



**Universidad Nacional de San
Agustín de Arequipa**

**Escuela Profesional de Ciencia de la
Computación**

**Computación Molecular Biológica
(Código: 1005155)**

Semestre 2020A



Indice

1	Curriculum Vitae en Formato ICACIT	3
2	Sílabo del Curso en Formato DUFA	6
3	Sílabo del Curso en Formato ICACIT	12
4	Prueba de Entrada	15
4.1	Evidencias	18
5	Evaluación Primer Parcial	19
5.1	Evidencias	22
6	Evaluación Continua 1	23
6.1	Evidencias	32



1. Curriculum Vitae en Formato ICACIT



UNIVERSIDAD NACIONAL DE SAN AGUSTÍN
Facultad de INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
Escuela Profesional de Ingeniería DE SISTEMAS

Curriculum Vitae del Docente

Nombre	VICENTE ENRIQUE MACHACA ARCEDA
Educación	<ul style="list-style-type: none"> · Magíster en ciencias informática con mención en tecnologías de información, 2016 / Universidad Nacional de San Agustín · Ingeniero de sistemas, 2017 / Universidad Nacional de San Agustín · Bachiller en Ingeniería de sistemas, 2011 / Universidad Nacional de San Agustín
Experiencia Académica	<ul style="list-style-type: none"> · Cátedra de Postgrado; Universidad Universidad Nacional de San Agustín; 2017, 2018 · Cátedra de Pregrado; Universidad Universidad Nacional de San Agustín; 2017, 2018 · Cátedra de Pregrado; Universidad La Salle; 2017, 2018
Experiencia No Académica	<ul style="list-style-type: none"> · Vex Soluciones E.I.R.L.; Analista; 2017 - 2017. · Tata Consultancy Services – Tcs; System Engennier; 2013 - 2014. · Superintendencia Nacional de Aduanas y Administración Tributaria (SUNAT); IPM: 2011 – 2014. · Coriing Eirl; Analista; 2011 – 2013 · Ctd V&C Sac; Analista; 2010 – 2011 · Regesa Scrl; Programador; 2009 – 2010
Registro Profesional	Colegio de Ingenieros del Perú, CIP: 211444
Membresía actual en organizaciones profesionales	· Colegio de Ingenieros del Perú, 2018. CIP: 211444

Honores y Premios	<ul style="list-style-type: none"> · Acreedor a una beca integral para estudiar una maestría en ciencias informática. · Alumno revelación de la maestría en ciencias informática
Actividades de Servicio (dentro y fuera de la Institución)	
Publicaciones y Presentaciones (últimos 5 años)	<p>Publicaciones – Artículos</p> <ul style="list-style-type: none"> · Small Ship Detection on Optical Satellite Imagery with YOLO and YOLT. Presentado en: FTC 2020 - Future of Information and Communication Conference San Francisco, EEUU, 2020. · Fast Car Crash Detection. Presentado en: CLEI 2018 - The Latin American Computing Conference, São Paulo, Brasil, 2018. · Fast Face Detection in Violent Video Scenes. Publicado en: ScienceDirect. Presentado en: CLEI 2016 - The Latin American Computing Conference, Valparaiso, Chile, 2016. · Real Time Violence Detection in Video with ViF and Horn-Schunck. Publicado en: LACCEI. Presentado en: The Latin American and Caribbean Consortium of Engineering Institutions, San Jose, Costa Rica, 2016. · Optimization model for face detection in video sequences. Publicado en: LACCEI. Presentado en: The Latin American and Caribbean Consortium of Engineering Institutions, San Jose, Costa Rica, 2016. · Real Time Violence Detection in Video. Publicado en: IEEE Explore y IET Digital Library. Presentado en: International Conference on Pattern Recognition Systems, Talca, Chile, 2015.
Actividades de Desarrollo Profesional (últimos 3 años)	<ul style="list-style-type: none"> · Docente con experiencia en los niveles de pregrado y postgrado en diversas la Universidad Nacional de San Agustín y la Universidad La Salle. · Ponencias en Conferencias Internacionales LACCEI 2018, CLEI 2016.



2. Sílabo del Curso en Formato DUFA

UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA



VICERRECTORADO ACADÉMICO

**FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
DEPARTAMENTO ACADÉMICO DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**

SILABO 2021 A

ASIGNATURA: COMPUTACIÓN MOLECULAR BIOLÓGICA

1. INFORMACIÓN ACADÉMICA

Periodo académico:	2021 - A	
Escuela Profesional:	Ciencias de la Computación	
Código de la asignatura	1005155	
Nombre de la asignatura	Computación molecular biológica	
Semestre:	VII	
Características:	Semestral	
Duración:	17 semanas	
Número de horas (Semestral)	Teóricas	2
	Teórico-práctico	2
	Prácticas	2
	Laboratorio	-
Número de Créditos:	6	
Prerrequisitos:	1003233	Estructuras de Datos Avanzadas

2. INFORMACIÓN ADMINISTRATIVA

PROFESOR: Vicente Machaca Arceda					
GRADO ACADÉMICO: Maestro en Ciencias Informática, con mención en Tecnologías de Información					
DEPARTAMENTO ACADÉMICO: Ingeniería de Sistemas e Informática					
HORARIO	Lunes	Martes	Miércoles	Jueves	Viernes
Total Semanal: Hrs.		14:00 a 15:40	10:40 a 12:20	14:00 a 15:40	
Tipo de clase		Teoría	Práctica	Laboratorio	
Grupo		A	A	A	

3. FUNDAMENTACIÓN (JUSTIFICACIÓN)

El uso de métodos computacionales en las ciencias biológicas se ha convertido en una de las herramientas claves para el campo de la biología molecular, siendo parte fundamental en las investigaciones de esta área. En Biología Molecular, existen diversas aplicaciones que involucran tanto al ADN, al análisis de proteínas o al secuenciamiento del genoma humano, que dependen de métodos computacionales. Muchos de estos problemas son realmente complejos y tratan con grandes conjuntos de datos. Este curso puede ser aprovechado para ver casos de uso concretos de varias áreas de conocimiento de Ciencia de la Computación como: Lenguajes de Programación (PL), Algoritmos y Complejidad (AL), Probabilidades y Estadística, Manejo de Información (IM), Sistemas Inteligentes (IS).

4. SUMILLA DEL CURSO

El curso tiene como objetivo que el alumno tenga un conocimiento sólido de los problemas biológicos moleculares que desafía la computación y que el alumno sea capaz de abstraer la esencia de los diversos problemas biológicos para plantear soluciones usando sus conocimientos de Ciencia de la Computación.

5. COMPETENCIAS

- La comprensión intelectual y la capacidad de aplicar las bases matemáticas y la teoría de la informática (Resultado [a] nivel 2).
- Analiza, diseña y propone soluciones frente a problemas bioinformáticos. (Resultado [b] nivel 1, Resultado [c] nivel 1, Resultado [d] nivel 1).
- Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas. (Resultado [a] nivel 2 y Resultado [h] nivel 2).

6. CONTENIDOS DE LA ASIGNATURA

PRIMERA UNIDAD: Introducción a la Biología Molecular

Capítulo I.- Biología molecular

1. Introducción
2. Biología de la célula
3. Bases de datos en bioinformática
4. Transcripción y traducción
5. Estructura del DNA y replicación de DNA
6. Secuenciamiento de DNA

SEGUNDA UNIDAD : Algoritmos en Bioinformática

Capítulo II.- Alineamiento de Secuencias

1. Alineamiento de secuencias de DNA y aminoácidos.
2. Dot matrix
3. Programación dinámica
4. BLAST

Capítulo III.- Árboles Filogenéticos

1. Introducción y relaciones filogenéticas
2. UPGMA
3. Neighbor joining
4. Métodos basados en caracteres

Capítulo IV.- Ensamblaje de Secuencias

1. El problema np-hard de ensamblaje de secuencias de ADN
2. Técnicas basadas en grafos y k-mer
3. Técnicas basadas en heurísticas

TERCERA UNIDAD : Tópicos en Bioinformática

Capítulo V.- Tópicos en Bioinformática

1. Métodos de clasificación *alignment-free*
2. Predicción de estructura de proteínas y función biológica de proteínas
3. *Drug discovery*
4. *Cancer genomics*

7. ESTRATEGIAS DE ENSEÑANZA

- a) **Métodos:** Expositivo en clases teóricas y desarrollo de un trabajo práctico.
- b) **Medios:** Classroom, Google meet.
- c) **Formas de organización:** Clases teóricas, exposición de clases magistrales.
- d) **Programación de actividades que integren investigación formativa y responsabilidad social:** Difusión del Pensamiento Computacional en la Región Arequipa.
- e) **Seguimiento del aprendizaje:** Desarrollo de los trabajos en el taller.

8. CRONOGRAMA ACADÉMICO

Semana	Tema / Evaluación	Docente	Avance
1	Tema 1	Vicente Machaca Arceda	6%
2	Tema 2	Vicente Machaca Arceda	12%
3	Tema 3	Vicente Machaca Arceda	20%
4	Tema 4	Vicente Machaca Arceda	25%
5	Examen parcial 1	Vicente Machaca Arceda	35%
6	Tema 5	Vicente Machaca Arceda	40%
7	Tema 6	Vicente Machaca Arceda	46%
8	Tema 7	Vicente Machaca Arceda	52%
9	Tema 8	Vicente Machaca Arceda	58%
10	Tema 9	Vicente Machaca Arceda	60%
11	Examen parcial 2	Vicente Machaca Arceda	65%
12	Tema 10	Vicente Machaca Arceda	71%
13	Tema 11	Vicente Machaca Arceda	80%
14	Tema 12	Vicente Machaca Arceda	86%
15	Tema 13	Vicente Machaca Arceda	92%
16	Tema 14	Vicente Machaca Arceda	95%
17	Examen final	Vicente Machaca Arceda	100%

9. ESTRATEGIAS DE EVALUACIÓN

Evaluación Continua.

Práctica y Laboratorios en cada clase sobre los temas realizados, tanto para el primer parcial (EC1), segundo parcial (EC2) y tercer parcial (EC3).

Evaluación Periódica.

Al ser un curso basado en lenguajes de programación, la evaluación periódica representa el promedio de las prácticas desarrolladas en clase.

Cronograma

Evaluación	Fecha de Evaluación	Examen Teoría	Evaluación Continua	Ponderación porcentual
Evaluación parcial	13/05/2021	15%	15%	30%
Evaluación parcial	17/06/2021	15%	15%	30%
Evaluación final	29/07/2021	20%	20%	40%

Tipo de evaluación

Heteroevaluación, es la evaluación que realiza una persona sobre otra respecto de su trabajo, rendimiento.

Instrumentos de evaluación

Exámenes, proyecto de curso, prácticas y exposiciones.

10. REQUISITOS DE APROBACIÓN

Para aprobar el curso se deberá haber presentado todos sus trabajos. Los trabajos o tareas deberán ser originales, la copia o plagio a cualquier tipo de nivel, o cualquier tipo de actitud deshonestas, será castigado con cero en todo el componente donde se haya detectado la copia.

11. BIBLIOGRAFÍA

Bibliografía básica obligatoria

[1] Aluru, S., editor (2006). Handbook of Computational Molecular Biology. Computer and Information Science Series. Chapman & Hall, CRC, Boca Raton, FL.

[2] Archibald, John M. *Genomics: A Very Short Introduction*. Vol. 559. Oxford University Press, 2018.

[3] Xiong, Jin. *Essential bioinformatics*. Cambridge University Press, 2006.

Bibliografía de consulta

[4] Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G. RNA-seq data analysis: a practical approach. CRC press; 2014 Sep 19.

[5] Kim IJ, editor. Cancer Genetics and Genomics for Personalized Medicine. CRC Press; 2017 Apr 11.



Vicente Machaca Arceda



3. Sílabo del Curso en Formato ICACIT



UNSA
UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA

UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS

Sílabos del Curso

ESCUELA PROFESIONAL DE CIENCIA DE LA COMPUTACIÓN

1. Nombre del curso:

Código	Nombre	Semestre
1005155	Bioinformática	2021-A

2. Créditos y horas semanales:

Nº créditos	H. Teoría	H. Práctica	H. T-P	H. Lab	T. Horas
6	2	2	2		6

3. Nombre del instructor o coordinador del curso:

MSc. Vicente Machaca Arceda

4. Libro texto: Título, autor y año:

a. Obligatoria

Título	Autor	Año
Essential bioinformatics	Essential bioinformatics	2006

b. Otros materiales suplementarios

Título	Autor	Año
A Very Short Introduction	Archibald, John M	2018



UNSA
UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA

UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS

5. Información específica del curso:

a. Breve descripción del contenido del curso:

El curso tiene como objetivo que el alumno tenga un conocimiento sólido de los problemas biológicos moleculares que desafía la computación y que el alumno sea capaz de abstraer la esencia de los diversos problemas biológicos para plantear soluciones usando sus conocimientos de Ciencia de la Computación.

b. Requisitos previos o correquisitos:

1703238 - Estructuras de datos avanzadas

c. Obligatorio o Electivo:

Obligatorio		Electivo	X
-------------	--	----------	---

6. Objetivos específicos del curso:

La comprensión intelectual y la capacidad de aplicar las bases matemáticas y la teoría de la informática (Resultado [a] nivel 2).
Analiza, diseña y propone soluciones frente a problemas bioinformáticos. (Resultado [b] nivel 1, Resultado [c] nivel 1, Resultado [d] nivel 1).
Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas. (Resultado [a] nivel 2 y Resultado [h] nivel 2).

7. Breve lista de temas a ser abordados en el curso:

Biología molecular
Alineamiento de Secuencias
Árboles Filogenéticos
Ensamblaje de Secuencias
Tópicos en Bioinformática



4. Prueba de Entrada

Examen de entrada

MSc. Vicente Machaca Arceda
13 de abril de 2021

Nombre:
Apellidos:
CUI:

Preguntas

1. Explique qué es el ADN. **(4 puntos)**
2. Explique qué son los genes y proteínas. **(4 puntos)**
3. ¿A que cree que se deban las mutaciones en el ADN?. **(4 puntos)**
4. Explique como es el proceso de aprendizaje de los modelos de *machine learning*. **(4 puntos)**
5. Implementar un programa en el lenguaje de su preferencia que reciba como entrada dos cadenas de texto y retorne un valor numérico indicando el grado de similitud entre dichas cadenas. Usted puede definir qué criterios tomar para retornar el grado de similitud. **(4 puntos)**

Ejemplos de cadenas similares:

- cadena_1: ACGT
- cadena_2: ACGGT

- cadena_1: GTAACGT
- cadena_2: GTAAGT

- cadena_1: ACGT
- cadena_2: AGGT

Examen de entrada

MSc. Vicente Machaca Arceda
13 de abril de 2021

Nombre:
Apellidos:
CUI:

Preguntas

1. Explique qué es el ADN. **(4 puntos)**

Solución: Cadena de moléculas basadas en Adenine, Cytosine, Guanine y Tyamine. Dicha cadena forma una doble helice, y son el sustrato para la creación de proteínas.

2. Explique qué son los genes y proteínas. **(4 puntos)**

Solución: Los genes son segmentos del ADN y cada gen forma entre una a más proteínas (isomorfos)

3. ¿A que cree que se deban las mutaciones en el ADN?. **(4 puntos)**

Solución: Las mutaciones son cambios en las bases nitrogenadas del ADN, estos cambios originan proteínas malformadas y funciones deterioradas. Algunas causas de su origen son los hábitos de fumar, exposición a radiación y algunas sustancias químicas.

4. Explique como es el proceso de aprendizaje de los modelos de *machine learning*. **(4 puntos)**

Solución: Mediante la gradiente descendiente, en este proceso el error va reduciendo mediante iteraciones sobre la función costo hasta llegar a un minimo local.

5. Implementar un programa en el lenguaje de su preferencia que reciba como entrada dos cadenas de texto y retorne un valor numérico indicando el grado de similitud entre dichas cadenas. Usted puede definir qué criterios tomar para retornar el grado de similitud. **(4 puntos)**

Ejemplos de cadenas similares:

- cadena_1: ACGT
- cadena_2: ACGGT

- cadena_1: GTAACGT
- cadena_2: GTAAGT

- cadena_1: ACGT
- cadena_2: AGGT

Solución:

```
index = 0
similitud = 0
for c in cadena_1:
    if index < len(cadena_2) and c == cadena_2[index]:
        similitud += 1
    index += 1
return similitud
```

4.1 Evidencias

Rindieron la Prueba de Entrada 25 estudiantes de los 25 estudiantes matriculados, lo que representa un 100%.

CUI	APELLIDOS	NOMBRES	EC1
20160748	CESPEDES/FUENTES	RENATO GONZALO	17
20170732	HERMOZA/LOAYZA	MIGUEL ANGEL	18
20160746	INCA/CHIPANA	GUSTAVO HERNAN	13
20173449	TORRES/RODRIGUEZ	JAIME FRANCISCO	16
20170737	VICENTE/CASTRO	RENZO OMAR	17
20153695	APAZA CHAVEZ	MARIA LOURDES	17
20123493	BARRIOS/CORNEJO	SELENE	18
20163427	GOMEZ/CONTRERAS	JUNIOR VALENTIN	16
20041749	PILCO/PANCCA	LUZ MARINA	18
20123377	BUSTINZA/CORNEJO	ALEJANDRA PAMELA	17
20151124	NIFLA/CATASI	WILLIAMS FIDEL	17
20160759	OXA/CACYA	SHIRLEY MICHELLE	18
20160750	PANIBRA/MAMANI	THALES GONZALO	18
20143484	AZA/MAMANI	NICOLL DEL ROSARIO	13
20143490	CHUCTAYA/ELME	MILAGROS	14
20143482	GUTIERREZ/GUTIERREZ	DIEGO ANTONY	15
20110202	TACORA/CRUZ	RICHARD JAVIER	14
20163436	CUEVA/FLORES	JONATHAN BRANDON	17
20170734	FERNANDEZ/MAMANI	BRAYAN GINO	18
20173462	GARCIA/DIAZ	GERMAN FLAVIO	18
20173453	QUISPE/MENOR	HERMOGENES	19
20132402	CRUZ/MAMANI	MILAGROS CELIA	15
20052826	ESPINEL QUISPE	INGRID SALLY	16
20170735	HERRERA/COOPER	MIGUEL ALEXANDER	15
20153709	QUISPE/QUISPE	YARA JEANETTE	15



5. Evaluación Primer Parcial

Primer examen parcial

MSc. Vicente Machaca Arceda

8 de junio de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencia de la Computación	Computación Molecular Biológica

1. Preguntas

1. ¿En que lugar de las celulas, NO esta presente el DNA?
2. ¿Cuáles son las bases nitrogenadas presentes en el RNA?
3. ¿Cuáles son las bases nitrogenadas presentes en el DNA?
4. What is bioinformatics?
5. ¿Qué carbono del azucar ribosa es utilizado para la unión de los nucleotidos durante la replicación de DNA?
6. ¿Qué son los "codons" durante la transcripción?
7. ¿Qué es un dNTP?
8. ¿Qué es un ddNTP?
9. ¿Cuáles son correctas respecto a gel electrophoresis y capillar electrophoresis
10. ¿Cuál es la característica del cDNA durante RNA sequencing?

Primer examen parcial - Solución

MSc. Vicente Machaca Arceda

8 de junio de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencia de la Computación	Computación Molecular Biológica

1. Preguntas

1. ¿En que lugar de las células, NO está presente el DNA?
Núcleo, mitocondrias y cloroplasto.
2. ¿Cuáles son las bases nitrogenadas presentes en el RNA?
Adenine (A), uracil (U), guanine (G), and cytosine (C)
3. ¿Cuáles son las bases nitrogenadas presentes en el DNA?
Adenine (A), guanine (G), thymine (T), and cytosine (C)
4. What is bioinformatics?
Bioinformatics involves the technology that uses computers for storage, retrieval, manipulation, and distribution of information related to biological macromolecules such as DNA, RNA, and proteins. Is limited to sequence, structural, and functional analysis of genes and its products. It is the same as Computational molecular biology
5. ¿Qué carbono del azúcar ribosa es utilizado para la unión de los nucleótidos durante la replicación de DNA?
Carbono 3
6. ¿Qué son los "codons" durante la transcripción?
Son conjuntos de 3 nucleótidos y son utilizados para sintetizar aminoácidos en los ribosomas
7. ¿Qué es un dNTP?
Es el nucleótido de DNA sin grupo hidroxyl en el carbono 2 de su ribosa. Es el sustrato utilizado por DNA polimerasa durante la replicación de DNA. Significa: deoxy Nucleoside triphosphate
8. ¿Qué es un ddNTP?
Es el nucleótido de DNA sin grupo hidroxyl en el carbono 3 de su ribosa. Significa: dideoxy Nucleoside triphosphate
9. ¿Cuáles son correctas respecto a gel electrophoresis y capillary electrophoresis?
Capillary electrophoresis utiliza unos tubos por donde pasan los fragmentos y son captados con un laser. Gel electrophoresis utiliza una lamina de rayos X para leer los fragmentos
10. ¿Cuál es la característica del cDNA durante RNA sequencing?
Es un DNA sin intrones

5.1 Evidencias

Rindieron la Segunda Evaluación Parcial 25 estudiantes de los 25 estudiantes matriculados, lo que representa un 100.0%. La nota promedio 13 puntos.

CUI	APELLIDOS	NOMBRES	EP1
20160748	CEPESDES/FUENTES	RENATO GONZALO	14
20170732	HERMOZA/LOAYZA	MIGUEL ANGEL	16
20160746	INCA/CHIPANA	GUSTAVO HERNAN	2
20173449	TORRES/RODRIGUEZ	JAIME FRANCISCO	12
20170737	VICENTE/CASTRO	RENZO OMAR	14
20153695	APAZA CHAVEZ	MARIA LOURDES	8
20123493	BARRIOS/CORNEJO	SELENE	10
20163427	GOMEZ/CONTRERAS	JUNIOR VALENTIN	6
20041749	PILCO/PANCCA	LUZ MARINA	12
20123377	BUSTINZA/CORNEJO	ALEJANDRA PAMELA	8
20151124	NIFLA/CATASI	WILLIAMS FIDEL	8
20160759	OXA/CACYA	SHIRLEY MICHELLE	12
20160750	PANIBRA/MAMANI	THALES GONZALO	10
20143484	AZA/MAMANI	NICOLL DEL ROSARIO	2
20143490	CHUCTAYA/ELME	MILAGROS	4
20143482	GUTIERREZ/GUTIERREZ	DIEGO ANTONY	6
20110202	TACORA/CRUZ	RICHARD JAVIER	2
20163436	CUEVA/FLORES	JONATHAN BRANDON	10
20170734	FERNANDEZ/MAMANI	BRAYAN GINO	12
20173462	GARCIA/DIAZ	GERMAN FLAVIO	12
20173453	QUISPE/MENOR	HERMOGENES	14
20132402	CRUZ/MAMANI	MILAGROS CELIA	4
20052826	ESPINEL QUISPE	INGRID SALLY	8
20170735	HERRERA/COOPER	MIGUEL ALEXANDER	4
20153709	QUISPE/QUISPE	YARA JEANETTE	4



6. Evaluación Continua 1

Práctica 1

MSc. Vicente Machaca Arceda

30 de julio de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencia de la Computación	Computación Molecular Biológica

PRÁCTICA	TEMA	DURACIÓN
01	Secuenciamiento de ADN	3 horas

1. Competencias del curso

- Aplica las bases matemáticas y la teoría de la informática en algoritmos de Bioinformática.
- Analiza, diseña y propone soluciones frente a problemas bioinformáticos.
- Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas.

2. Competencias de la práctica

- Comprender las bases teóricas del secuenciamiento de ADN.

3. Equipos y materiales

- Editor de texto Latex

4. Entregables

- Se debe elaborar un informe en Latex donde se desarrolle el trabajo solicitado.
- El informe se desarrollará en grupos de 4.
- El informe deberá estar correctamente citado utilizando las normas APA.

5. Desarrollo

1. Desarrolle un informe sobre secuenciamiento de ADN con la siguiente estructura:
 - Definición de secuenciamiento de ADN.
 - Historia del secuenciamiento de ADN (cite los paper donde se publicó cada técnica).
 - Método de Sanger.
 - Problemas actuales del secuenciamiento de ADN.
 - Conclusiones
 - Referencias

Práctica 2

MSc. Vicente Machaca Arceda

12 de mayo de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencia de la Computación	Computación Molecular Biológica

PRÁCTICA	TEMA	DURACIÓN
02	Secuenciamiento de ADN	3 horas

1. Competencias del curso

- Aplica las bases matemáticas y la teoría de la informática en algoritmos de Bioinformática.
- Analiza, diseña y propone soluciones frente a problemas bioinformáticos.
- Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas.

2. Competencias de la práctica

- Comprender los métodos utilizados en *Next Generation Sequence* (NGS)

3. Equipos y materiales

- Editor de texto Latex

4. Entregables

- Se debe elaborar un informe y una presentación en Latex donde se desarrolle el trabajo solicitado.
- El informe se desarrollará en grupos de 4.
- El informe deberá estar correctamente citado utilizando las normas APA.

5. Desarrollo

1. Escoga uno de los siguiente temas:
 - Pyrosequencing (454 Life Sciences)
 - Semiconductor sequencing (Ion Torrent).
 - Reversible chain-termination sequencing (Illumina).
 - Single-molecule sequencing (Pacific Biosciences and MinION)
2. Desarrolle un informe con la siguiente estructura como minimo:
 - Introducción
 - Descripción del método.
 - Pasos o etapas del método.
 - Longitud de fragmentos leídos.
 - Ventajas y desventajas.
 - Conclusiones
 - Referencias

Práctica 3

MSc. Vicente Machaca Arceda

20 de agosto de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencias de la Computación	Computación molecular Biológica

PRÁCTICA	TEMA	DURACIÓN
01	Thresholding	3 horas

1. Competencias del curso

- Aplica las bases matemáticas y la teoría de la informática en algoritmos de Bioinformática.
- Analiza, diseña y propone soluciones frente a problemas bioinformáticos.
- Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas.

2. Competencias de la práctica

- Utilizar herramientas de Dot plot.
- Implementar el algoritmo Dot plot para alineamiento de secuencias.

3. Equipos y materiales

- Python
- Matplotlib
- Numpy
- BioPython
- Cuenta en Github

4. Entregables

- Se debe elaborar un informe en Latex donde se responda a cada ejercicio de la Sección 5.
- En el informe se debe agregar un enlace al repositorio Github donde esta el código.
- En el informe se debe agregar el código fuente así como capturas de pantalla de la ejecución y resultados del mismo.

5. Ejercicios

1. Descargue dos secuencias de proteínas (las utilizadas en clases) y ejecute el siguiente código. Comente sus resultados.

```
# dot_matrix.py
from Bio import SeqIO

sequences = SeqIO.parse("P21333.fasta", "fasta")
for record in sequences:
    data1 = str(record.seq.upper()) # the fasta file just have one sequence

sequences = SeqIO.parse("Q8BTM8.fasta", "fasta")
for record in sequences:
    data2 = str(record.seq.upper()) # the fasta file just have one sequence

print(data1)
print(data2)
```

2. Ahora usted debe implementar un programa que genere un Dot matrix. Se recomienda utilizar Matplot para la gráfica, de igual manera no es necesario dibujar las líneas, basta con dibujar los puntos por cada coincidencia (no olvide incluir el *windows size* y un *threshold*).
3. Descargue otras secuencias y genere el *Dot matrix*, evalúe sus resultados y comente.

Práctica 05

MSc. Vicente Machaca Arceda

22 de junio de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencia de la Computación	Computación Molecular Biológica

PRÁCTICA	TEMA	DURACIÓN
05	Alineamiento de Secuencias con Programación Dinámica	3 horas

1. Competencias del curso

- Aplica las bases matemáticas y la teoría de la informática en algoritmos de Bioinformática.
- Analiza, diseña y propone soluciones frente a problemas bioinformáticos.
- Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas.

2. Competencias de la práctica

- Aplica las bases matemáticas y la teoría de la informática en algoritmos de Alineamiento de Secuencias con Programación Dinámica.

3. Equipos y materiales

- Latex
- Conexión a internet
- Python
- Matplotlib
- Numpy
- BioPython
- Cuenta en Github

4. Entregables

- Se debe elaborar un informe en Latex donde se responda a cada ejercicio de la Sección 5.
- En el informe se debe agregar un enlace al repositorio Github donde está el código.
- En el informe se debe agregar el código fuente así como capturas de pantalla de la ejecución y resultados del mismo.

5. Ejercicios

1. Encuentre el mejor alineamiento global entre las secuencias **AAAC** y **AGC**, con el siguiente *scoring scheme*: *+1 for match, -1 for mismatch and -2 for an alignment with a gap*.
2. Encuentre el mejor alineamiento global entre las secuencias **ATAG** y **TTCG**, con el siguiente *scoring scheme*: *+1 for match, -1 for mismatch and -1 for an alignment with a gap*.
3. Encuentre el mejor alineamiento local entre las secuencias **ATACTGGG** y **TGACTGAG**, con el siguiente *scoring scheme*: *+1 for match, -1 for mismatch and -2 for an alignment with a gap*.

6.1 Evidencias

La evidencia de los Laboratorios realizados por los estudiantes en la Evaluación Continua 1 se muestran en la siguiente tabla:

Laboratorio	Evidencia
Laboratorio 1	Link
Laboratorio 2	Link
Laboratorio 3	Link
Laboratorio 4	Link

Tabla 6.1: Evidencia Evaluación Continua 1