Teaching transposon classification as a means to crowd source the curation of repeat annotation – a tardigrade perspective

Valentina Peona[1], Jacopo Martelossi[2], Transposable Element Course Participants , and Alexander Suh*[3]

[1]Swiss Ornithological Institute – 6204 Sempach, Switzerland
[2]University of Bologna – Piazza di Porta S.Donato 1, Bologna, Italy
[3]Leibniz Institute for the Analysis of Biodiversity Change – Adenauerallee 160, D-53113 Bonn, Germany

Abstract

The advancement of the third-generation sequencing era is providing an unprecedented opportunity for the global genomic community. Because of the increasing cost-effectiveness of long-read sequencing, an exponential number of high-quality genomes are being released rapidly for non-model organisms, providing valuable resources for conservation and evolutionary studies. Within this context, the significance of in-depth analyses of repetitive elements, particularly transposable elements (TEs), is increasingly recognized in understanding genome evolution. However, while a plethora of bioinformatic tools is available for identifying and annotating TEs, any automated annotation effort is constrained by the phylogenetic distance of the target species from a curated and classified database of repetitive element sequences. Furthermore, manual curation of raw repeat libraries is deemed essential due to the frequent truncation and incomplete status of automatically generated consensus sequences. Nonetheless, manual curation and classification are time-consuming processes, they offer limited short-term academic rewards, and are typically confined to a few research groups where they are taught through hands-on experience. Crowdsourcing efforts could offer a significant opportunity to bridge this gap and empower the scientific community with high-quality, reusable genomic resources. Here, we present an example of such a process with both in-person and online courses, each attended by one or two dozen participants. These courses focused on tardigrades, a phylum for which no TE libraries existed, and nearly all repeats were previously annotated as "Unknown". After a series of theoretical lectures on TEs biology and practical examples of manual curation, participants were divided into groups curating hundreds of consensus sequences each. After a peer-review process we discovered a huge diversity especially of non-autonomous TEs, demonstrating that a peer-reviewed classroom setting can yield substantial benefits for both students and for the entire scientific community. A hidden treasure awaits discovery within non-model organisms.

Keywords: manual curation, genome evolution, crowd sourcing

---

*Speaker