of Recombinase-in-Trio (RIT) elements.

2021/06/04

Student Name: Jason Moggridge

Student Number: 1159229

University Advisor, Biological Expertise: Dr. Nicole Ricker, Department of Pathobiology

University Advisor, Informatics Expertise: Dr. Dan Tulpan, Department of Animal Biosciences

Mobile genetic elements (MGE) are DNA sequences capable of intracellular and/or intercellular transfer, as well as that of any 'cargo genes' that they encapsulate, which can include resistance genes, virulence factors and other adaptive traits [1]. The mobility of MGEs such as temperate phages, integrative and conjugative elements, transposons, and others can be enabled by the DNA breaking-rejoining activity of various site-specific tyrosine recombinases (YR; also referred to as integrases). YRs are a diverse and widespread family of proteins that are also involved in essential cellular functions including replication and differentiation [2]. MGEs may either exploit the YRs of their host or encode one or more YRs to facilitate their mobility. Recombinase-in-trio (RIT) elements are a recently described class of MGE distributed across diverse bacterial taxa [3,4]. RIT elements encode a triplet of YRs, often with overlapping reading frames. While the function of the three YRs has not yet been determined, it seems likely that they are involved in RIT element mobilization inferred from duplication within strains.

It has been previously demonstrated that YR domains can be clustered into subfamilies, the majority of which are associated with specific classes of MGEs [2,5]. However, YRs are currently labelled generically as 'phage integrases' by functional annotation tools and in public databases. As such, classification of YRs in a more precise ontology could provide valuable information about the presence of specific types of MGEs (and the relationships between them). The proposed work has two related objectives: (1) to classify YRs in known MGEs from reference databases into 20 subfamilies described by Smyshlyaev *et al.* [5] aimed at improving their functional annotation and characterization of their distribution in MGEs; and (2) to discover and characterize new RIT elements from public sequence databases using the pipeline from objective 1, and then to describe their distribution within genomes and across taxa.

To accomplish the first objective, I will create a pipeline to assign YRs of MGEs to subfamilies based on the alignment of conserved domains. For this, I will use sets of reference proteins from the SMART database to develop and evaluate the accuracy of candidate methods for classification (such as through Hidden Markov Models or NCBI's CD-search/RPS-BLAST). After classification, the distribution of YR subfamilies among specific MGEs and across taxa will be characterized. To complete the second objective, the Short Read Archive (SRA) will be searched for sequences encoding the three YR domains (RitA, RitB, and RitC) associated with RIT elements, using methods previously developed for objective 1. Novel RIT elements will be characterized through various sequence analyses in terms of their protein architectures, ORF overlaps, the presence of flanking inverted repeats, and their distribution within genomes and across taxa.

The proposed work will provide a critical update of the functional annotation of YRs in MGEs, which could provide new insights into the mobilization of MGEs and relationships between them. It is of interest to improve our knowledge of MGEs since these are critical to microbial adaptation and evolution, particularly in the horizontal transfer of antibiotic resistance, virulence factors, and other accessory traits. As antibiotic resistance is a growing problem due to the selective pressure of widespread antibiotic use, understanding how resistance is transferred and maintained in the microbial environment is paramount for surveillance of AMR pathogens and developing new strategies for antimicrobial therapies.

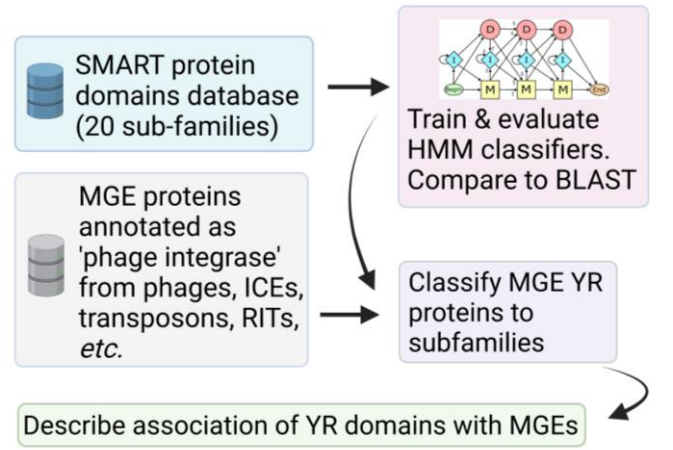

Figure 1. Workflow for objective 1 – classification of tyrosine recombinases from mobile genetic elements.

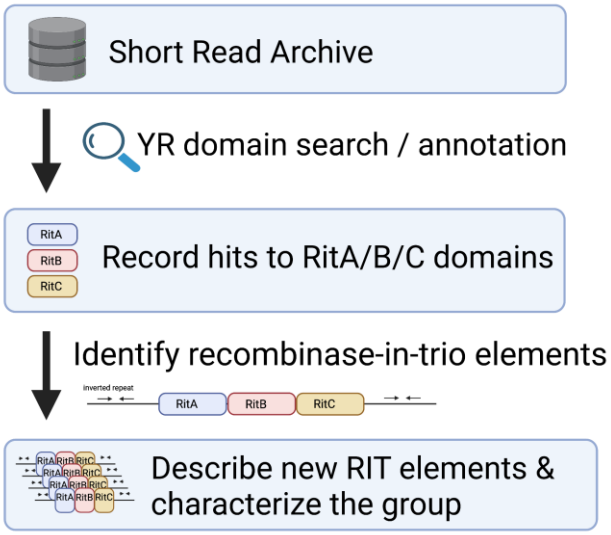

Figure 2. Workflow for objective 2 – discovery of RIT elements from SRA

References

1. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev. 2018 Aug 1;31(4):e00088-17. doi: 10.1128/CMR.00088-17. PMID: 30068738; PMCID: PMC6148190.
2. Van Houdt R, Leplae R, Lima-Mendez G, Mergeay M, Toussaint A. Towards a more accurate annotation of tyrosine-based site-specific recombinases in bacterial genomes. Mob DNA. 2012 Apr 13;3(1):6. doi: 10.1186/1759-8753-3-6. PMID: 22502997; PMCID: PMC3414803.

3. Janssen PJ, Van Houdt R, Moors H, *et al*. The complete genome sequence of Cupriavidus metallidurans strain CH34, a master survivalist in harsh and anthropogenic environments. PLoS One. 2010 May 5;5(5):e10433. doi: 10.1371/journal.pone.0010433. PMID: 20463976; PMCID: PMC2864759.
4. Ricker N, Qian H, Fulthorpe RR. Phylogeny and organization of recombinase in trio (RIT) elements. Plasmid. 2013 Sep;70(2):226-39. doi: 10.1016/j.plasmid.2013.04.003. Epub 2013 Apr 28. PMID: 23628708.
5. Smyshlyaev G, Bateman A, Barabas O. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. Mol Syst Biol. 2021 May;17(5):e9880. doi: 10.15252/msb.20209880. PMID: 34018328.

Timeline

Week	Tasks/Milestones	Completed?
1 (05/13 - 05/16)	**Scheduling** of weekly meeting **Literature review**	Yes Yes / Ongoing
2 (05/17 - 05/23)	**Literature review** **Proposal:** outline/draft **Research preliminary**: - Conceptualize pipelines - Data exploration: MGE and protein ref. databases - Data acquisition: references & target	Yes / Ongoing Outline started Discussed Started Started
3 (05/24 - 05/30)	**Proposal:** to advisors for review (~26th) **Research obj 1**: - Obtain ref data for MGE proteins (filter YRs) - Plan methods to identify RITs from YRs - Planning how to identify domains in MGE YRs (classifier using BLAST or HMM)	- Drafted - Started data acquisition and methods planning
4 (05/31 - 06/06)	***Edit & submit proposal (06/04 4pm)*** **Research obj 1**: - Create + evaluate tools/models to classify MGE YRs - Build alignments, hmms for SMART proteins. - Finish obtaining targets YR seqs from known MGE databases. - Train / evaluate models. - Classify target YRs / Preliminary results for obj 1.	- Submitted - Started hmm building workflow
5 (06/07 - 06/13)	**Research**: - Finish classification pipeline for objective 1 - Begin developing pipeline to identify RIT elements (obj 2) - hmm scan	

Week	Tasks/Milestones	Completed?
6 (06/14 - 06/20)	**Research**: - Ensure that obj 1 pipeline is reproducible. - Summarize results & create figures. - Review results with advisors. - Continue developing RIT pipeline	
7 (06/21 - 06/27)	**Research**: - Apply RIT pipeline & collect results from SRA & other sequence archives - Summarize initial results for review by advisors	
8 (06/28 - 07/04)	**Draft self-reflection essay** **Research**: - Apply RIT pipeline, finish collecting results - Summarize RIT discoveries for review by advisors	
9 (07/05 - 07/11)	Proof & submit: ***Self-reflection essay (07/09 4pm)*** **Research**: - Describe / characterize RIT elements – further analysis, *e.g.* alignment, clustering, taxonomic distribution, *etc.*	
10 (07/12 - 07/18)	Finalize the research work (changes or further analyses). **Write report**: - Methods/Results/Figures	
11 (07/19 - 07/25)	**Report outline –** discuss with advisors Finalize research work & Ensure reproducibility **Write report**: Discussion & Introduction *Deadline: apply for fourth semester or to inform coordinators of intent to transfer to MSc (June 25)*	
12 (07/26 - 08/01)	**Presentation**: prepare slides and writing **Write report**: Introduction / Discussion / Abstract	
13 (08/02 - 08/08)	Practice presentation with advisors & peers for feedback **Write report** Send draft report to advisors ***Oral presentations (08/05-06)***	
14 (08/09 - 08/15)	Edit report ***Final report (08/13)***	

Meetings

I will have arranged for weekly meetings with Drs. Ricker and Tulpan on Wednesday afternoons at 1:30pm for the duration of the project when possible (alternative means of communication will be used when schedules do not allow for an online in-person meeting).