**Project Title:** Understanding Gene Prediction Programs

**Authors and Affiliations:**

|  |  |
| --- | --- |
| **Author** | **Affiliation** |
| **Ellen Wisner** | **University of North Carolina at Charlotte** |
| **Beckie Bortz** | **University of Pittsburgh** |
| **Kirk Anders** | **Gonzaga University** |
| **Bill Davis** | **Washington State University** |
| **Lee Hughes** | **University of North Texas** |

**Keywords:** gene prediction, genemark, coding potential, codon bias, six frame translation

**Description:** The field of modern genomics is built upon a foundation of big data and the use of computational resources. The teaching materials in this packet are aimed at helping students apply the key biological concepts underlying gene prediction, understand how gene prediction programs work, and how to analyze and apply the output of gene prediction/analysis programs. We also include an activity that allows students to reflect on their engagement with big data and computational output.

**Learning objectives:**

After completing this module, students should be able to:

1. Identify open reading frames in a 6-frame translation, in both directions.
2. Create a 6-frame translation from a DNA sequence and identify potential starts and stop coordinates.
3. Explain the relationship between an ORF and a gene.
4. Describe how coding potential helps to identify putative genes.
5. Interpret GeneMark output.
6. Recognize the role of genomics (Big Data) in biological research.

**How is the module structured to promote student development as a scientist?**

This module promotes students’ development as scientists by encouraging independence, facilitating peer collaboration and through explicit discussion of how the SEA-PHAGES program contributes to big data. The activities have students work with each other to interpret real data output on their phage. In addition, we have included two readings and discussion prompts that will allow students to reflect upon big data, and how their research within the SEA-PHAGES program relates and contributes to the field.

**Intended Teaching Setting**

**Course level:** This resource is adaptable for both major and non-major students

**Instructional Setting:** The module is designed for in-class teaching, but has suggestions for modification for use in online teaching.

**Implementation Time Frame:** This would likely be implemented over a few class days. The entire module would likely take between 2.5 - 3.5 hours to complete.

**Acknowledgments:** HHMI Science Education Alliance, Denise Monti

**Project Documents**

**Facilitator document:** Teaching Notes - Gene Prediction Programs

**Learning activity documents:**

Slides for Gene Prediction Programs.pptx

Creating and Reading a 6-Frame Translation.docx

Find ORFs in a 6-Frame Translation.docx

Article discussion prompts.docx

**Assessment documents:**

Creating and Reading a 6-Frame Translation KEY.docx

Find ORFs in a 6-Frame Translation KEY.docx

Question bank for assessment

Question bank for assessment - KEY