



Tecnológico Nacional de México

Centro Nacional de Investigación y Desarrollo Tecnológico

Tesis de Maestría

**Modelo conceptual del proceso de anotación del gen
V de anticuerpos basado en la Ciencia del Diseño**

presentada por

Ing. Noemi Rossana Chavez Serret

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis

Dr. Javier Ortiz Hernández

Codirector de tesis

Dra. María Yasmín Hernández Pérez

Cuernavaca, Morelos, México. Agosto de 2024



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Centro Nacional de Investigación
y Desarrollo Tecnológico
Departamento de Ciencias Computacionales

Cuernavaca, Mor. **23/Agosto/2024**

OFICIO No. DCC/139/2024

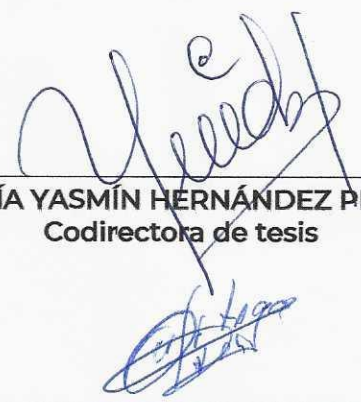
Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de **NOEMI ROSSANA CHÁVEZ SERRET** con número de control **M22CE046**, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Modelo conceptual del proceso de anotación del gen V de anticuerpos badaso en la Ciencia del Diseño"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.



JAVIER ORTIZ HERNÁNDEZ
Director de tesis



MARÍA YASMÍN HERNÁNDEZ PÉREZ
Codirectora de tesis



HUGO ESTRADA ESQUIVEL
Revisor 1



JOAQUÍN PÉREZ ORTEGA
Revisor 2

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante

cenidet
Centro Nacional de Investigación
y Desarrollo Tecnológico



Interior Internado Palmira S/N, Col. Palmira, C. P. 62490, Cuernavaca, Morelos
Tel. 01 (777) 3627770, ext. 3202, e-mail: dcc_cenidet@tecnm.mx | tecnm.mx | cenidet.tecnm.mx



2024
AÑO DE
Felipe Carrillo
PUERTO
GOBIERNO DEL PUEBLO
RECONSTRUCCIÓN Y DESARROLLO
DEL MUNDO



Cuernavaca, Mor.,

No. De Oficio:

Asunto:

29/agosto/2024

SAC/253/2024

Autorización de
impresión de tesis

**NOEMI ROSSANA CHÁVEZ SERRET
CANDIDATA AL GRADO DE MAESTRA
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Modelo conceptual del proceso de anotación del gen V de anticuerpos basado en la Ciencia del Diseño"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y tecnología al servicio de México"



**CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO**

C. c. p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz

Contenido

Resumen.....	vi
Abstract.....	vi
Glosario.....	vii
Presentación.....	ix
Capítulo 1 Introducción.....	1
1.1 Planteamiento del problema.....	1
1.2 Propuesta de solución.....	2
1.2.1 Objetivo general	2
1.2.2 Objetivos específicos	2
1.2.3 Alcances.....	2
1.2.4 Limitaciones.....	2
Capítulo 2 Marco teórico.....	3
2.1 Dominio genómico	3
2.1.1 Gen.....	3
2.1.3 Aminoácidos.....	3
2.1.4 Proteínas.....	3
2.1.5 Proceso de Traducción.....	4
2.1.6 Anticuerpos.....	4
2.1.7 Segmento genético V(D)J.....	4
2.1.8 Recombinación somática	5
2.1.9 Segmento funcional	5
2.2 Dominio computacional.....	6
2.2.2 Propósitos del modelado conceptual.....	6
2.2.3 Enfoques en modelado conceptual.....	7
2.2.4 Ontologías	8
2.2.5 Ciencia del Diseño	8
Capítulo 3 Estado del arte	9
3.1 Antecedentes	9
3.1.1 Discusión de la literatura.....	9
3.2 Estado del arte	10

3.2.1 Artículos relevantes de modelos conceptuales representando al genoma	10
3.2.1.1 Discusión de la literatura	12
3.2.2 Artículos enfocados a modelar diversas problemáticas por medio de modelos conceptuales	18
3.2.2.1 Discusión de la literatura	19
Capítulo 4 Desarrollo del modelo del proceso de anotación del gen V	21
4.1 Marco metodológico	21
4.2 Desarrollo del modelo basado en la metodología ciencia del diseño	23
4.2.1 Ciclo de relevancia	24
4.2.2 Ciclo de rigor	26
4.2.3 Ciclo de diseño	29
4.2.3.1 Selección de modelos conceptuales para la representación del gen V	30
4.2.3.1.1 Diagrama de Clases	30
4.2.3.1.2 Diagrama de actividades	30
4.2.3.1.3 Diagramas BPMN	31
4.2.3.1.4 Modelado SADT	32
4.2.3.2 Comparación de modelos	34
4.2.3.3 Modelo BPMN del proceso de anotación del gen V	37
Capítulo 5 Evaluación de resultados	44
5.1 Fase 1 Aplicación de la Ciencia del Diseño	44
5.2 Fase 2 Conformidad del modelo BPMN	45
5.3 Fase 3 Correctes del modelo por parte de expertos de biología molecular	52
5.4 Fase 4 Aprendizaje del proceso de anotación con el modelo conceptual	60
Capítulo 6 Conclusiones y recomendaciones	70
Referencias	71

CONTENIDO DE FIGURAS

Figura 1 Aminoácidos	3
Figura 2 Representación de receptores de células B	4
Figura 3 Proceso de generación de linfocitos B inmaduros	5
Figura 4 Artifacts of conceptual modeling.....	7
Figura 5 Metodología de solución (Ciencia del Diseño).....	21
Figura 6 Ciclos de Ciencia del Diseño	23
Figura 7 Ontología	29
Figura 8 Ejemplo de clase	30
Figura 9 Elementos de los diagramas de actividades en UML 2.0	31
Figura 10 Elementos BPMN	32
Figura 11 Elementos del modelo SADT	33
Figura 12 Diagrama SADT del proceso de anotación general	34
Figura 13 Diagrama de actividades v1 del proceso de anotación de los genes V	35
Figura 14 Modelo conceptual BPMN del proceso de anotación de los genes V	39
Figura 15 Modelo BPMN del proceso de anotación del gen V (Partes del gen V).....	40
Figura 16 Modelo BPMN del proceso de anotación del gen V (Reglas de la señal de recombinación)	41
Figura 17 Modelo BPMN del proceso de anotación del gen V (Reglas del exón)	42
Figura 18 Modelo BPMN del proceso de anotación del gen V (Reglas del exón L1; péptido señal).....	42
Figura 20 Ejemplo de secuencia de flujo.....	49
Figura 19 Ejemplo de secuencia de flujo	49
Figura 21 Programa Bizagi Modeler	50

CONTENIDO DE TABLAS

Tabla 1 Tesis de maestría de Ciencias en computación desarrolladas en el TecNM CENIDET	9
Tabla 2 Artículos de modelos conceptuales representando al genoma	10
Tabla 3 Artículos modelando diversas problemáticas por medio de modelos conceptuales.....	18
Tabla 4 Artículos relacionados con el modelado conceptual del genoma.....	27
Tabla 5 Artículos representando el proceso de anotación.....	27
Tabla 6 Comparación de modelos conceptuales que representan procesos	34
Tabla 7 Comparación de diagrama de actividades y BPMN.....	36
Tabla 8 Evaluación de la Ciencia del Diseño.....	44
Tabla 9 Elementos y criterios de conformidad de modelado de procesos	47
Tabla 10 Reglas de secuencia de flujo	48
Tabla 11 Base de datos de los encuestados del cuestionario de los expertos	58
Tabla 12 Frecuencia de elementos del cuestionario de los expertos	58
Tabla 13 Porcentaje de la encuesta del cuestionario de los expertos	59
Tabla 14 Base de datos de los encuestados del cuestionario de los aprendices	67
Tabla 15 Frecuencia de elementos del cuestionario de los aprendices	68
Tabla 16 Porcentaje de la encuesta del cuestionario de los aprendices.....	68

Dedicatoria

Principalmente quiero dedicar mi tesis a mi familia por confiar siempre en mí, a mi abuelita Rosa María García por siempre apoyarme y ayudarme en todo, sin ella no estaría aquí.

A mi hermana Nayelli Chavez por siempre admirarme y darme la fuerza para continuar.

A mi mejor amiga Jessica Martínez por siempre estar a mi lado apoyándome, confiando en mí y ayudarme a no rendirme.

A mi familia Mi mamá, mi papá y mi hermano por siempre estar conmigo en cada decisión.

A Frankie por qué hiciste que estos dos años fueran increíbles y por todas esas veces que estuviste a mi lado.

A Aria mi pequeñito motor que me impulsa a seguir adelante, para darle siempre lo mejor.

Agradecimiento

Quiero agradecer al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por otorgar una beca que me permitió cumplir mi objetivo personal de realizar una maestría.

Al Tecnológico Nacional de México (TecNM/CENIDET), por facilitar sus instalaciones para el desarrollo académico, personal y educativo de mi formación estudiantil.

A la Dra. Yasmín, por sus valiosos consejos y conocimientos, siempre dispuesta a brindarlos de la mejor manera, tan amable y dulce.

Al Dr. Joaquín por sus sabias palabras, su sabiduría y su manera de abrirte diversos panoramas.

Al Dr. Hugo por su manera de enseñar, facilitando el aprendizaje y demostrado que siempre hay tiempo para disfrutar.

Al doctor Javier Ortiz Hernández por su paciencia y dedicación, por cada uno de los consejos y gran apoyo durante del desarrollo de esta tesis. Siendo un gran Doctor siempre impulsándonos a seguir adelante, y transmitiéndonos su amor a la ciencia.

Resumen

El modelado conceptual es el proceso para desarrollar un modelo abstracto o una representación de un sistema, así como las entidades involucradas y sus relaciones. Este proceso se puede aplicar a diversos dominios buscando comprender cómo están constituidos y cómo funcionan dichos sistemas. Un dominio de gran interés es el campo de la genómica, el cual ha llevado a la necesidad de atender problemas relacionados con la medicina de precisión, el desarrollo de nuevos medicamentos, el mejoramiento de los cultivos, entre otros. Para desarrollar soluciones basadas en el genoma se requiere adjudicar información biológica para comprender cómo está estructurada y cómo funcionan las secuencias del genoma. Se trata de una actividad compleja realizada por biólogos moleculares a la que se denomina anotación genómica. El proceso de anotación genómica consume una gran cantidad de tiempo y esfuerzo debido a que las anotaciones se realizan en gran parte de manera manual utilizando como herramienta un software para visualizar y anotar el genoma de cualquier especie. En vista de estandarizar este proceso, en este proyecto se desarrolló un modelo conceptual basado en el diagrama de actividades UML, mismo que sirvió de base para construir un modelo BPMN del proceso de anotación que utilizan los biólogos moleculares para determinar la ubicación del gen V de los anticuerpos en quirópteros. Por la naturaleza multidisciplinaria del proyecto, estos trabajos fueron conducidos a través de la metodología Ciencia del Diseño.

Abstract

Conceptual modeling is the process of developing an abstract model or representation of a system, as well as understanding the entities involved and their relationships. This process can be applied to various domains seeking to understand how these systems are constituted and how they function. One domain of great interest is the field of genomics, which has led to the need to address problems related to precision medicine, the development of new drugs, crop improvement, among others. In order to develop genome-based solutions, biological information is required to understand how genome sequences are structured and how they function. This is a complex activity performed by molecular biologists and is called genomic annotation. The genomic annotation process consumes a great deal of time and effort because the annotations are largely done manually using software as a tool to visualize and annotate the genome of any species. In order to standardize this process, this project developed a model of the annotation process by means of the rules used by molecular biologists to determine the location of the V gene of antibodies in chiroptera, using the Design Science methodology.

Glosario

A

ADN: Ácido desoxirribonucleico contiene la información genética de cada ser humano.

Aminoácidos: Un aminoácido es la unidad base que actúa como estructura fundamental de las proteínas. Hay 20 aminoácidos distintos.

Anticuerpos: Los anticuerpos son unas proteínas que forman parte del sistema inmune y circulan por la sangre. Cuando reconocen sustancias extrañas para el organismo, como los virus y las bacterias o sus toxinas, las neutralizan.

B

Bioinformática: La Bioinformática es una subdisciplina de la biología y las ciencias computacionales que se encarga de adquirir, almacenar, analizar y diseminar la información biológica, en gran parte correspondiente a las secuencias de ADN y aminoácidos.

Blast: Sistema para alineamiento de secuencias, enfocado en la detección del exón y el péptido señal.

C

Código genético: Son las instrucciones que le dicen a la célula cómo hacer una proteína específica. A, T, C y G, son las "letras" del código del ADN; representan los compuestos químicos adenina (A), timina (T), citosina (C) y guanina (G), respectivamente, que constituyen las bases de nucleótidos del ADN.

Cromosoma: Son estructuras altamente organizadas, formadas por ADN y proteínas, que contiene la mayor parte de la información genética de un ser vivo.

E

Exón: Poseen la información, molde para la síntesis de proteínas.

Exonerate: Es una herramienta genérica para la comparación de secuencias por pares. Se enfoca en la detección del exón y el péptido señal.

F

Fasta: Constan de una extensión "fa". Los archivos son ficheros de texto utilizados para representar secuencias de ácidos nucleicos, consta de una línea de cabecera donde se aloja el identificador, el nombre y los comentarios y la siguiente línea comienza la secuencia biológica.

G

Gen: Es una secuencia lineal de nucleótidos con la información para la síntesis de una proteína.

Genoma: Es el conjunto de instrucciones genéticas que se encuentra en una célula. En los seres humanos, el genoma consiste en 23 pares de cromosomas, que se encuentran en el núcleo.

Genómica: La genómica se refiere al estudio del genoma completo de un organismo, mientras que la genética se refiere al estudio de un gen en concreto.

Gen Funcional: Codifican para proteínas específicas, o segmentos de proteínas, que tienen diferentes funciones en el cuerpo.

Gen no Funcional: No codifican para crear una proteína.

GFF: Los archivos de Formato de función general (GFF) se trata de ficheros con extensión "gff", son archivos de texto plano.

H

Hmmer: Es un sistema enfocado en identificar la señal de recombinación (RSS) a nivel de nucleótidos.

I

Intrones: No contienen información para las proteínas, pero juegan un papel importante para que estas se fabriquen de forma correcta.

M

Marco abierto de lectura: El código genético lee secuencias de ADN en grupos de tres pares de bases, esto significa que, en una molécula de ADN de doble hebra, hay 6 posibles sentidos en los que pueden abrirse marcos de lectura, tres en dirección hacia adelante y tres en reverso.

N

Nucleótido: El ARN y el ADN son polímeros formados por largas cadenas de nucleótidos. Un nucleótido está formado por las bases en el ADN son la adenina (A), citosina (C), guanina (G) y timina (T). En el ARN, la base uracilo (U) ocupa el lugar de la timina.

P

Proteína: Son una clase importante de moléculas que se encuentran en todas las células vivas. Una proteína se compone de una o más cadenas largas de aminoácidos, cuya secuencia corresponde a la secuencia de ADN del gen que la codifica. Las proteínas desempeñan gran variedad de funciones en la célula, como los anticuerpos.

Péptido señal: El péptido señal (SP) decide sobre el destino, la ruta de transporte y la eficiencia de secreción de una proteína.

T

Traducción: Es el proceso de traducir la secuencia de una molécula de ARN mensajero (ARNm) a una secuencia de aminoácidos durante la síntesis de proteínas.

Transcripción: Es el proceso por el cual se genera una copia de ARN a partir la secuencia de un gen. Esta copia, llamada una molécula de ARN mensajero (ARNm), deja el núcleo de la célula y entra en el citoplasma, donde dirige la síntesis de la proteína, que codifica.

S

Segmento: Es una sección específica de la cadena de ADN que contiene información genética. Puede ser corto o largo y está delimitado por un punto de inicio y un punto final. Los segmentos de ADN contienen secuencias de ADN, pero una secuencia de ADN puede abarcar diferentes segmentos o partes de la cadena.

Segmentos V(D)J: Es una región específica del ADN que se encarga de la diversidad de anticuerpos en el sistema inmune.

Secuencia: Es el orden específico de los nucleótidos que forman al ADN. Como son A, T, C y G y contienen la información genética de la célula. La secuencia de ADN es una "cadena de letras" que representa la información genética en forma de código.

Señal de recombinación: La secuencia de señal de recombinación (RSS) sirve como el sitio de división de ADN.

Presentación

Las secuencias del ácido desoxirribonucleico (ADN) contienen las instrucciones genéticas que involucran el desarrollo y funcionamiento de un organismo (Julia María Lesher Gordillo & Carlos Alfonso Tovilla Zárate, 2013). Los cuatro componentes básicos del ADN son Adenina (A), Citosina (C), Guanina (G) y Timina (T) denominados nucleótidos (Lewin, 2008). Al combinar tres nucleótidos se obtienen los aminoácidos, los cuales son capaces de formar proteínas. Un ejemplo de proteínas son los anticuerpos que se encuentran en el sistema inmune.

El sistema inmune se compone de dos mecanismos, el primero es el sistema inmune innato, el cual es la primera línea de defensa. El segundo mecanismo, es el sistema inmune adaptativo, el cual se va desarrollando con el paso del tiempo a través del contacto de individuos con su ambiente y los antígenos presentes en él: virus, bacterias, polen, polvo, etc. (Brandan et al., 2007). Los antígenos se combaten y eliminan a través del sistema inmune adaptativo, el cual es el encargado de producir anticuerpos.

Los anticuerpos pueden combatir a los antígenos gracias a su estructura compuesta por cuatro cadenas, dos ligeras y dos pesadas, en forma de “Y”, las cuales a su vez se dividen en zona variable (V) y zona constante (C). Las regiones variables de cada una de estas cadenas son las encargadas de la identificación de los antígenos. Las regiones variables están determinadas por distintas combinaciones génicas, como son los genes V (variable), D (diversidad) y J (unión). Es decir, son los encargados de generar un anticuerpo funcional capaz de combatir a un antígeno (García Mariscal et al., 2013).

Por razones que se desconocen, los organismos producen genes funcionales y genes no funcionales. Un gen funcional es aquel que produce un anticuerpo, mientras que los genes no funcionales no logran producir anticuerpos. La producción de anticuerpos es “Un proceso de recombinación V(D)J que involucra numerosos elementos” (Fotini et al., 1997).

La determinación de la funcionalidad de los genes V, D, J es una tarea compleja debido a las partes y reglas que intervienen en el funcionamiento del genoma humano para producir genes funcionales. Los biólogos moleculares son los encargados de realizar un proceso denominado “proceso de anotación” en el cual identifican los genes funcionales y no funcionales con base a las reglas y elementos definidos en biología molecular. Se trata de un proceso que requiere mucho esfuerzo y que puede necesitar hasta un año para la identificación de los genes de un solo individuo. Uno de los genes más complejos es el gen V, debido a su estructura y elementos que lo componen. Por estas razones, es de interés de los biólogos moleculares estandarizar cada una de las actividades que se llevan a cabo, para en un futuro sistematizar aquellas que resulten repetitivas o que impliquen procesar grandes cantidades de datos. Por ello, en este trabajo se propuso realizar una conceptualización del proceso de anotación del gen V, que pueda servir de base para análisis, mejora y una eventual sistematización.

Debido a la naturaleza multidisciplinaria de este proyecto, se utilizó la metodología Ciencia del Diseño. Esta metodología busca la construcción de artefactos que brinden una solución útil y efectiva a un problema en un dominio dado. Se distinguen cuatro tipos de artefactos: constructores (vocabularios y símbolos), modelos (abstracciones y representaciones), métodos (algoritmos y prácticas) e instanciación (implementaciones y prototipos). Para la

construcción de cualquier artefacto se llevan a cabo ciertas actividades de diseño, construcción y evaluación, que iteran tantas veces como sean necesarias antes que el artefacto sea finalmente generado y comunicado para su utilización (Tebes et al., 2019).

El artefacto que se propuso desarrollar fue un modelo conceptual, debido a que los modelos conceptuales permiten representar problemas complejos que para su comprensión requieren un alto nivel de abstracción. Se obtuvieron diferentes vistas del dominio con el objetivo de facilitar la comprensión del problema. Durante el proceso de construcción del modelo los expertos en computación trabajaron a la par con biólogos moleculares. Los modelos conceptuales permiten una representación única del conocimiento, una mejor comprensión del fenómeno tanto biológico como computacional y la posibilidad desarrollar una sistematización del proceso de anotación.

Si bien se han encontrado diversos trabajos que representan el genoma humano mediante modelos conceptuales, existen muy pocos trabajos que expliquen y describan el proceso de anotación del genoma. El proceso de anotación genómica adjudica información biológica para comprender cómo está estructurada y cómo funcionan las secuencias del genoma.

En vista de representar y estandarizar este proceso, en este proyecto se modeló el proceso de anotación del gen V de los anticuerpos en quirópteros, identificando sus partes, reglas, funcionalidad, así como el inicio y fin de este gen. Estos modelos fueron evaluados y corregidos por los expertos del CISEI/INSP en el transcurso del desarrollo de este proyecto.

A continuación, se presenta el documento de tesis, el cual se encuentra dividido en 7 capítulos

Capítulo 1 Introducción: Presenta el planteamiento del problema y el enfoque de solución que se siguió durante el desarrollo del proyecto.

Capítulo 2 Marco teórico: Se definen conceptos básicos para una mejor comprensión de los términos empleados en este proyecto.

Capítulo 3 Estado del arte: Se describen investigaciones afines al tema, se muestran diversos trabajos de tesis realizados en el CENIDET y otras instituciones.

Capítulo 4 Marco metodológico: Se presenta la metodología de solución que se utilizó a lo largo del proyecto para resolver la problemática planteada.

Capítulo 5 Aplicación de la metodología Ciencia del Diseño: Se presenta la metodología Ciencia del Diseño, la cual se utilizó para el desarrollo del modelo conceptual del proceso de anotación del gen V.

Capítulo 6 Evaluación de resultados: Se describen las fases que fueron esenciales para la evaluación del modelo conceptual desarrollado.

Capítulo 7 Conclusiones: Se presentan las conclusiones que se obtuvieron respecto al proyecto.

Capítulo 1 Introducción

1.1 Planteamiento del problema

Uno de los mayores desafíos en el campo de las ciencias de la computación es la falta de entendimiento de los problemas como requisito necesario para desarrollar una adecuada solución. En este sentido, los modelos conceptuales son una herramienta para la representación del dominio del problema a diversos niveles de abstracción, la identificación de las propiedades relevantes de un sistema, así como su comportamiento, llevándolo a un alto nivel de abstracción (O. Pastor et al., 2012).

El desarrollo de un modelo conceptual requiere conocimiento en el dominio del problema, en este caso para un modelado biológico se requiere conocer el proceso biológico y su comportamiento con relación al genoma humano, los elementos que interactúan, sus funciones y relaciones (Frederick W et al., 1992).

En el CISEI (Centro de Investigación Sobre Enfermedades Infecciosas) del INSP (Instituto Nacional de Salud Pública), necesitan reducir el tiempo para realizar la anotación es el problema práctico que tienen los biólogos moleculares para realizar su trabajo. Una forma de contribuir a atender el problema es mediante el desarrollo de un modelo que representa las actividades relevantes que realizan y el orden en que son ejecutadas, de una manera estándar y mediante reglas precisas. En este sentido el problema computacional consiste en comprender e interpretar en términos lógicos y de procesamiento de datos, cada una de las actividades que realizan cada uno de los diversos actores involucrados. Representar este modelo en términos de un proceso, entradas de datos (fuentes de datos) y salidas (conjuntos de datos o resultados obtenidos), así como reglas decisión. Este modelo para ser computacional debe ser computable y debe cumplir con ciertas reglas de conformidad respecto a alguno modelo para representar procesos como es el caso de BPMN. Hay diversas formas de representar un modelo computacional, pueden ser modelos formales como son las gramáticas, mediante algoritmos o mediante naciones gráficas, lo importante es que sigan reglas de construcción basadas en estándares industriales.

1.2 Propuesta de solución

1.2.1 Objetivo general

Desarrollar un modelo conceptual del proceso de anotación del gen V, para identificar las partes y las reglas que determinan la funcionalidad de anticuerpos con base a la Ciencia del Diseño.

1.2.2 Objetivos específicos

- Realizar un estudio del proceso de anotación manual que realizan los biólogos moleculares para la determinación de la funcionalidad del gen V de anticuerpos en quirópteros.
- Realizar un estudio comparativo de trabajos relacionados con el modelado conceptual del genoma para la selección de un modelo apropiado al objetivo general.
- Desarrollar un modelo conceptual de las partes y reglas de composición del gen V de anticuerpos en quirópteros.
- Realizar una validación empírica de los modelos conceptuales obtenidos por parte de expertos en biología molecular.

1.2.3 Alcances

- La investigación estará dirigida por la metodología Ciencia del Diseño.
- El modelo propuesto será fácil de entender por expertos en biología molecular y por expertos en computación.
- El modelo propuesto servirá como base para otros trabajos de investigación en curso, relacionados con el análisis y caracterización de la funcionalidad de segmentos génicos V(D)J.
- El modelo desarrollado apoyará el aprendizaje de biólogos en el proceso de anotación del gen V.

1.2.4 Limitaciones

- El modelo se basará en la información del dominio que brindan los expertos en biología molecular del CISEI/INSP.
- La utilidad de los modelos desarrollados será únicamente evaluada por un grupo de expertos en biología molecular del CISEI/INSP y bioinformáticos.
- La correctes del modelo conceptual del proceso de anotación del gen V solo será evaluado por tres **expertos**.
- La evaluación del modelo para efectos de aprendizaje solo será evaluada por 12 aprendices de un grupo de estudiantes que se encuentran colaborando con el CISEI/INSP.

Capítulo 2 Marco teórico

2.1 Dominio genómico

2.1.1 Gen

Un gen es una secuencia de ADN que codifica a un ARN; en los genes codificadores de proteínas (Lewin, 2008).

Los genes son secuencias que son transcritas en proteínas. Estos poseen diferentes números de nucleótidos en sus secuencias. Cada gen puede ser localizable dentro del cromosoma y es posible utilizar la información genética que posee (O. Pastor et al., 2012).

2.1.2 Nucleótidos

Son moléculas de ADN que proporcionan nuestros rasgos genéticos únicos. Se componen de (A) adenina, (C) citosina, (G) guanina o (T) timina (Stephen A. White et al., 2009).

2.1.3 Aminoácidos

Son moléculas que se combinan para formar proteínas. Los aminoácidos y las proteínas son los pilares fundamentales de la vida.

2.1.4 Proteínas

Las proteínas se forman a partir de la información suministrada por los genes. Se describen como la combinación de 20 tipos de aminoácidos (la base de las proteínas). Por lo tanto, una proteína es la secuencia en la que se disponen dichos aminoácidos de acuerdo con las instrucciones indicadas por el ADN.

Los aminoácidos Fig.1 que componen las proteínas son 20:

Ala (A)	GCU, GCC, GCA, GCG	Lys (K)	AAA, AAG
Arg (R)	CGU, CGC, CGA, CGG, AGA, AGG	Met (M)	AUG
Asn (N)	AAU, AAC	Phe (F)	UUU, UUC
Asp (D)	GAU, GAC	Pro (P)	CCU, CCC, CCA, CCG
Cys (C)	UGU, UGC	Sec (U)	UGA
Gln (Q)	CAA, CAG	Ser (S)	UCU, UCC, UCA, UCG, AGU, AGC
Glu (E)	GAA, GAG	Thr (T)	ACU, ACC, ACA, ACG
Gly (G)	GGU, GGC, GGA, GGG	Trp (W)	UGG
His (H)	CAU, CAC	Tyr (Y)	UAU, UAC
Ile (I)	AUU, AUC, AUA	Val (V)	GUU, GUC, GUA, GUG
Leu (L)	UUA, UUG, CUU, CUC, CUA, CUG		
Comienzo	AUG	Parada	UAG, UGA, UAA

Figura 1 Aminoácidos (O. Pastor et al., 2012)

2.1.5 Proceso de Traducción

La traducción es el proceso de fabricar una proteína siguiendo las instrucciones almacenadas en el ARNm que pasa de un lenguaje de 4 letras (los 4 nucleótidos), en que está transcrito el ARNm, al lenguaje de 20 letras (los 20 aminoácidos) en que están escritas las proteínas (O. Pastor et al., 2012).

2.1.6 Anticuerpos

Los anticuerpos se dividen en una región variable y una región constante, la primera consiste en una sección del anticuerpo alojada en sus extremos, el cual cumple la función de reconocer y unirse a los antígenos. La región constante da forma al anticuerpo, determinando su función. La Figura 2 muestra una representación del anticuerpo.

La estructura de la región variable de cada cadena sea pesada o ligera está determinado por un conjunto de segmentos génicos. La cadena pesada (IgH) se codifica por tres segmentos génicos V, D y J. La cadena ligera (IgL) se compone de los segmentos V y J.

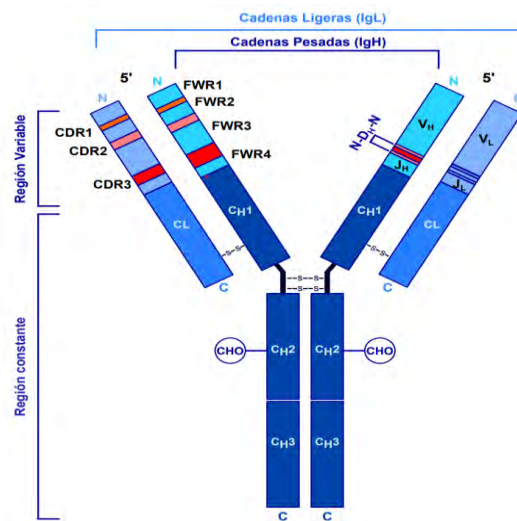


Figura 2 Representación de receptores de células B (Godoy Lozano, 2015)

2.1.7 Segmento genético V(D)J

La diversidad combinatoria de los segmentos génicos que codifican las regiones variables de las cadenas pesadas y ligeras es el proceso de mayor relevancia en la creación de la diversidad de anticuerpos.

La recombinación es un proceso aleatorio mediante el cual 46 segmentos V, 23 segmentos D y 6 segmentos J de la cadena pesada; 41 segmentos V y 5 segmentos J de la cadena ligera kappa; y 74 segmentos V y 7 segmentos J de la cadena ligera lambda permiten la diversidad de la estructura de la zona variable. La recombinación de los segmentos V, D y J puede

generar aproximadamente 6,000 regiones variables distintas para las cadenas pesadas humanas.

Durante el desarrollo de los linfocitos B, el primer evento de recombinación se da entre los segmentos D y J del clúster de la cadena pesada eliminando el fragmento de ácido desoxirribonucleico (ADN) entre los segmentos seleccionados. Posteriormente, se lleva a cabo la recombinación con el segmento V para dar lugar a un gen V(D)J reorganizado (García Merino, 2011)(Godoy Lozano, 2015).

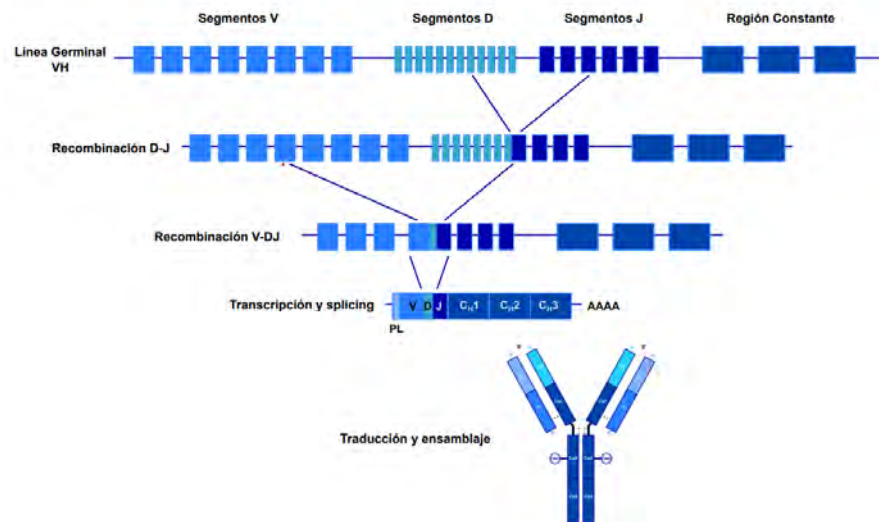


Figura 3 Proceso de generación de linfocitos B inmaduros (Godoy Lozano, 2015)

2.1.8 Recombinación somática

“Recombinación somática”, que se lleva a cabo por un complejo enzimático cuyo nombre es “*Recombinasas V(D)J*” en la médula ósea. La recombinación somática es un proceso por el cual diferentes segmentos de ADN se mezclan al azar para formar un gen, de este modo se explica cómo unos pocos segmentos al combinarse de todas las maneras posibles producen dicha cantidad de Ig (Inmunoglobulinas) (Brandan et al., 2007).

2.1.9 Segmento funcional

Los segmentos funcionales son todos aquellos que logran codificar un gen, transformándolo en un anticuerpo o también conocido como proteínas, listo para combatir a los antígenos.

2.2 Dominio computacional

2.2.1 Fundamentos del modelado conceptual

En sus orígenes el modelado conceptual se centró en el campo de las bases de datos. Se desarrollaron dos propuestas: en 1974 el modelo semántico propuesto por Abrial y en 1976 el modelo entidad-relación (ER) propuesto por Chen.

- El modelo semántico es un modelo conceptual de datos que incluye información semántica. Quiere decir que el modelo describe el significado de sus instancias. Este modelo es una abstracción que define cómo los símbolos almacenados (los datos de la instancia) se relacionan con el mundo real.
- El modelo entidad-relación: permite definir las necesidades de información de una organización. Identificar los asuntos de importancia dentro de una organización (entidades), las propiedades de esos asuntos (atributos) y cómo se relacionan entre sí (relaciones) (Richard Barker, 1990). Es una herramienta para el modelado de datos que facilita la representación de entidades en la construcción de bases de datos.

Más tarde surgen nuevas aplicaciones del modelado conceptual bajo el concepto de modelado de procesos de negocios, incorporando un nuevo papel en el campo de la ingeniería de software (Robinson et al., 2016). En 1995 es desarrollado el Lenguaje Unificado de Modelado (UML, por sus siglas en inglés). UML permitiendo modelar aplicaciones en diversos dominios, tales como telecomunicaciones, comercio, salud, entre otros (Fuentes & Vallecillo, 2004).

- Modelo orientado a objetos: son útiles para captar los aspectos de comportamiento de los sistemas. Describe el comportamiento en términos de los ciclos de vida de los objetos (a menudo representados como una máquina de estados) y sus interacciones con otros objetos (Liddle, 2011). Permite la identificación y comunicación de objetos y admite las propiedades de abstracción, herencia y encapsulación de datos.
- Los modelos UML: es un lenguaje estándar utilizado para especificar y documentar sistemas de forma precisa. Dependiendo de los aspectos del sistema que nos interesa observar y describir ofrece diversos diagramas como: diagrama de estructura, diagrama de comportamiento, diagramas de interacción.

2.2.2 Propósitos del modelado conceptual

El modelado conceptual es una parte importante de la ingeniería de requisitos, la primera y más importante fase en el desarrollo de un sistema de información (Olivé, 2007).

Un modelo conceptual es una representación visual de las clases conceptuales u objetos del mundo real en un dominio de interés (Jacobson et al., 2000). Permite identificar, organizar y realizar razonamientos sobre los componentes y comportamiento del sistema, son la guía para el proceso de diseño del software (Gea et al., 2003) (Grady Booch & James Rumbaugh, 2011).

Los modelos conceptuales proporcionan una definición clara del dominio, lo que permite que se puedan entender mucho mejor las entidades involucradas y sus relaciones. Por eso es

ampliamente aceptado que la aplicación de modelado conceptual facilita la comprensión de dominios complejos (Reyes Román & Pastor López, 2018).

Un modelo conceptual evoluciona en dos etapas de acuerdo con (Birta & Arbez, 2013).

- La etapa inicial “alto nivel” proporciona una perspectiva preliminar para el modelo, sin excesivos detalles.
- La segunda etapa “nivel detallado” profundiza en los aspectos del comportamiento a través de vistas. Es una representación que puede servir de especificación para un programa.

2.2.3 Enfoques en modelado conceptual

En los modelos conceptuales existen diversos enfoques que establecen, qué es, cuál es el propósito y cómo se construye un modelo conceptual (Reyes Román, 2016) propone el siguiente enfoque:

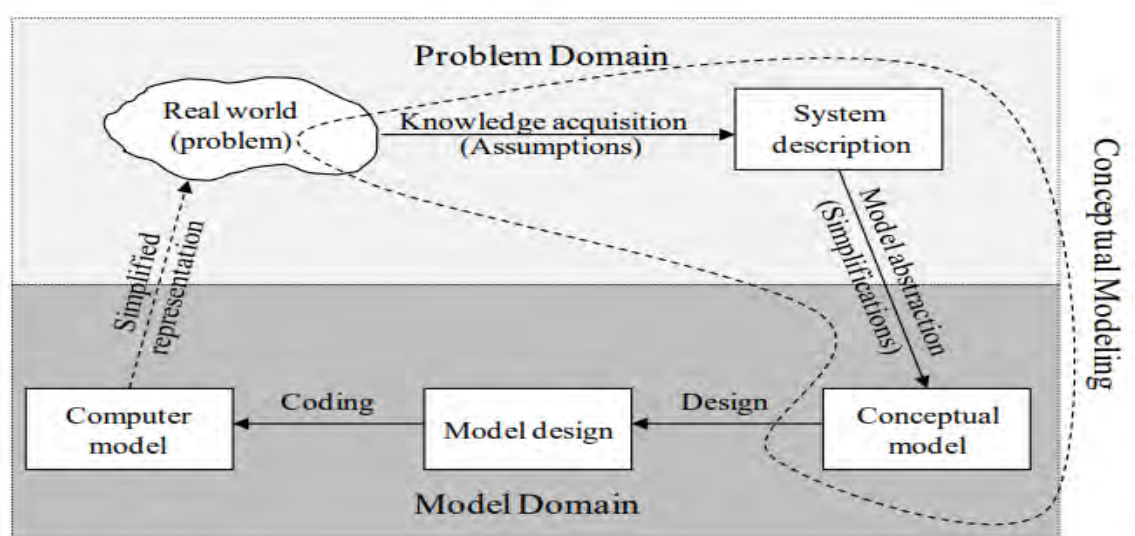


Figura 4 Artifacts of conceptual modeling (Robinson, 2010, 2013, 2014).

En la figura 4 se muestran los artefactos clave del modelado conceptual. La nube representa el mundo real o problemática. Los cuatro rectángulos representan el proceso del modelado conceptual:

- Descripción del sistema: describe el problema y los elementos que se relacionan.
- Modelado conceptual: artefactos claves del modelo.
- Diseño del modelado: son los detalles técnicos, contiene datos, componentes, ejecución del modelo, entre otros.
- Modelo informático: representación específica en software del modelo conceptual.

Las flechas muestran el flujo de información que se describe como adquisición de conocimiento, abstracción de modelos, diseño y codificación. Principalmente, destacan las funciones específicas y diferentes de las suposiciones y simplificaciones. Las suposiciones están relacionadas con la adquisición de conocimientos y las simplificaciones están relacionadas con la abstracción del modelo.

Este enfoque es de gran ayuda cuando se plantea realizar un modelo conceptual. Se debe contemplar los artefactos claves que describen el enfoque para tener una mejor visión del dominio que se desea representar. El enfoque se puede utilizar en diversas aplicaciones dentro de campos específicos.

2.2.4 Ontologías

Un modelo ontológico hace suposiciones específicas sobre lo que existe y cómo se comportan las cosas (Evermann & Wand, 2001). En el ámbito de la Ciencia de la Computación, para un sistema lo que “existe” solo es aquello que puede ser “representado” (Guarino et al., 2009).

La ontología es un sistema de representación del conocimiento que resulta de la selección de un ámbito o dominio del conocimiento, las ontologías se pueden organizar en estructuras jerárquicas, una de las mejores formas para representar el conocimiento (María Jesús Lamarca Lapuente., 2013). Los tipos de ontologías son:

- **Ontologías de alto nivel o genéricas:** Describen conceptos más generales como tiempo, espacio, objetos
- **Ontologías de dominio:** Describen un vocabulario relacionado con un dominio genérico.
- **Ontologías de tareas o técnicas básicas:** Describen una tarea, actividad o artefacto, por ejemplo, componentes, procesos o funciones.
- **Ontologías de aplicación:** Describen conceptos que dependen tanto de un dominio específico como de una tarea específica, y generalmente son una especialización de ambas.

2.2.5 Ciencia del Diseño

La ciencia del diseño busca el desarrollo de artefactos útiles en un contexto establecido. Un contexto se define como el conjunto de circunstancias, factores o condiciones que rodean una situación particular y que influyen en ella. La ciencia del diseño busca comprender y definir el entorno en el que se llevará a cabo el diseño, desarrollo e implementación de un artefacto (De Sordi, 2021); (Dresch, 2015); (Insfrán et al., 2014).

En la actualidad se ha utilizado la ciencia del diseño en diversas investigaciones: el ámbito eléctrico, sistemas de información, ingeniería civil, arquitectura, entre otros, pero actualmente no se ha utilizado la ciencia del diseño en el ámbito genómico. En específico en el proceso de la anotación genómica

Capítulo 3 Estado del arte

3.1 Antecedentes

En la tabla 1 se describen los trabajos realizados en el TecNM CENIDET relacionados con modelado conceptual y con ciencias genómicas.

Tabla 1 Tesis de maestría de Ciencias en computación desarrolladas en el TecNM CENIDET

<i>Año</i>	<i>Artículo</i>	<i>Objetivo</i>	<i>Herramienta de modelado</i>
2022	Desarrolló del modelado de un almacén de datos como parte integral de un sistema de inteligencia de negocios (Suarez Mendez, 2018)	Desarrollar un modelo para implementar un almacén de datos en una empresa, con el propósito de apoyar la toma de decisiones estratégicas y tácticas.	Modelo conceptual basado en clases para representar objetos, relaciones y atributos
2019	Análisis de datos genómicos para el diagnóstico temprano de osteosarcoma (Moncada C., 2019)	Desarrollar un sistema de aprendizaje automático para el manejo de datos genómicos relacionados con el diagnóstico temprano de cáncer de hueso en humanos.	Modelos de aprendizaje automático
2009	Recuperación y procesamiento de datos genómicos para la carga de información clasificada en una base de datos basada en un modelo conceptual del genoma humano (Pliego, 2009)	Desarrollar una metodología para la extracción automática y remota de datos de diversas bases de datos genéticas e integrarlas en una base de datos basada en un esquema conceptual del genoma humano.	Modelo conceptual basado en clases del genoma humano

3.1.1 Discusión de la literatura

De esta búsqueda y análisis de los trabajos encontrados principalmente dentro del TecNM CENIDET se puede concluir que se han utilizado modelados conceptuales para su aplicación en diversas disciplinas, además de utilizarlos como herramientas de apoyo para el desarrollo de sistemas de aprendizaje automático. No se encontraron tesis relacionadas con modelos conceptuales del genoma.

3.2 Estado del arte

Para la revisión bibliográfica se planteó resolver la siguiente pregunta ¿Cómo se ha modelado conceptualmente el proceso de anotación genómica y sus reglas?, enfocándose a trabajos que hayan representado conceptualmente el proceso de la anotación. Se generaron diversas cadenas de búsquedas como "genoma humano" and "modelo conceptual", “annotation process” “modelo” AND “conceptual genes”, utilizando la base de datos de IEEE, Springerlink y Science Direct.

Los resultados que se obtuvieron se dividen en dos: modelos conceptuales del genoma en general y modelos conceptuales que resuelven una problemática.

3.2.1 Artículos relevantes de modelos conceptuales representando al genoma

Tabla 2 Artículos de modelos conceptuales representando al genoma

No	Año	Artículo	Descripción general	Modelo utilizado
1	2011	Del Esquema Conceptual al Mundo Real, una Base de Datos Genómica	Se presentó un modelo conceptual que se utilizó en la creación de una base de datos para el diseño e implementación de un sistema de información para resolver los problemas de integración de datos. Enfocado a la búsqueda y recuperación de información sobre la secuenciación de genomas de los seres humanos.	Modelo conceptual de clases
2	2011	Integración de bases de datos genómicas: una aproximación basada en modelado conceptual	El trabajo se centró en el diseño de un modelado conceptual del genoma humano, a partir de modelos ya existentes como el trabajo de (Reyes Román & Pastor López, 2018). El modelo muestra nuevas vistas, obteniendo un modelo conceptual del genoma v3.	Modelo conceptual de clases
3	2013	Sistema de información genómico: Extensión del módulo de carga e integración de información genómica	Se desarrolló un modelo conceptual en donde se integró toda la información asociada a los cromosomas del genoma humano, así como la partición de estos en elementos del cromosoma y un nuevo gen.	Modelo conceptual de entidad-relación
4	2016	Applying Conceptual Modeling to Better Understand the Human Genome	El trabajo representó los beneficios de la aplicación de Modelado Conceptual (CM) en dominios complejos como lo es la genómica, con el objetivo de entender y lograr explicar a través del modelo, la evolución del genoma humano buscando una comprensión clara y precisa del genoma.	Modelo conceptual de entidad-relación
5	2018	Diseño y desarrollo de un sistema de información genómica basado en un modelo conceptual holístico del genoma humano.	Se utilizó un modelo conceptual del genoma humano que permitió un mejor entendimiento para la generación de sistemas de información genómico.	Modelo conceptual de clases
6	2018	Gestión de datos genómicos basada en Modelos Conceptuales	Se realizó un modelo conceptual para definir los elementos que participan en los procesos de análisis en el desarrollo de datos genómicos de cítricos.	Modelo conceptual de clases
7	2018	A Method to Identify Relevant Genome Data: Conceptual Modeling for the Medicine of Precision	Se realizó un método sistemático basado en técnicas de modelado conceptual y gestión de la calidad de los datos con el fin de ayudar al diagnóstico genómico de una enfermedad.	Modelo conceptual de clases

8	2020	Using conceptual modeling to improve genome data management	Se utilizó el modelo conceptual para gestionar de forma eficiente los datos genómicos por medio de un sistema de información, el caso de estudio es la medicina de precisión “epilepsia”.	Modelo conceptual de clases
9	2020	Towards the Understanding of the Human Genome: A Holistic Conceptual Modeling Approach	Se actualizó el modelo conceptual del genoma humano proporcionando un enfoque más flexible ampliando interacciones entre las diferentes partes, agregando nueva información de las proteínas.	Modelo conceptual de clases
10	2020	Conceptual Modeling of Proteins Based on UniProt	Este trabajo se centró el desarrollo de un modelo conceptual enfocado en las proteínas y el papel que desempeñan en las diferentes funciones que se llevan a cabo dentro de la célula de cualquier sistema vivo.	Modelo conceptual de clases
11	2020	Towards the Generation of a Species-Independent Conceptual Schema of the Genome	Se desarrolló un esquema conceptual holístico para trabajar con cada especie eucariota, creando un esquema conceptual global, por medio de una comparación ontológica de dos esquemas conceptuales existentes.	Modelo conceptual de clases
12	2021	On how to generalize species-specific conceptual schemes to generate a species-independent Conceptual Schema of the Genome	Se compararon dos modelos, uno del genoma humano enfocado en medicina de precisión y el segundo enfocado en el genoma de los cítricos con el objetivo de unificar un modelo que sirva para cualquier tipo de especie.	Modelo conceptual de clases, nodos
13	2021	Towards a Shared, Conceptual Model-Based Understanding of Proteins and Their Interactions	Se realizó un modelo conceptual centrándose en las proteínas, identificar algunos problemas y debilidades que dificultaban la comprensión y la representación correcta de conceptos importantes en las proteínas.	Modelo conceptual de clases
14	2021	A Conceptual Model-Based Approach to Improve the Representation and Management of Omics Data in Precision Medicine	Se desarrolló un esquema conceptual que permitió superar limitaciones que se presentan en la medicina de precisión, proporcionando una mejor representación de los datos proteómicos y facilita su integración.	Modelo conceptual de clases
15	2022	Diseño de un Modelo Conceptual para caracterizar proteoma y pathways humanos: aplicaciones al metabolismo de la CoVid-19.	Se creó una nueva vista en el estudio de enfermedades infecciosas (COVID) para el modelo conceptual del genoma humano ya desarrollado.	Modelo conceptual de clases
16	2022	A Comprehensive Approach for the Conceptual Modeling of Genomic Data	Se integraron dos modelos existentes, para lograr una representación más completa y holística de la información genómica, permitiendo establecer vínculos entre los conceptos y los datos genómicos, lo que facilita la expresión de consultas y análisis de manera más eficiente y precisa.	Modelo conceptual de clases

3.2.1.1 Discusión de la literatura

1. Del Esquema Conceptual al Mundo Real, una Base de Datos Genómica (Eraso Schattka, 2011)

Resumen: Se presenta un modelo conceptual que se utilizó en el desarrollo de una base de datos para el diseño e implementación de un sistema de información que contribuye a resolver los problemas de integración de datos. Enfocado a la búsqueda y recuperación de información valiosa acerca de los estudios realizados sobre la secuenciación de genomas de los seres humanos.

Modelo utilizado: Modelo conceptual de clases

Resultados: El modelo conceptual del genoma humano logró ser la base para la transformación de un esquema de base de datos relacional.

Conclusiones: En este trabajo se utilizó como apoyo un modelo conceptual del genoma humano como referencia. El modelo fue modificado para su utilización en una base de datos.

2. Integración de bases de datos genómicas: una aproximación basada en modelado conceptual (Martín Mayordomo, 2011)

Resumen: El trabajo se centra en el diseño de un modelado conceptual del genoma humano, a partir de modelos ya existentes como el trabajo de (Reyes Román & Pastor López, 2018). El modelo mostró nuevas vistas, además se complementaron las vistas ya existentes, obteniendo un modelo conceptual del genoma v3.

Modelo utilizado: Modelo conceptual de clases

Resultados: Se realizó un modelo conceptual del genoma representando y unificando el conocimiento del genoma. El modelo no solo abarca la secuencia del genoma humano, si no contempla de cualquier especie.

Conclusiones: El modelo proporciona vistas que representan el dominio: estructura del genoma, transcripción, variaciones, rutas metabólicas y fuentes de datos y referencias bibliográficas.

3. Sistema de información genómico: Extensión del módulo de carga e integración de información genómica (Guerola Martínez, 2013)

Resumen: Se desarrolla un modelo conceptual que integra toda la información asociada a los cromosomas del genoma humano y la partición de estos en elementos del cromosoma y un nuevo gen.

Modelo utilizado: Modelo conceptual entidad-relación

Resultados: El modelo redujo la gran longitud de secuencias, la secuencia del gen y el posicionamiento tanto a nivel génico como cromosómico, además de la carga de un nuevo gen APC (Adenomatous polyposis coli) es un gen tumoral que codifica una proteína que juega un importante papel en la supresión de tumores.

Conclusiones: El modelo conceptual se centró en ampliar las entidades y relaciones del cromosoma siendo un avance fundamental debido a su importancia dentro del genoma.

4. Applying Conceptual Modeling to Better Understand the Human Genome (Reyes Román, 2016)

Resumen: El trabajo busca representar los beneficios de la aplicación de Modelado Conceptual (CM) en dominios complejos como lo es la genómica, con el objetivo de entender y lograr explicar a través del modelo, la evolución del genoma humano buscando una comprensión clara y precisa del genoma.

Modelo utilizado: Modelo conceptual de entidad-relación

Resultados: El modelo se centró en la arquitectura conceptual, buscando la creación de una visión más actualizada del modelo conceptual del genoma humano v1. Capaz de destacar todas las ventajas de la aplicación del MC en un dominio complejo como el de los Sistemas de Información Genómica (SIGG).

Conclusiones: El uso del modelo conceptual fue una estrategia para el diseño de un sistema de información, esta versión número uno se centró en el modelado de “Genotipado” que logró unificar el modelo conceptual.

5. Diseño y desarrollo de un sistema de información genómica basado en un modelo conceptual holístico del genoma humano (Reyes Román & Pastor López, 2018)

Resumen: Este trabajo se centra en el diseño y análisis de un sistema de información genómica basado en un modelo conceptual del genoma humano. El modelo conceptual utilizado es un diagrama de clases enfocado a la medicina de precisión, el cual muestra cinco vistas, *Structural view* es la encargada de describir la estructura del genoma, *Transcription view* esta vista muestra los componentes implicados en el ADN, ARN y los conceptos relacionados con la síntesis de proteínas, *Variation view*: describe los cambios en la secuencia de referencia, *Pathways view* describe información sobre las rutas metabólicas y *Bibliography and data bank view*, esta vista se encarga de evaluar la fuente de cualquier información con el objetivo de establecer de dónde proviene cualquier dato.

Modelo utilizado: Modelo conceptual de clases

Resultado: En esta investigación se desarrolló un modelo que conceptualizó el genoma, representándolo por medio de vistas para una mejor comprensión del dominio.

Conclusiones: El trabajo descrito es uno de los primeros en representar un modelo conceptual del genoma humano, logró mostrar de manera general los elementos que involucra el desarrollo de la vida, enfocándolo en la medicina de precisión.

6. Gestión de datos genómicos basada en Modelos Conceptuales (García Simón et al., 2018)

Resumen: El trabajo plantea la creación de un modelo conceptual del genoma de los cítricos con el fin de identificar y definir correctamente los elementos que participan en los cítricos y el desarrollo de una aplicación de explotación de datos genómicos.

Modelo utilizado: Modelo de clases

Resultados: Se desarrolló el modelo con base al conocimiento e información de especialistas en el tema, enfocándose a una nueva especie no modelada anteriormente como son los cítricos.

Conclusiones: Este trabajo utilizó diversos archivos para poder obtener los datos y posteriormente crear un modelo, este trabajo se relaciona bastante con el proceso de anotación.

7. A Method to Identify Relevant Genome Data: Conceptual Modeling for the Medicine of Precision (Palacio et al., 2018)

Resumen: El trabajo propone un método sistemático para identificar datos genómicos relevantes para la medicina de precisión utilizando técnicas de modelado conceptual y gestión de calidad de datos. El método consta de dos fases: seleccionar fuentes de datos confiables e identificar variantes de ADN. Los autores validan el método utilizando un estudio de caso sobre el diagnóstico genómico de la enfermedad de Crohn.

Modelo utilizado: Modelo de clases

Resultados: El método presentó una forma sistemática de identificar la información relevante para la curación de variantes o el análisis terciario para proporcionar un diagnóstico genómico preciso.

Conclusiones: El trabajo abordó el proceso inicial donde los biólogos realizan la extracción de información genética, llamado análisis primario y secundario, para posteriormente pasar dicha información al modelo conceptual.

8. Using conceptual modeling to improve genome data management (Ó. Pastor et al., 2020)

Resumen: El trabajo presenta la integración de diferentes perspectivas de los datos genómicos, con base a los modelos desarrollados en trabajos anteriores.

Se realizaron tres versiones de modelado:

- Modelo conceptual del Genoma Humano (CSHG) v1: Fue el primer modelo enfocado en conceptos básicos, ignorando aspectos más complejos, sobre todo se centra en el análisis de genes individuales.
- CSHG v1.1: Este incluye la visión fenotípica, es de importante valor de investigación y aumenta la completitud del modelo.

- CSHG v2: Este modelo contiene una visión gene centrista en el concepto de cromosoma.

Modelo utilizado: Modelo de clases

Resultados: Se presenta un marco de trabajo para el uso del modelo conceptual en la gestión eficiente de datos genómicos, con el objetivo de desarrollar e implementar un sistema de información con datos de alta calidad.

Conclusiones: El trabajo discutió la importancia del modelado conceptual en la genómica y el desarrollo de herramientas como ontologías y bases de datos para unificar el conocimiento. El trabajo ayudará a los investigadores a organizar, almacenar y procesar la información, centrándose solo en los datos que son relevantes.

9. Towards the Understanding of the Human Genome: A Holistic Conceptual Modeling Approach (García S & Casamayor, 2020)

Resumen: El trabajo se centra en el diseño de un modelo conceptual del genoma humano que pueda utilizarse como artefacto clave para compartir, integrar y comprender los distintos datos utilizados en el ámbito genómico. Se proporcionó un enfoque más flexible, ampliando interacciones entre las diferentes partes del modelo, agregando nuevos datos que en los modelos anteriores no se habían considerado.

Modelo utilizado: Modelo conceptual de clases

Resultados: Propuso un nuevo enfoque el cual utilizan un modelo conceptual unificado para representar todos los aspectos del genoma, desde las secuencias de ADN hasta la expresión génica, profundizando en las proteínas.

Conclusiones: Proporcionó una visión más holística, precisa y coherente del genoma. Se creó diversos modelos para representar la jerarquía de los datos en el modelado conceptual, haciendo más comprensibles el dominio.

10. Conceptual Modeling of Proteins Based on UniProt (León Palacio & Pastor López, 2020)

Resumen: Este trabajo se centra en el desarrollo de un modelo conceptual enfocado en las proteínas y el papel que desempeñan en las diferentes funciones que se llevan a cabo dentro de la célula de cualquier sistema vivo.

Modelo utilizado: Modelado conceptual de clases

Resultados: Se desarrolló cuidadosamente un modelo conceptual para las proteínas tomando la base de datos UniProtKB como fuente de datos.

Conclusiones: Se mostró la complejidad conceptual de la base de datos utilizada, y permitió una base conceptual sólida y bien fundamentada, enfocada en las proteínas, siendo uno de los primeros trabajos enfocado en dominio.

11. Towards the Generation of a Species-Independent Conceptual Schema of the Genome (García S & Casamayor, 2020)

Resumen: Se desarrolla un esquema conceptual para trabajar con cada especie eucariota, creando un esquema conceptual global, por medio de una comparación ontológica de dos esquemas conceptuales existentes.

Modelo utilizado: Modelo conceptual de clases, ontologías

Resultados: El modelo conceptual se desarrolló con el objetivo de tener un modelo conceptual único para trabajar con todas las especies eucariotas porque, aunque sean únicas y diversas, su genoma y su comportamiento es el mismo.

Conclusiones: Este modelo conceptual podría funcionar como un elemento global y genérico a partir del cual se puede trabajar con cualquier dominio particular.

12. On how to generalize species-specific conceptual schemes to generate a species-independent Conceptual Schema of the Genome (García S and Casamayor, 2021)

Resumen: El trabajo se centra en la comparación de dos modelos ya desarrollados, el primer modelo es del genoma humano enfocado en medicina de precisión y el segundo se enfoca en el genoma de los cítricos. Los dos trabajos se unificaron para generar un nuevo modelo conceptual del genoma. El esquema conceptual es único y holístico que puede ser válido para trabajar con todas las especies eucariotas debido a que se estudió y se concluyó que es el mismo comportamiento.

Modelo utilizado: Modelo de clases

Resultados: El modelo funcionó como un elemento genérico global a partir del cual se pueden crear vistas conceptuales para trabajar con cualquier especie.

Conclusiones: Se formalizó el modelo a partir de dos modelos ya existentes con el objetivo de adaptarse a cualquier contexto de trabajo específico basado en el genoma. Se utilizaron ontologías para definir los conceptos y relaciones que participaron en dicho modelo.

13. Towards a Shared, Conceptual Model-Based Understanding of Proteins and Their Interactions (Leon & Pastor, 2021)

Resumen: El trabajo se centra en el estudio de las proteínas, se analizan los datos proporcionados por una base de dato de proteínas que es utilizada por los especialistas en biología molecular. Se realizó una conceptualización sólida para identificar los conceptos relevantes del dominio, creando un modelo conceptual.

Modelo utilizado: Modelo conceptual holístico

Resultados: El trabajo se centra en un modelo conceptual enfocado en las proteínas y sus interacciones. Propuso un marco para desarrollar dicho modelo, basado en el uso de ontologías y otras técnicas de representación del conocimiento.

Conclusiones: Un modelo conceptual de las proteínas y sus interacciones es esencial para avanzar en la comprensión de los sistemas biológicos. Además, se representó dentro del modelo, cómo se producen las proteínas, las funciones específicas y la estructura de estas.

14. A Conceptual Model-Based Approach to Improve the Representation and Management of Omics Data in Precision Medicine (Garcia S et al., 2021)

Resumen: El trabajo presenta un esquema conceptual proporcione una mejor representación de los datos proteómicos, se encontraron nuevos descubrimientos y datos como la integración de proteínas y aminoácidos.

Modelo utilizado: Modelo de clases

Conclusiones: El uso de modelado conceptual ayudó a desarrollar herramientas de modelado más precisión. El modelo conceptual puede utilizarse para crear un vocabulario común para representar los datos ómicos. Esto puede ayudar a mejorar la comunicación entre investigadores y clínicos.

Resultados: El modelo conceptual desarrollado profundiza en las proteínas, mostrando los elementos que intervienen en dicha instancia, fue de ayuda para comprender a mayor detalle la estructura de las proteínas.

15. Diseño de un Modelo Conceptual para caracterizar proteoma y *pathways* humanos: aplicaciones al metabolismo de la CoVid-19 (Jaque et al., 2022)

Resumen: El trabajo propone una nueva vista en el modelo conceptual del genoma humano CSHG v3 buscando ampliar el marco de aplicación al estudio de enfermedades infecciosas de origen vírico. Por medio de la instanciación de un caso real de patogénesis de la enfermedad del coronavirus (COVID19).

Modelo utilizado: Modelo conceptual de clases

Resultados: El modelo fue de apoyo para la comprensión de enfermedades infecciosas, el cual hace énfasis en el desarrollo de nuevas instancias, las cuales fueron clave para conceptualizar el problema que se plantea. Principalmente, la instancia de proteoma la cual se enfocó en la generación de proteínas.

Conclusiones: El modelo conceptual fue una herramienta fundamental para el desarrollo del sistema de información. Sin el desarrollo del SI se debía integrar y analizar manualmente la información, lo cual es propenso a errores.

16. A Comprehensive Approach for the Conceptual Modeling of Genomic Data (Bernasconi et al., 2022)

Resumen: El trabajo aborda dos modelos ya creados, el Esquema Conceptual del Genoma Humano (CSG) ofrece una representación orientada a conceptos del comportamiento del genoma independiente de los datos y el Modelo Conceptual Genómico (MCG) ofreció una representación orientada a los datos. Se propuso unir estos dos enfoques para lograr una visión más completa, vinculando (1) una capa de conceptos, que describa los elementos del genoma y sus conexiones conceptuales, con (2) una capa de datos, que describa los conjuntos de datos derivados de la secuenciación del genoma con tecnologías específicas.

Modelo utilizado: Modelo conceptual de clases

Resultados: El modelo logró que sea posible expresar consultas que empleen una perspectiva conceptual holística del genoma, traducida directamente a términos orientados a los datos y a su organización.

Conclusiones: El modelo desarrollado logró representar el dominio genómico orientado a dos vistas: la vista conceptual, la cual indica los elementos que participan dentro del genoma y la segunda vista los datos que representan los elementos. Esta forma de representarlo logra entender de manera general al genoma desde diferentes perspectivas.

3.2.2 Artículos enfocados a modelar diversas problemáticas por medio de modelos conceptuales

Tabla 3 Artículos modelando diversas problemáticas por medio de modelos conceptuales

No	Año	Artículo	Descripción general	Modelo utilizado
1	2018	A Method to identify relevant genome data: conceptual modeling for the medicine of precision	El trabajo presentó un método sistemático basado en técnicas de modelado conceptual con el objetivo de ayudar al diagnóstico genómico de la enfermedad de Crohn.	Modelo conceptual de clases
2	2020	Towards an Ontology for Tertiary Bioinformatics Research Process	El trabajo mostró cómo puede aprovecharse la creación de una ontología para diseñar interfaces que no se centren únicamente en los datos, si no considerar tanto el flujo de trabajo de la investigación como los requisitos de los investigadores.	Modelado conceptual basado en ontologías
3	2022	Conceptual modelling for life sciences based on systemist foundations	El trabajo propuso unir dos enfoques para lograr una visión más completa, vinculando una capa de conceptos, que describa los elementos del genoma y sus conexiones conceptuales, con una capa de datos, que describa los conjuntos de datos derivados de la secuenciación del genoma con tecnologías específicas.	Diagrama de composición
4	2022	Using conceptual models in agile software development: a possible solution to requirements engineering challenges in agile projects	Se propuso un enfoque para utilizar modelos conceptuales en proyectos respetando los principios ágiles, centrándose en los requisitos como historias de usuarios para el desarrollo de software con métodos ágiles.	Modelos conceptuales

3.2.2.1 Discusión de la literatura

1. A Method to identify relevant genome data: conceptual modeling for the medicine of precision (Palacio et al., 2018)

Resumen: En este trabajo se presenta un método sistemático basado en técnicas de modelado conceptual y gestión de la calidad de los datos de la calidad con el objetivo de ayudar al diagnóstico genómico de la enfermedad de Crohn. Para sistematizar el proceso de curación de variantes se propone dos fases: (i) Selección de fuentes de datos fiables e (ii) identificación de las variantes de ADN.

Una vez fijado el modelo conceptual, se agrupa los atributos en tres categorías: "obligatorios", "recomendados" y "otros".

Modelo utilizado: Modelo conceptual de clases

Resultados: En este trabajo, se presentó un método sistemático por medio de un modelo conceptual para identificar la información relevante para facilitar la curación de variantes con el fin de proporcionar un diagnóstico genómico preciso.

Conclusiones: El modelado conceptual sirvió para extraer información de los repositorios públicos, esto facilitó el proceso de curación de variantes debido a que el proceso se ve dificultado por su heterogeneidad y dispersión.

2. Towards an Ontology for Tertiary Bioinformatics Research Process (Crovari et al., 2020a)

Resumen: El trabajo muestra cómo puede aprovecharse la creación de una ontología para diseñar interfaces que no se centren únicamente en los datos, también considerar tanto el flujo de trabajo de la investigación como los requisitos de los investigadores.

Modelo utilizado: Ontologías

Resultados: Se creó un árbol jerárquico de tareas como ontología para describir el proceso de investigación terciaria bioinformática a partir de los resultados de los estudios de usuarios.

Conclusiones: La representación del proceso de investigación en forma de árbol, logra analizarse en diferentes niveles de abstracción. Un modelo de este tipo permite a los diseñadores de herramientas tener una visión completa del proceso, viendo las herramientas no como piezas de software autónomas sino como parte de un proceso más amplio.

3. Conceptual modelling for life sciences based on systemist foundations (Lukyanenko et al., 2022)

Resumen: El trabajo propone unir dos enfoques para lograr una visión más completa, vinculando una capa de conceptos, que describa los elementos del genoma y sus conexiones conceptuales, con una capa de datos, que describa los conjuntos de datos derivados de la secuenciación del genoma con tecnologías específicas.

Modelo utilizado: Diagrama de composición

Resultados: El diagrama mostró cómo esta extensión semántica puede hacer que el diseño y la gestión de sistemas de información en ciencias de la vida sean más precisos y completos.

Conclusiones: Se desarrolló el diagrama con ejemplos basados en el genoma humano y enfermedades víricas, utilizando dos modelos, un modelo de clases que representa los elementos del genoma y un Pathways Analysis Diagram (PAD) en el que se representa la descripción funcional de los diferentes procesos moleculares. Así obteniendo un Diagrama de composición.

4. Using Conceptual Models in Agile Software Development: A Possible Solution to Requirements Engineering Challenges in Agile Projects (Gupta et al., 2022)

Resumen: Se propone un enfoque para utilizar modelos conceptuales en proyectos respetando los principios ágiles, centrándose en los requisitos como historias de usuarios para el desarrollo de software con métodos ágiles.

Modelo utilizado: Modelo de procesos basado en historias de usuarios, Modelo de Caso de Uso y Modelo de Dominio basado en las historias de usuario.

Resultados: El trabajo demuestra que se pueden utilizar eficazmente modelos conceptuales en proyectos que utilizan la metodología ágil, sin embargo, se deben cumplir varias condiciones a partir de un conjunto determinado de historias de usuarios.

Conclusiones: Los modelos conceptuales apoyaron el análisis de requisitos y el diseño de sistemas. Las representaciones visuales facilitaron la comunicación entre los miembros del equipo del proyecto y ayudando a crear una comprensión compartida de cómo un sistema debe soportar un dominio.

Capítulo 4 Desarrollo del modelo del proceso de anotación del gen V

4.1 Marco metodológico

La metodología y los trabajos derivados que resultaron durante el transcurso del proyecto, se desarrollaron y validaron en conjunto con un equipo de expertos en biología molecular del Centro de Investigación Sobre Enfermedades Infecciosas (CISEI) del Instituto Nacional de Salud Pública (INSP).

Dada la naturaleza de la problemática abordada y el enfoque multidisciplinario de este proyecto, se seleccionó la metodología Ciencia del Diseño. Este enfoque de investigación propone la construcción de artefactos para brindar una solución útil y efectiva a un problema de un dominio dado.

La Ciencia del Diseño consta de tres ciclos:

- Ciclo de relevancia: Principalmente, se identifica el problema y su posible solución
- Ciclo de rigor: Se basa en una búsqueda de información relevante al problema.
- Ciclo de diseño: Se finaliza diseñando y desarrollando la solución por medio de un artefacto

La metodología de solución consta de 5 fases, mismas que integran los tres ciclos de la Ciencia del Diseño. Es decir, ciclo de relevancia FASE 1 y FASE 2. Ciclo de rigor FASE 3, ciclo de diseño FASE 4 y FASE 5.

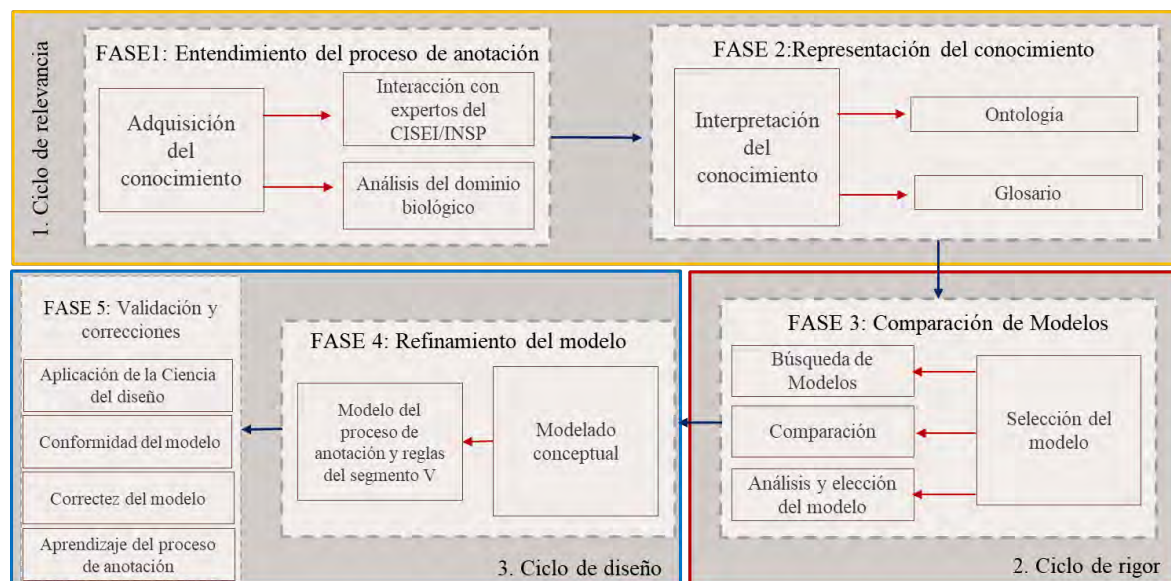


Figura 5 Metodología de solución (Ciencia del Diseño)

FASE 1: Entendimiento del proceso de anotación

Esta fase se constituye a partir de la adquisición del conocimiento, la interpretación de los datos y conceptos biológicos por medio de interacciones con los expertos en biología molecular de CISEI/INSP. Se realizó por medio de reuniones y sesiones con los expertos, quienes compartieron sus conocimientos del dominio. Adicionalmente, colaboramos en las actividades que desempeñan los biólogos moleculares del CISEI/INSP para realizar el proceso de anotación en una determinada especie.

FASE 2: Representación del conocimiento

A partir de la interacción con los expertos y de manera conjunta, se construyeron ontologías, glosarios y diagramas para representar el proceso de anotación con información genética los quirópteros. Estos artefactos se estuvieron actualizando y validando de manera continua a medida del avance en esta investigación.

Fase 3: Comparación de modelos

Esta fase se divide en dos secciones importantes. La primera sección consiste en una búsqueda de los diferentes modelos que permitan representar el proceso de anotación del gen V. En la segunda sección se realizó una comparación de los diferentes modelos y se eligió el más adecuado para representar el proceso de anotación del gen V.

FASE 4: Desarrollo del modelo

Se desarrolló el modelo conceptual del proceso de anotación del gen V con base en los modelos elegidos en la fase 3.

FASE 5: Validación y correcciones

La fase cinco es un proceso iterativo en el cual se validaron y se realizaron correcciones conforme a las fases mencionadas anteriormente. Además de una validación empírica conforme a los expertos del CISEI/INSP.

4.2 Desarrollo del modelo basado en la metodología ciencia del diseño

La metodología Ciencia del Diseño es un enfoque de investigación que propone la construcción de cuatro artefactos para brindar una solución útil y efectiva a un problema de un dominio dado, mediante ciclos.

Existen diversos trabajos de la Ciencia del Diseño. De acuerdo con (Offermann & Levina, 2009) se estructura en tres fases: 1) Identificación del problema, 2) Diseño de la solución y 3) Evaluación.

Otro autor (Johannesson & Perjons, 2014) menciona siete guías para la aplicación de la ciencia del diseño: 1) El diseño como un artefacto; 2) Relevancia del problema; 3) Evaluación del diseño; 4) Contribuciones de la investigación; 5) Rigor de la investigación; 6) El diseño como un proceso de búsqueda; y 7) Comunicación de la investigación. Conforme a (Hevner, 2007), la aplicación del enfoque de Ciencia del Diseño consiste en tres ciclos de investigación, denominados ciclo de Relevancia, de Diseño y de Rigor.

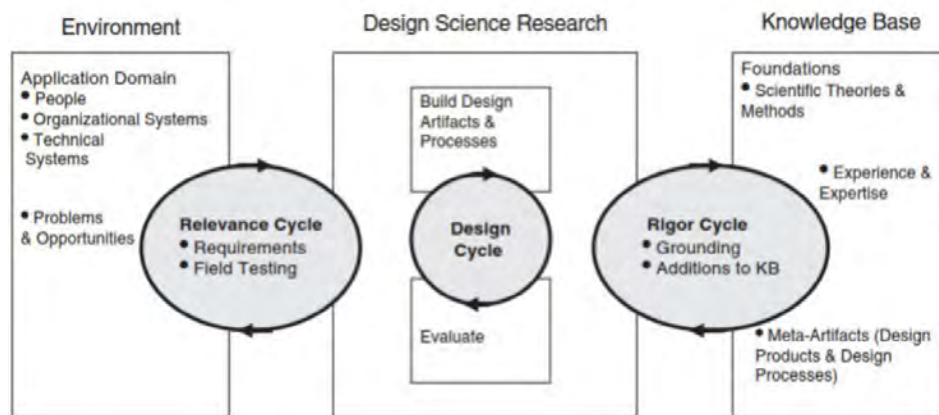


Figura 6 Ciclos de Ciencia del Diseño (Hevner & Chatterjee, 2010)

El objetivo principal de la Ciencia del Diseño es dar soporte al diseño y construcción de artefactos útiles e innovadores para resolver problemas de dominio. La Ciencia del Diseño permite abordar problemas de difícil resolución debido a su complejidad, a que están incompletos, son contradictorios o presentan requisitos cambiantes que a menudo, son difíciles de relevar. Además de requerir un trabajo multidisciplinario para la adecuada comprensión del problema y para el desarrollo de su solución.

En este proyecto se aplicó el enfoque (Hevner, 2007) por el orden que propone conforme a sus tres ciclos. 1) El entendimiento del dominio de manera profunda, 2) Adquisición del conocimiento a partir de las posibles soluciones que han utilizado para resolver este problema, 3) Hasta llegar a la construcción del artefacto adecuado para la solución.

Este enfoque aplica la Ciencia del Diseño por medio de tres ciclos que constan de diversas actividades para su correcta aplicación.

1. Ciclo de relevancia:

- A. Definir el problema
- B. Especificar los requisitos

2. Ciclo de rigor

- C. Plantear las preguntas de investigación
- D. Investigar soluciones actuales
- E. Identificar la relevancia

3. Ciclo de diseño

- F. Diseñar y desarrollar el artefacto
- G. Obtener retroalimentación con expertos

4.2.1 Ciclo de relevancia

A. Definir el problema

El ciclo de relevancia se centra en entender el dominio, detectando problemas y oportunidades de mejora. Este proyecto se enfoca en el proceso de anotación que realizan los biólogos moleculares (expertos) del Centro de Investigación sobre enfermedades infecciosas (CISEI) del Instituto Nacional de Salud Pública (INSP).

Una de las principales actividades del CISEI es realizar un proceso denominado “proceso de anotación”. En este proceso se identifican los genes funcionales y no funcionales para formar anticuerpos en secuencias de ADN de diversas especies, con base a las reglas y elementos definidos en biología molecular. Las actividades de este proceso se realizan de manera manual con apoyo de varios sistemas de software. Se trata de un trabajo que requiere de tiempo, precisión y esfuerzo. Los biólogos pueden llegar a tardar hasta un año para la identificación y clasificación de los genes V, D y J de un solo individuo. Por razones de tiempo este proyecto se enfocó en identificar y clasificar un solo gen. Se seleccionó el gen V por ser uno de los más complejos. El gen V es el que contiene una mayor cantidad de elementos y reglas para su formación, además de diferentes causas para determinar su no funcionalidad.

Para abordar el ciclo de relevancia, se formó un equipo de trabajo conformado por un biólogo molecular, un bioinformático y tres informáticos encargados de realizar las diversas actividades relacionadas con el entendimiento del problema.

- **Curso de inmunología:** El curso fue impartido en el CISEI/ INSP por el Dr. Jesús Martínez Barnetche. Tuvo una duración de 30 días con un total de 60 hrs efectivas. Participaron estudiantes del CISEI/INSP, UAEM y CENIDET con la finalidad de conocer los aspectos claves de la biología molecular enfocados a los genes V(D)J.
- **Capacitación para la anotación:** Posteriormente se realizó una capacitación, donde participaron cuatro estudiantes en biología en el proceso de anotación. Los cursos ayudaron a conocer y familiarizarse acerca de cómo realizar el proceso de anotación, centrándose en la identificación del gen V, sus elementos y sus reglas.

- **Revisión de la literatura en el ámbito genómico:** Para ampliar el conocimiento del gen V, se investigó de manera independiente el ámbito genómico y varios trabajos desarrollados en el área bioinformática. Algunos de los más relevantes fueron (Davis, 2004; Garcia Mariscal et al., 2013; Julia María Leshner Gordillo & Carlos Alfonso Tovilla Zárate, 2013; Martínez-Frías, 2010; Schroeder & Cavacini, 2010; Sirupurapu et al., 2022).
- **Acompañamiento del proceso de anotación:** Uno de los principales retos fue el entender cómo los expertos ubican dentro del genoma de una especie al gen V. Con la finalidad de entender este proceso se realizaron diversos ejercicios con el software IGV (*Integrative Genomics Viewer*) que los expertos utilizan para buscar y clasificar los genes.
- **Entrevistas estructuradas para resolver dudas:** Periódicamente se realizaron entrevistas con los expertos para resolver las dudas que fueran surgiendo a lo largo de la investigación. Principalmente enfocadas al proceso de anotación, las reglas del gen V y los conceptos que iban surgiendo en el transcurso.
- **Presentación por informáticos:** Se comenzó a representar el conocimiento adquirido en ontologías para entender la relación entre los conceptos y se elaboró un glosario para estandarizar los conceptos a utilizar a lo largo del proyecto. Se validó por parte de los expertos que todo se represente correctamente.

En este punto se tenía una idea del artefacto que se utilizaría para representar el proceso de anotación del gen V. De acuerdo con (Tebes et al., 2019), un proceso que no presenta sus actividades modeladas es difícil de comprender, debido a que puede generar cierta ambigüedad en las descripciones de las actividades y es difícil también de comunicar, por lo cual se propuso realizar un modelo conceptual. Un modelo conceptual permite representar problemas complejos que requieren un alto nivel de abstracción, obteniendo vistas del dominio con el objetivo de facilitar la comprensión del problema.

B. Especificar los requisitos

El modelo conceptual tenía que ser capaz de representar el proceso de anotación, así como los elementos y reglas que se requieren para identificar un gen V funcional y no funcional. Para esto, se definieron los siguientes requisitos:

- Debe ser un modelo conceptual que permita representar procesos
- Debe ser capaz de representar cada una de las actividades del proceso de anotación del gen V que realizan los biólogos moleculares con el software IGV
- Debe representar las reglas que se requieren para formar un gen V
- Debe representar las partes que forman un gen V
- Debe representar la funcionalidad de un gen V
- Debe ser un modelo conceptual claro y fácil de seguir para los biólogos moleculares
- Debe ser validado y aprobado por los biólogos moleculares
- Debe representar correctamente el proceso de anotación de acuerdo con los biólogos moleculares del CISEI/INSP

- Debe ser un modelo computable

El ciclo de relevancia permitió entender el dominio del problema y definir una posible solución, en este caso un modelo conceptual. Se continúa con el siguiente ciclo.

4.2.2 Ciclo de rigor

En el ciclo de rigor se debe investigar a fondo y consultar el estado del arte para garantizar que los diseños producidos sean aportaciones a la investigación y no diseños rutinarios.

C. Plantear las preguntas de investigación

El ciclo de rigor busca dentro del estado del arte otras soluciones que se han realizado para resolver la problemática. Se plantean preguntas de investigación para acotar la búsqueda. La pregunta que se planteó para este proyecto es la siguiente: ¿Cómo se ha modelado conceptualmente el proceso de anotación y sus reglas?

D. Investigar soluciones actuales

Se realizó una búsqueda exhaustiva en el estado del arte, donde se encontraron diversos trabajos relacionados al modelo conceptual del genoma humano. De estos trabajos se seleccionaron 16 artículos que proponen diversas conceptualizaciones del genoma humano y sus relaciones. Todos estos trabajos representan el genoma por medio de diagramas de clases.

En ninguno de los trabajos seleccionados encontramos una representación del proceso de anotación, sin embargo, fueron de utilidad para entender cómo modelar el ámbito genómico. En la tabla 1 se muestran los cinco trabajos más relevantes para nuestro proyecto. En ellos podemos ver cómo se modelan la relación y la función que tienen las proteínas, anticuerpos e incluso segmentos V(D)J dentro del genoma. Principalmente los modelos conceptuales desarrollados muestran una vista general del genoma, ayudando a comprender los conceptos y relaciones de manera biológica, sin embargo, no representan el proceso de anotación de genes.

Tabla 4 Artículos relacionados con el modelado conceptual del genoma

NO.	REF	AÑO	ARTÍCULO	DESCRIPCIÓN GENERAL	MODELO EMPLEADO
1	(Palacio et al., 2018)	2018	Un método para identificar datos genómicos relevantes: modelo conceptual para la medicina de precisión	Se realizó un método sistemático basado en técnicas de modelado conceptual y gestión de la calidad de los datos con el fin de ayudar al diagnóstico genómico de una enfermedad	Modelo conceptual de clases
2	(García S et al., 2020)	2020	Hacia la comprensión del genoma humano: Un enfoque holístico del modelo conceptual	Se actualizó el modelo conceptual del genoma humano, proporcionando un enfoque más flexible, ampliando interacciones entre las diferentes partes, agregando nueva información de las proteínas	Modelo conceptual de clases
3	(Leon & Pastor, 2021)	2021	Hacia una comprensión común de las proteínas y sus interacciones basada en modelos conceptuales	Se realizó un modelo conceptual centrándose en las proteínas, identificando algunos problemas y debilidades que dificultan la comprensión y representación correcta de los conceptos importantes en las proteínas	Modelo conceptual de clases
4	(García S et al., 2021)	2021	Un enfoque basado en modelos conceptuales para mejorar la representación y gestión de datos ómicos en medicina de precisión	Se desarrolló un esquema conceptual que permitió superar limitaciones que se presentan en la medicina de precisión, proporcionando una mejor representación de los datos proteómicos y facilita su integración	Modelo conceptual de clases
5	(Sirupurapu et al., 2022)	2022	Un enfoque global para el modelado conceptual de datos genómicos	Se integraron dos modelos existentes para lograr una representación más completa y holística de la información genómica, permitiendo establecer vínculos entre los conceptos y los datos genómicos.	Modelo conceptual de clases

Posteriormente se realizó otra revisión al estado del arte, centrándose en cómo se ha representado el proceso de anotación de manera general, sin ninguna palabra clave “modelo conceptual”. En esta revisión se encontraron los siguientes trabajos:

Tabla 5 Artículos representando el proceso de anotación

NO.	REF	AÑO	ARTÍCULO	TEMA	REPRESENTACIÓN
1	(Kitts & Box, n.d.)	2012	Proceso de ensamblaje y anotación del genoma	Muestra cómo están organizadas las principales etapas del proceso de ensamblaje y anotación del genoma humano.	Diagrama de Flujo
2	(Ejigu & Jung, 2020)	2020	Revisión sobre la anotación computacional del genoma de secuencias obtenidas mediante secuenciación de próxima generación	Todo el flujo de trabajo de anotación después del ensamblaje de la secuencia	Diagrama de Flujo
3	(Crovari et al., 2020b)	2020	Hacia una ontología del proceso terciario de investigación bioinformática	Muestra las tareas que describen el proceso de investigación terciaria bioinformática	Árbol jerárquico (ontología)
4	(Victoria María Posner, 2022)	2022	Uso de datos de transcriptómica para la anotación de Genes en el genoma de pacú (Piaractus mesopotamicus)	Se realizó un análisis de calidad de las secuencias crudas y luego se ensamblaron de manera conjunta usando el programa Trinity.	Descripción del código

El proceso de anotación se ha representado de diferentes maneras, por ejemplo, en (Kitts & Box, n.d.) y (Ejigu & Jung, 2020) por medio de diagramas de flujo donde muestran el proceso de manera general que lleva desde la obtención de datos del genoma hasta el procesamiento de estos datos para encontrar los genes. También en (Crovari et al., 2020a) se representa mediante un árbol jerárquico, como una ontología, para mostrar cómo los biólogos moleculares clasifican los elementos según su nivel de abstracción y las tareas que realizaban. Otro trabajo (Victoria María Posner, 2022) muestra el proceso de anotación por medio de un pseudocódigo, mostrando cómo anotar el ARNm y la extracción de los genes. Este proceso es muy similar al primer proceso que realizan los expertos del CISEI/INSP para identificar dentro del genoma los genes que desean clasificar y nosotros representamos con un diagrama SADT (fig. 12).

E. Identificar la relevancia

En esta búsqueda se concluyó que los trabajos publicados solo representan las etapas previas al proceso de anotación, pero no describen las reglas que siguen los biólogos moleculares para realizarla, y en particular no encontramos trabajos de anotación del gen V. Conforme a esto, se identificó que existe un área de oportunidad para el desarrollo de un modelo conceptual que represente este dominio.

4.2.3 Ciclo de diseño

F) Diseñar y desarrollar el artefacto

Una vez identificado el problema, se continuó con el desarrollo de la solución y posteriormente se evaluó para demostrar que resuelve la problemática planteada.

En este ciclo no solamente se desarrolló el modelo conceptual entregable, sino que además se realizaron otros modelos fundamentales para el entendimiento del problema que ayudaron a representar el proceso de manera adecuada.

Se desarrolló una ontología como primera representación, principalmente para conocer los conceptos y relaciones importantes que intervienen en el proceso de anotación.

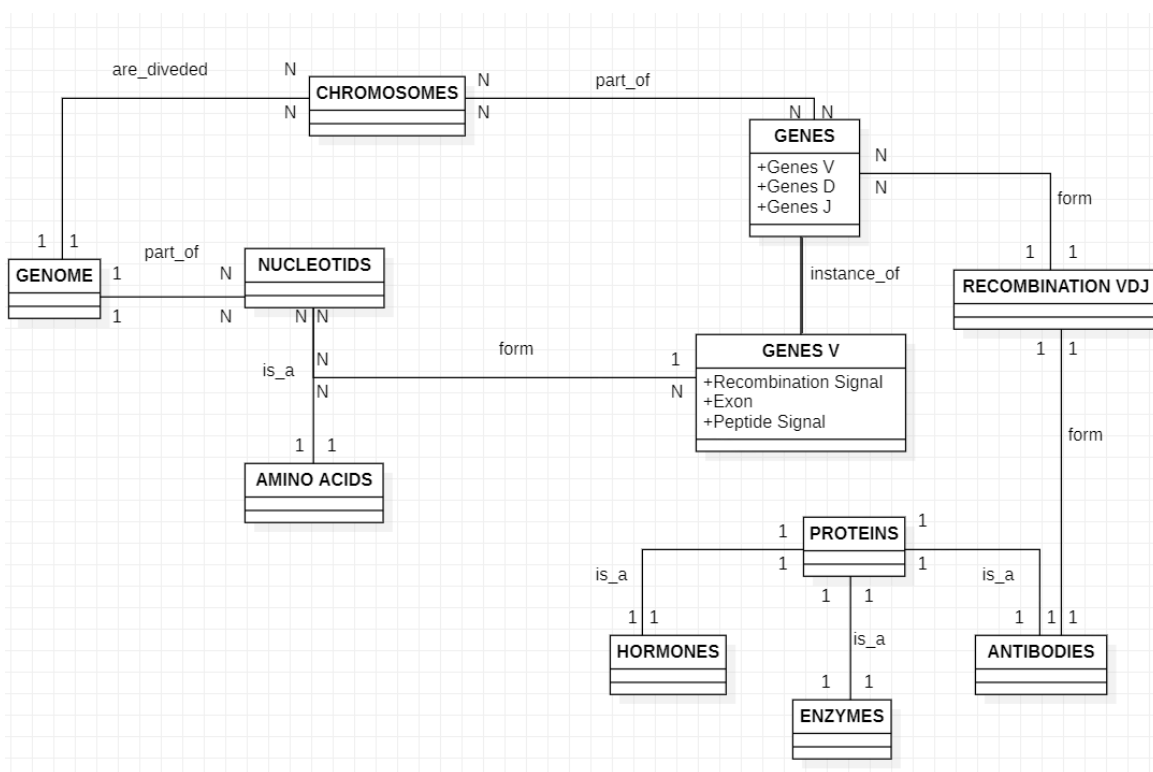


Figura 7 Ontología

4.2.3.1 Selección de modelos conceptuales para la representación del gen V

En este ciclo se buscó principalmente identificar modelos conceptuales capaces de representar procesos. Se incluyó el diagrama de clases por ser uno de los más utilizados en el estado del arte para representar el genoma.

4.2.3.1.1 Diagrama de Clases

El diagrama de clases describe la estructura de un sistema en el cual se muestran sus clases, atributos y relaciones entre ellos. Estos diagramas se utilizan normalmente durante el proceso de análisis y diseño de los sistemas, donde se crea el diseño conceptual de la información que se planea manejar en el sistema, así como los componentes que interactúan en él para su correcta funcionalidad y su relación entre uno y otro (O. Pastor et al., 2012)

En el diagrama de clases se identifican los siguientes elementos:

- Clase: La clase es la agrupación de un conjunto de propiedades y características comunes que presentan elementos del dominio.
 - Parte superior: Contiene el nombre de la clase
 - Parte intermedia: Contiene los atributos que caracterizan a la clase, los atributos pueden representarse solo mostrando su nombre.
- Asociación: Representa un tipo de relación bidireccional entre dos o más conceptos relevantes del dominio.
- Relación: Es una característica que vincula conceptualmente varias clases distintas.
- Composición: Se trata de una asociación especial que representa el concepto ser parte de, una parte solo puede pertenecer a un todo, una parte no puede existir sin pertenecer a un todo, la eliminación del todo implica la eliminación de todas sus partes.

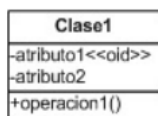


Figura 8 Ejemplo de clase

4.2.3.1.2 Diagrama de actividades

Los diagramas de actividades (DA) son parte de los diagramas de comportamiento UML. Es una herramienta clave para el diseño y análisis de sistemas. Permite la comprensión de flujos de trabajo y permite modelar actividades condicionales y paralelas. Se utiliza para representar actividades ejecutadas por actores diferentes, lo que permite una visión clara y organizada del sistema (Percy Edwin De la Cruz Vélez de Villa et al., 2010).

Un diagrama de actividades está compuesto por los siguientes elementos:

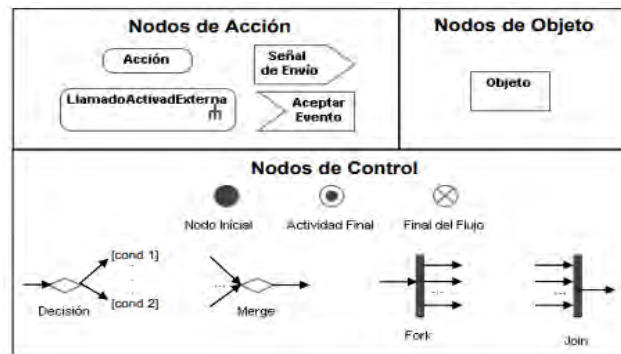


Figura 9 Elementos de los diagramas de actividades en UML 2.0 (Percy Edwin De la Cruz Vélez de Villa et al., 2010)

4.2.3.1.3 Diagramas BPMN

BPMN (*Business Process Model and Notation*) es un lenguaje de notación gráfica, ampliamente aceptado para el modelado de procesos de negocio, creado en el 2001 por BPMI (*Business Process Management Initiative*) desde el 2005. BPMN es mantenido por el OMG (*Object Management Group*). BPMN utiliza un conjunto de elementos gráficos especializados para describir un proceso y de qué manera es realizado, de modo que puedan comunicar una amplia variedad de información (Stephen A. White et al., 2009).

BPMN se puede dividir en tres categorías:

- Orquestación: Un proceso de orquestación describe cómo una única entidad de negocio lleva a cabo las cosas.
- Coreografía: Un proceso de coreografía es una clase de contrato, procedimiento o protocolo entre los participantes que interactúan, pueden ser roles o entidades específicas.
- Colaboración: Un proceso de colaboración es cualquier BPMN que contenga dos o más participantes. Los pools tienen flujo de mensaje entre ellos.

En la fig.10 se muestran algunos de los elementos que contiene BPMN

Nombre	Símbolo	Nombre	Símbolo	Nombre	Símbolo
Actividad		Flujo de secuencias		Asociación	
Subproceso colapsado		Actividad en bucle		Mensaje	
Subproceso ampliado		Anotación de texto (adjunta con una Asociación)		Unión	
Puerta de enlace		Piscinas		Bifurcación	
Tipos de control de puerta de enlace		Caniles		Objeto de datos	

Figura 10 Elementos BPMN (Object Management Group, 2013)

4.2.3.1.4 Modelado SADT

El modelo SADT (Structured Analysis and Design Technique), permite representar las actividades de un proceso, definir las dependencias y relaciones entre dichas actividades, así como los controles que determinan o limitan su ejecución, los mecanismos que los ponen en marcha, así como los datos que se utilizan, comparten o transforman en los procesos (Fahim Ahmed et al., 2014). Su estilo genérico y sus notaciones la hacen apta para su uso en diversos ámbitos problemáticos.

Los diagramas SADT incorporan los procesos de la organización en orden secuencial, de acuerdo con su lógica de ejecución mediante una numeración que se refleja en la esquina inferior de cada actividad. Esto refleja el nivel de influencia de una actividad sobre el resto del proceso.

Los flujos que interconectan actividades se clasifican en cuatro tipos de acuerdo con su significado:

- Entrada: Se refiere a la información que se utilizara para producir las salidas de la actividad, y su flecha va de lado izquierdo.
- Salida: Se trata de información que se produce en la actividad, y su flecha corresponde al lado derecho.
- Control: Se trata de restricciones que afectan a una actividad, y su flecha se agrega por la parte superior.
- Mecanismo: Normalmente se refiere a máquinas, personas, recursos o sistemas existentes que ejecutan la actividad, y su flecha se agrega por la parte inferior

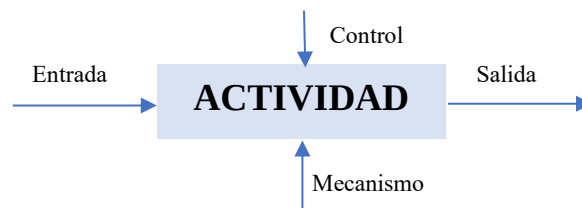


Figura 11 Elementos del modelo SADT

4.2.3.2 Comparación de modelos

Se realizó un cuadro comparativo de los modelos descritos anteriormente

Tabla 6 Comparación de modelos conceptuales que representan procesos

Modelo Conceptual	Concepto	Enfoque	Primitivas
Clases	Describe la estructura de un sistema	El proceso de análisis y diseño de los sistemas	Conceptos, atributos y relaciones
Actividades	Identifican las acciones que se realizan durante la ejecución de un proceso o actividad.	Describir flujos de procesos y tareas de software	Objetos, acciones, decisiones, entre otros
SADT	Permite representar las actividades de un proceso	Procesos de la organización en orden secuencial	Entrada, salida, control, mecanismo
BPMN	Captura la vista de un proceso, incluidos sus participantes, interacciones y el flujo de datos e información.	Identifica y controla los procesos manuales como automatizados	Actividades, eventos, compuertas, entre otros

Al tener en cuenta las aplicaciones y características de cada modelo, se identificó la aportación que podrían brindar algunos los modelos seleccionados para la representación del proceso de anotación. Se optó por utilizar el modelo SADT como una primera representación para entender el proceso general que siguen los biólogos moleculares. Este modelo representa desde la extracción del genoma de una especie, la transformación de los datos obtenidos, hasta la anotación del genoma en el software IGV. El software IGV permite visualizar secciones del genoma con la finalidad de ubicar, clasificar y realizar anotaciones de determinados genes de interés de los biólogos moleculares. El modelo conceptual SADT, además de representar el proceso general, ayudó a ubicar nuestra aportación en este proceso (fig. 12). El modelo conceptual del proceso de anotación del gen V que se desarrolla más adelante corresponde a las actividades 1.4 y 1.5.

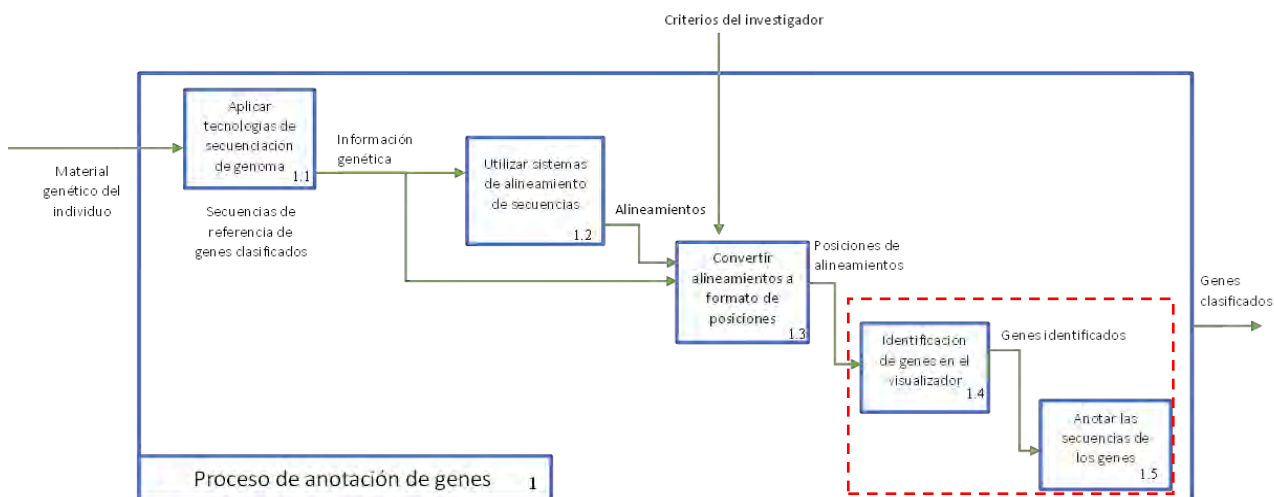


Figura 12 Diagrama SADT del proceso de anotación general

Una vez identificado donde se ubica nuestra aportación, se realizó un diagrama de actividades para representar el proceso de anotación del gen V. Se eligió el diagrama de actividades por su similitud con el diagrama de flujo, además contiene una semántica más fácil de seguir para los expertos.

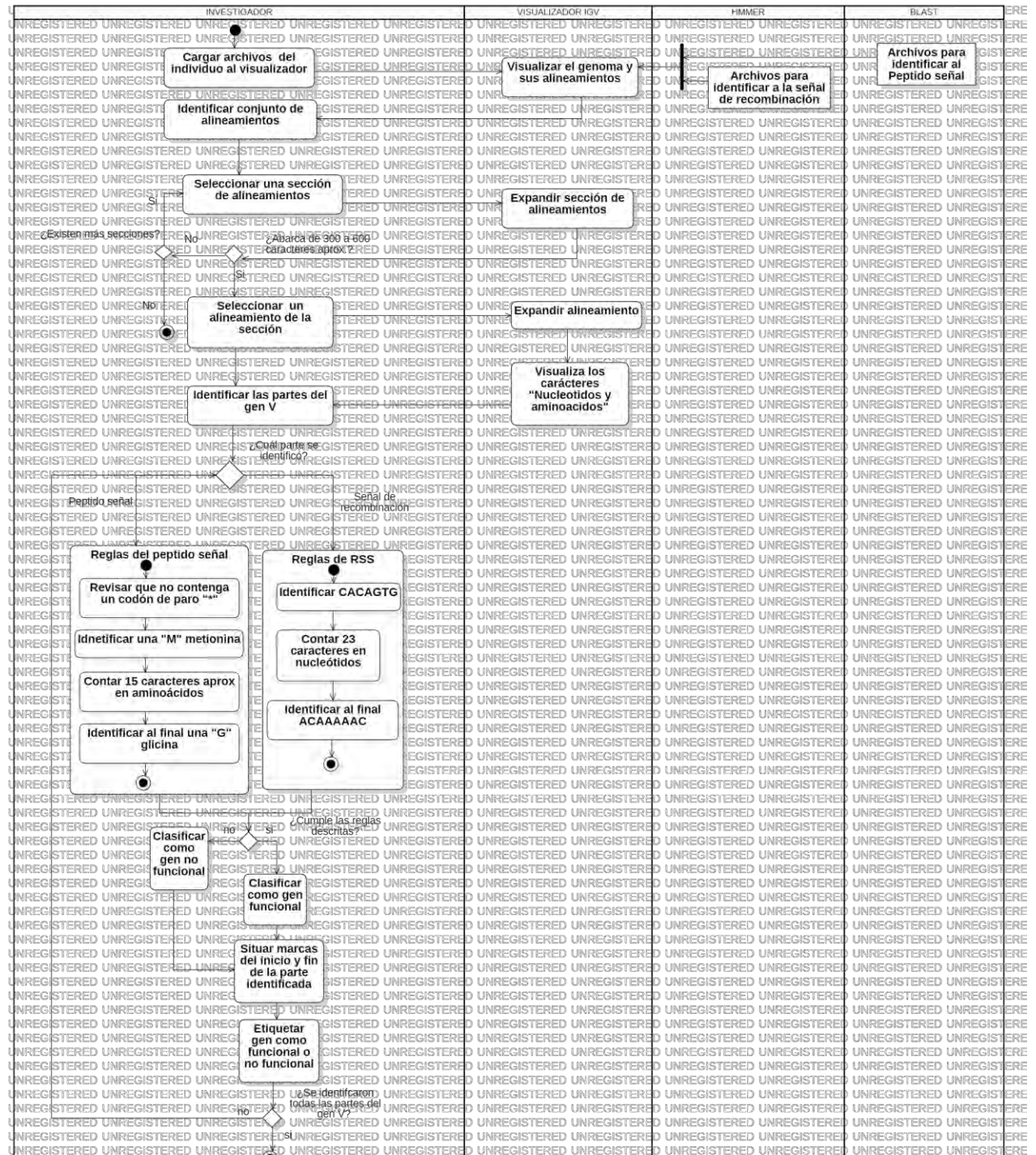


Figura 13 Diagrama de actividades v1 del proceso de anotación de los genes V

Para continuar con la representación se realizó una comparación más específica entre el diagrama de actividades y BPMN (tabla 7) para determinar qué modelo podría representar de manera más completa la complejidad del problema, sin dificultar su comprensión por parte de los biólogos encargados del proceso de anotación.

Tabla 7 Comparación de diagrama de actividades y BPMN

Aspecto	Diagrama de actividades	BPMN
<i>Objetivo</i>	Modelado de flujos de trabajo del sistema y procesos internos	Modelo de procesos de negocio de un extremo a otro
<i>Elementos</i>	Actividades, acciones, flujos de control, nodos de decisión	Actividades, eventos, puertas de enlace, grupos, carriles, objetos de datos
<i>Notación</i>	Rectángulos (actividades), rombos (decisiones), flechas (flujo de control), carriles	Círculo (eventos), rectángulo con esquinas redondeadas (actividades), rombos (puertas de entrada), piscinas, carriles
<i>Flexibilidad</i>	Versátil, adecuado para procesos detallados y de alto nivel, flujo de secuencias y paralelos	Diseñado para modelar procesos complejos, proporciona construcciones avanzadas para escenarios complejos
<i>Alcance</i>	De uso general, se puede utilizar tanto para el modelado de sistemas como de procesos de negocio	Específico para el modelado de procesos de un negocio, captura interacciones y flujo de datos
<i>Complejidad</i>	Más simple, ideal para representar acciones dentro de un sistema	Más sofisticado, adecuado para procesos complejos
<i>Colaboración</i>	Puede representar colaboraciones, pero con menos granularidad	Proporciona construcciones específicas para representar colaboraciones entre participantes u organizaciones
<i>Flujo de datos</i>	Se centra en representar el flujo de control	Proporciona un mejor soporte para modelar datos y flujo de información dentro de los procesos

Si bien el diagrama de actividades y el BPMN tienen el mismo enfoque de representar un proceso. El diagrama de actividad presentaba diversas limitantes para poder representar todo el proceso que realizan los biólogos moleculares para anotar los genes V y sus reglas, debido que se requiere de una semántica amplia, capaz de tener elementos adecuados para la representación que se desea, requiere un modelo que englobe subprocesos amplios y principalmente sea fácil de seguir, además que pueda presentar colaboraciones, el diagrama de actividades lo hace, pero con menos granularidad.

Aun así, el diagrama de actividades fue una herramienta de gran utilidad para plasmar una primera versión del proceso. El modelo elegido para representar el proceso de anotación del gen V fue el modelo BPMN (fig.14).

4.2.3.3 Modelo BPMN del proceso de anotación del gen V

En la presente sección se describe el artefacto desarrollado a partir de la interacción con los especialistas y de la revisión del estado del arte. El artefacto desarrollado es un modelo BPMN realizado en el sistema Bizagi. El modelo busca representar el proceso de anotación que realizan los especialistas para los genes V.

La principal función del proceso de anotación es anotar y clasificar los genes de una determinada especie. Identificando las reglas y la funcionalidad de dicho gen.

El modelo desarrollado es un BPMN de colaboración, el cual se compone de dos *pools*:

- **El experto:** es el encargado de realizar el proceso y de interactuar con el sistema para realizar la anotación de los genes V.
- **El software *Integrative Genomics Viewer* (IGV):** es el programa encargado de administrar los diferentes archivos que contienen la información como son los alineamientos de secuencias y el genoma del individuo a analizar. El sistema es el encargado de ejecutar las instrucciones dadas por el experto en este caso: a) delimitar las posiciones del gen y b) asignar una etiqueta al gen V y sus secciones.

En la (fig.14) se muestra el modelo conceptual representando el proceso de anotación del gen V. El modelo se dividió en cuatro secciones con el fin de facilitar la explicación. Cada una representa las diferentes actividades que realiza el especialista y el sistema IGV. La primera sección del modelo inicia desde el punto de vista del experto con la selección de los diferentes archivos necesarios para el proceso de anotación. Estos archivos provienen de los programas *Hmmer*, *blast*, *exonarate*.

- *Hmmer:* Es el encargado de identificar la Señal de Recombinación
- *Blast:* Es el encargado de identificar el Péptido señal
- *Exonarate:* Es el encargado de identificar el Exón

Una vez seleccionados los archivos, el sistema se encarga de cargar los datos y mostrarlos. El sistema IGV muestra los alineamientos del genoma. Posteriormente el experto debe seleccionar uno a uno los diferentes alineamientos a ser analizados.

Con la selección del alineamiento finaliza la primera sección del modelo. La segunda sección comienza con la revisión de la longitud de nucleótidos (caracteres) que se encuentran en el conjunto de alineamientos. El conjunto debe contar con una longitud de aproximadamente 450 a 480 nucleótidos. Esta longitud pertenece a un gen V en la totalidad con todas sus partes. Si no cuenta con esta longitud se pasa a identificar otro conjunto de alineamiento, de no existir otro, se finaliza el proceso.

Si el conjunto de alineamientos abarca la longitud antes mencionada, se continuará con el análisis de los alineamientos. Se debe identificar alguna de las secciones del gen, eligiendo el flanco derecho, izquierdo o la zona central del alineamiento. Según el alineamiento seleccionado se debe realizar una ampliación de la imagen o zoom. El software IGV mostrará

la secuencia de caracteres en nucleótidos y aminoácidos (caracteres). La secuencia que se muestra en el sistema IGV se utiliza para identificar la sección del gen y colocar sus marcas de inicio y fin. Con la ampliación del alineamiento finaliza la segunda sección y comienza la tercera sección del proceso la identificación de las partes del gen y la asignación de marcas de inicio y fin del gen.

Primera sección

Segunda sección

Tercera sección

Cuarta sección

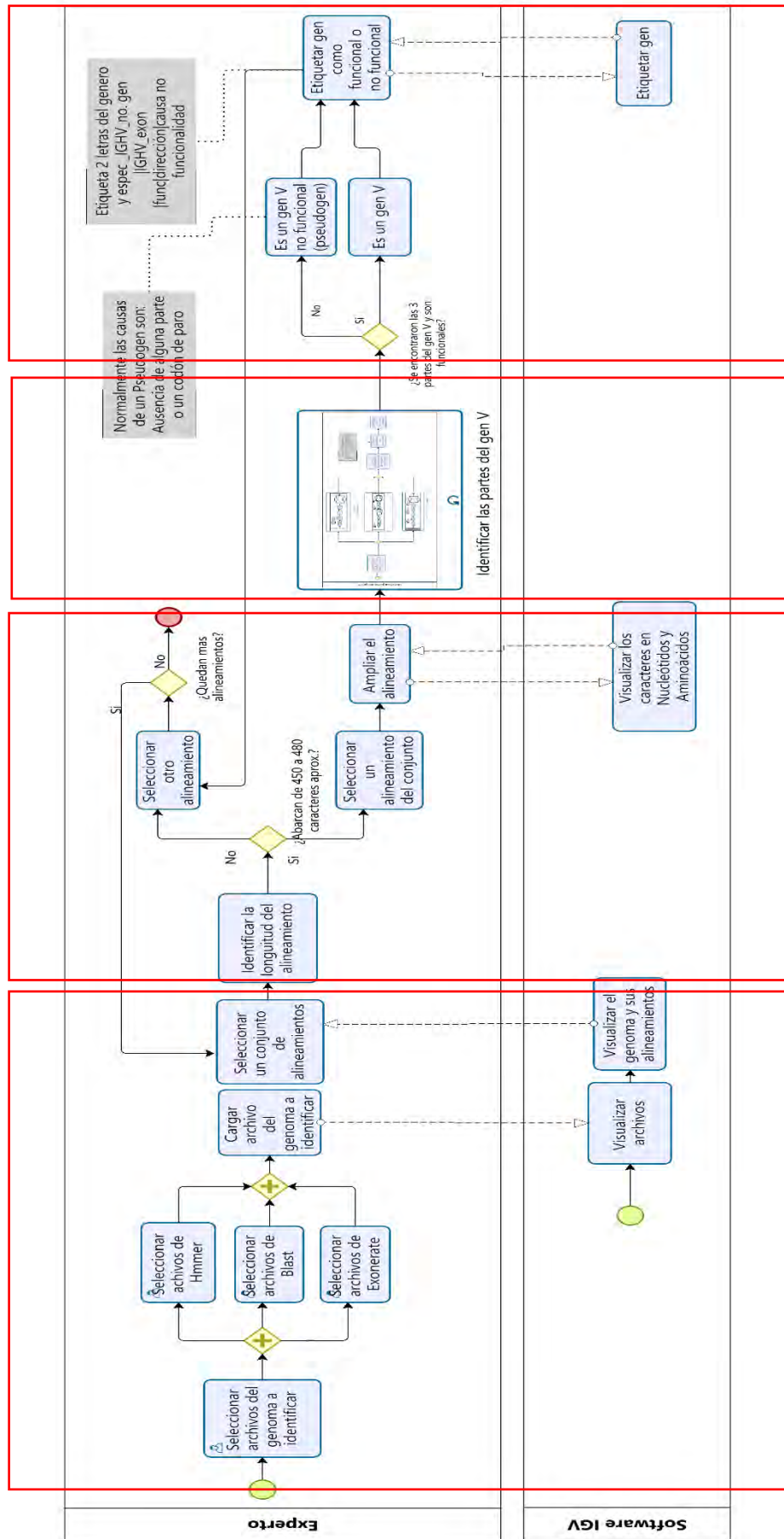


Figura 14 Modelo conceptual BPMN del proceso de anotación de los genes V

Haciendo un acercamiento a la tercera sección del modelo (fig.15), tenemos el subproceso de la identificación de la parte del gen V. En este subproceso se muestran cada una de las reglas posibles para cada una de las secciones del gen V. Lo que busca este subproceso es determinar qué parte se analiza y asignar las coordenadas de inicio y fin. Las reglas que se analizan y que están en función de la sección son las siguientes: a) Exón V, b) Péptido señal o c) Señal de recombinación (fig. 16,17,18).

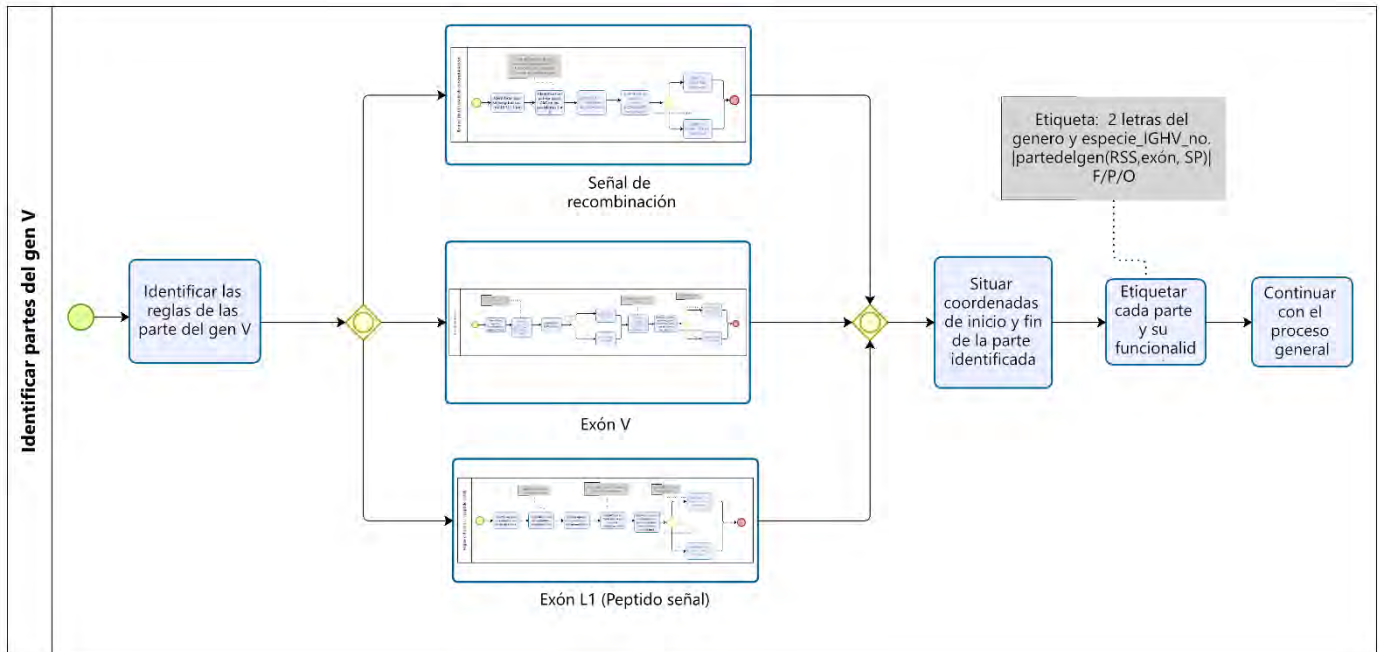


Figura 15 Modelo BPMN del proceso de anotación del gen V (Partes del gen V)

- **Reglas de la señal de recombinación:** Se comienza identificando una longitud de 39 ($\pm$) 1 de caracteres, posteriormente una regla fundamental que debe contener es un patrón similar a CAC, y con este patrón se puede asumir que es una señal de recombinación funcional, sin embargo, se recomienda verificar las demás reglas. Contar 23 caracteres de nucleótidos y finalizar con un nonamero con el patrón ACAAAAACC. Si la secuencia analizada cuenta con las reglas antes mencionadas se procede a determinar que es una RSS funcional, en el caso de que no cumpla con las reglas será una RSS no funcional. Finalizando con las reglas de la RSS se retorna al subproceso de la Figura 15 y se asignan sus coordenadas de inicio y fin. Asignadas sus coordenadas se debe verificar la siguiente sección el Exón.

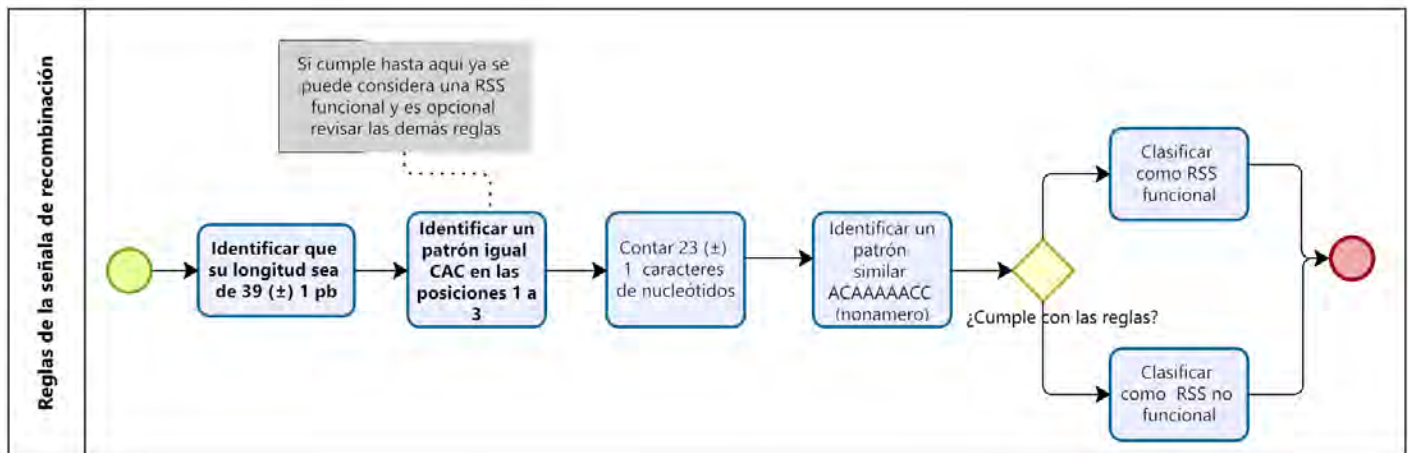


Figura 16 Modelo BPMN del proceso de anotación del gen *V* (Reglas de la señal de recombinación)

- **Reglas de exón V:** Se comienza identificando una longitud de 300 caracteres aproximadamente y solo existen dos reglas esenciales para determinar la funcionalidad del exón. Primero debe contener una Cisteína 104, si no cuenta con ella, es un exón no funcional. Y la segunda no debe contener ningún codón de paro (codón de paro es aquel que no codifica correctamente y contiene el símbolo “*”), si contiene un codón de paro, se clasifica como no funcional.

Para identificar un exón son las siguientes características opcionales debido a que se pueden encontrar o no. Una de ellas es encontrar la señal de *splicing* (es el proceso que se encarga de eliminar algunos intrones que son partes que no codifican nada), donde se encontrará al final un AG y al lado del exón se puede encontrar la señal de recombinación. Una vez finalizando con las reglas del exón se retorna al subproceso de la Figura 15 y se asignan sus coordenadas de inicio y fin. Asignadas sus coordenadas se debe verificar la siguiente sección el exón L1 péptido señal.

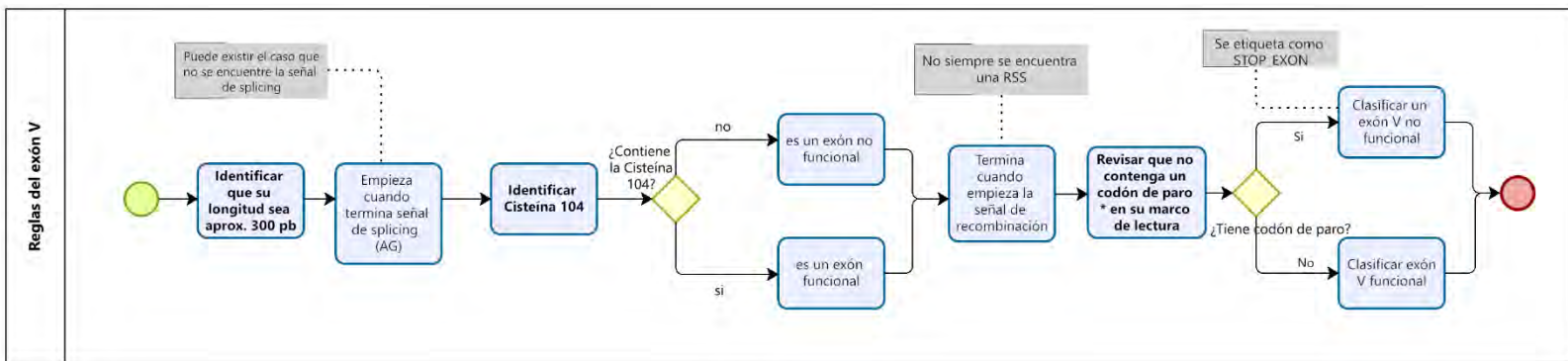


Figura 17 Modelo BPMN del proceso de anotación del gen V (Reglas del exón)

- **Reglas del exón L1 péptido señal:** Se comienza identificando una longitud de 46 caracteres aproximadamente, se debe identificar una M de metionina (aminoácidos) o un ATG (nucleótidos), contar 15 caracteres aproximadamente, y deben finalizar con una “G” de glicina (aminoácidos) o GT (nucleótidos). Normalmente se identifican las reglas en aminoácidos. Para finalizar y determinar su funcionalidad, no debe contener un codón de paro; si contiene uno, se clasifica como no funcional.

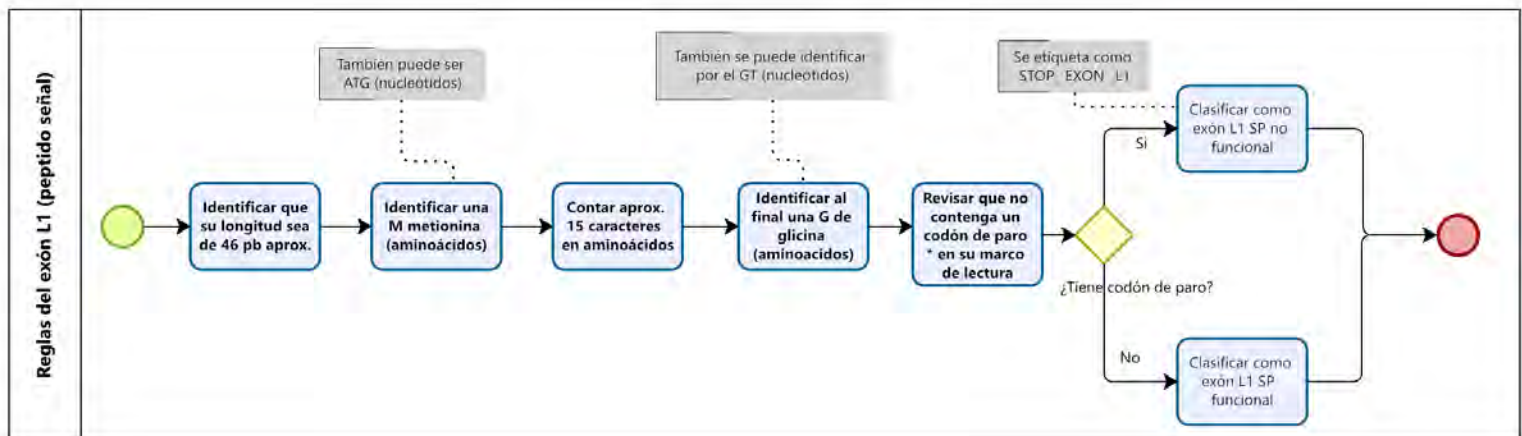


Figura 18 Modelo BPMN del proceso de anotación del gen V (Reglas del exón L1; péptido señal)

Una vez identifique cada parte conforme a sus reglas, y se situaron sus coordenadas de inicio y fin de cada parte identificada. Se continúa con el proceso general de la sección cuatro para determinar si se trata de un pseudogén (no funcional) o un gen V funcional, para determinar esto se debe responder la pregunta: ¿Se encontraron las 3 partes del gen y son funcionales? En esta pregunta se pueden dar los siguientes escenarios:

1. Que falte alguna parte – Se clasifica como no funcional (pseudogén)
2. Que no falte ninguna parte, pero alguna parte sea no funcional – Se clasifica como no funcional (pseudogén)
3. Que no falte ninguna parte y todas las partes sean funcionales – Se clasifica como gen V (funcional)

Cual sea su clasificación, se debe etiquetar al gen encontrado (etiqueta general).

Etiqueta: Especie_IGHV_no. gen |IGHV_exon |func|dirección|causa no funcionalidad

- Especie: Primeras dos letras del género y dos primeras letras de la especie
- No. del gen: Se comienza con 001 y así consecutivamente se vayan encontrando los genes
- Funcionalidad: F (funcional), P(Pseudogen) u O (Falta la señal de recombinación)
- Dirección: FOR (derecha) o REV (izquierda, reversa)
- Causa de no funcionalidad (si es funcional, este espacio queda en blanco): STOP_V, PARTIAL_RSS, STOP_SP, FS_V, NO_CYS, NO_RSS, NO_SPLICE_DONADOR.
- No debe tener acentos, comas, puntos y debe estar bien escrito

Los siguientes ejemplos son etiquetas utilizadas en el proceso de anotación del gen V:

- Péptido señal no funcional: Mymy_IGHV_006|IGHV_exon|P|REV|NO_SP
- Señal de recombinación no funcional
Mymy_IGHV_006|IGHV_exon|P|REV|NO_RSS

Una vez que se clasificó un gen V se pasa a identificar otro alineamiento para encontrar las partes de otro gen V, y seguir el mismo proceso hasta clasificar y etiquetar el gen. Este proceso es cíclico y se repite hasta no tener otro alineamiento y haber clasificado todos los genes V de la especie.

G) Obtener retroalimentación con expertos

La última actividad del ciclo del diseño es la retroalimentación de los expertos. Esta actividad se realizó conforme se fue desarrollando el modelo, asegurando que el conocimiento se represente correctamente. Adicionalmente, en el punto 8. Evaluación de resultados, se realizó una evaluación de 4 fases del modelo conceptual.

Capítulo 5 Evaluación de resultados

Una vez desarrollado el modelo conceptual del proceso de anotación se realizó su evaluación de acuerdo con las siguientes 4 fases:

- Fase 1: Aplicación de la Ciencia del Diseño
- Fase 2: Conformidad del modelo BPMN
- Fase 3: Correctes del modelo por parte de expertos de biología molecular
- Fase 4: Aprendizaje del proceso de anotación con el modelo conceptual

En la fase 3 El Dr. Jesús Martínez Barnette, biólogo molecular, fue el principal experto en la anotación de genes. En conjunto con otros dos investigadores de su equipo de trabajo, participaron en la evaluación de esta fase. En la fase 4, 12 aprendices del proceso de anotación con formación en el área biológica evaluaron el modelo conceptual.

La totalidad de la población que participó en la evaluación pertenece al Centro de Investigación sobre Enfermedades Infecciosas (CISEI) del Instituto Nacional de Salud Pública (INSP).

5.1 Fase 1 Aplicación de la Ciencia del Diseño

Esta fase se evaluó conforme a la aplicación de la metodología Ciencia del Diseño que se utilizó para el desarrollo del modelo conceptual. Cada uno de los ciclos cuenta con diversas actividades, las cuales ayudan para una correcta aplicación de la metodología. La evaluación consta en identificar si cada una de las actividades que se mencionan dentro de la Ciencia del Diseño conforme al autor (Hevner, 2007) se cumplieron o no, como se muestra en la tabla 8.

Tabla 8 Evaluación de la Ciencia del Diseño

CIENCIA DEL DISEÑO	ACTIVIDADES	UBICACIÓN		SE REALIZÓ	
		INCISO	PÁGINA	SI	NO
Ciclo de relevancia	Definir problema	A	Pág.35	x	
	Especificar requisitos de diseño	B	Pág.36	x	
Ciclo de rigor	Plantear preguntas de investigación	C	Pág.37	x	
	Investigar soluciones actuales	D	Pág.37	x	
	Identificar relevancia	E	Pág.39	x	
Ciclo de diseño	Diseñar y desarrollar el artefacto	F	Pág.40	x	
	Obtener retroalimentación con expertos	G	Pág. 54	x	

La tabla 9 muestra las actividades con sus incisos y la ubicación en la que se desarrolló en este documento. Indicando qué actividades se realizaron y cuáles no, teniendo la totalidad de las actividades que propone la metodología de la Ciencia del Diseño, podemos concluir que se aplicó y siguió correctamente la metodología para desarrollar el artefacto en este caso el modelo conceptual del proceso de anotación del gen V.

5.2 Fase 2 Conformidad del modelo BPMN

En esta fase se evaluó que el modelo conceptual BPMN (Business Process Model and Notation) se encuentre correctamente construido de acuerdo con las reglas establecidas para el tipo de modelo utilizado.

Se realizaron dos evaluaciones: La primera evaluó la conformidad con base a la correcta construcción del modelo, respecto a la norma internacional vigente ISO/IEC 19510 del Object Management Group Business Process Model and Notation. Como segunda evaluación del modelo BPMN, se evaluó la correcta construcción de acuerdo con el software Bizagi Modeler que se utilizó para modelar el proceso.

A) Conformidad de acuerdo a la norma ISO/IEC 19510:2013 - Object Management Group business process model and annotation

Esta norma tiene como objetivo proporcionar una notación que sea fácilmente comprensible para todos los usuarios, y que pueda constituir un puente estándar entre el diseño de procesos de negocios y su implementación.

La ISO contiene cuatro tipos de conformidades y se siguen conforme al alcance del modelo.

- ❖ Conformidad de modelado de procesos
- ❖ Conformidad de ejecución de procesos
- ❖ Conformidad de procesos BPEL
- ❖ Conformidad de modelado de coreografías

En nuestro caso nos centramos en la conformidad de modelado de procesos debido al alcance del proyecto limitado al desarrollo de un modelo. La ISO especifica que puede ser completa la conformidad si y solo si se cumple con cada uno de los puntos que contenga la conformidad aplicada (Object Management Group / ISO/IEC 19510:2013, 2013).

La conformidad de modelado de procesos contiene 7 puntos, los cuales se revisaron para desarrollar el modelo conceptual.

1. Elementos del proceso BPMN

Los elementos centrales de BPMN: Consta de todos los elementos de diagramas de colaboración y procesos como tareas, eventos, carriles, participantes, textos, mensajes de grupo, anotaciones, flujo de mensajes, entre otros y se dividen en:

- ❖ Descriptivo: Se ocupa de los elementos y atributos visibles utilizados en el modelado de alto nivel.
- ❖ Analítico: Contiene todos los constructos descriptivos, como elementos visibles y auxiliares.
- ❖ Ejecutable común: Se centra en lo necesario para los modelos de procesos ejecutables.

En este primer punto se presenta una lista de los elementos de BPMN, los cuales se clasifican conforme a cada conformidad descriptivo, analítico y ejecutable común. Esta lista solamente contiene los elementos que se aplicaron para este proyecto (tabla 5).

Se utilizaron 10 de 24 elementos de la conformidad descriptiva.

Se utilizaron 2 de 33 elementos de la conformidad analítica.

Se utilizaron 6 de 40 elementos de la conformidad de ejecutable común.

Tabla 9 Elementos y criterios de conformidad de modelado de procesos

CONFORMIDAD	ELEMENTOS	ATRIBUTOS	SE REALIZÓ	
			SI	NO
Conformidad descriptiva	Participant (pool)	id, name, processRef	X	
	sequenceFlow	id, name, sourceRef, targetRef	X	
	messageFlow	id, name, sourceRef, targetRef	X	
	exclusiveGateway	id, name	X	
	parallelGateway	id, name	X	
	subProcess (expanded)	id, name, flowElement	X	
	subProcess (collapsed)	id, name, flowElement	X	
	Text Annotation	id, text	X	
	startEvent (None)	id, name	X	
	endEvent (None)	id, name	X	
Conformidad analítica	exclusiveGateway	Add default attribute	X	
	eventBasedGateway	id, name, eventGatewayType	X	
Conformidad de ejecutable común	sequenceFlow (conditional)	id, name, sourceRef, targetRef, conditionExpression	X	
	subProcess (expanded)	id, name, flowElement, loopCharacteristics, boundaryEventRef	X	
	exclusiveGateway	d, name, gatewayDirection (only converging and diverging), default	X	
	parallelGatewayd	id, name, gatewayDirection (only converging and diverging)	X	
	startEvent (None)	id, name	X	
	endEvent (None)	id, name	X	

2. Aspecto visual

Un elemento clave de BPMN es la elección de formas e iconos utilizados para los elementos gráficos. La intención es crear un lenguaje visual estándar que todos los modeladores de procesos reconozcan y entiendan.

BPMN permite ampliar los elementos estándar con atributos adicionales. Los modeladores y las herramientas de modelado pueden utilizarlo para añadir elementos o artefactos no estándar con el fin de satisfacer una necesidad específica, como los requisitos únicos de un dominio.

Además, menciona que existe flexibilidad en el tamaño, color, estilo de línea y posición del texto de los elementos gráficos definidos, excepto cuando se especifique lo contrario. Algunos ejemplos son:

- ❖ Los elementos BPMN (por ejemplo, objetos de flujo) pueden tener etiquetas (por ejemplo, su nombre y/u otros atributos) colocadas dentro de la forma, o por encima o por debajo de la forma, en cualquier dirección o ubicación, dependiendo de la preferencia del modelador o proveedor de herramientas de modelado.

- ❖ Los objetos de flujo y los marcadores pueden ser de cualquier tamaño que se adapte a los propósitos del modelador o de la herramienta de modelado.
- ❖ Las líneas que se utilizan para dibujar los elementos gráficos pueden ser negras.

3. Conformidad estructural

La conformidad estructural garantiza la correcta interpretación del diagrama como una especificación del proceso, en términos de flujos de control e información.

BPMN se puede ajustar a las especificaciones y restricciones con respecto a las conexiones y otras relaciones diagramáticas entre elementos gráficos (reglas de secuencias de flujo).

Reglas de secuencia de flujo

Un flujo de secuencia entrante puede conectarse a cualquier ubicación de un Objeto de Flujo (izquierda, derecha, arriba o abajo). BPMN permite esta flexibilidad; sin embargo, también se recomienda que los modeladores usen su juicio o las mejores prácticas acerca de cómo los Objetos de Flujo deben ser conectados para que los lectores de los diagramas encuentren el comportamiento claro y fácil de seguir.

La Tabla 11 muestra los Objetos de Flujo BPMN y cómo estos objetos pueden conectarse entre sí a través de Flujos de Secuencia. El símbolo indica que el objeto listado en la fila puede conectarse con el objeto listado en la columna.

Tabla 10 Reglas de secuencia de flujo

From\To						
		↗	↗	↗	↗	↗
		↗	↗	↗	↗	↗
		↗	↗	↗	↗	↗
		↗	↗	↗	↗	↗
		↗	↗	↗	↗	↗
						

Un ejemplo para entender mejor las conexiones es la fig. 19, donde observamos conectados dos elementos *activity*, para rectificar que se encuentra correctamente conectado observamos

la tabla 11, la cual muestra que si se pueden conectar. En la fig. 20 observamos un elemento *activity*, un *gateway*, y otros *activity*, conforme a la tabla si están correctamente conectados

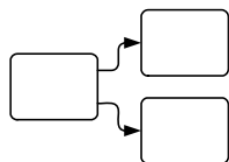


Figura 20 Ejemplo de secuencia de flujo

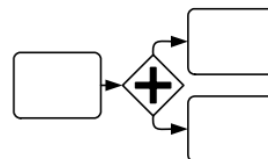


Figura 19 Ejemplo de secuencia de flujo

4. Semántica del proceso

En BPMN existe un paquete Proceso que contiene clases que se utilizan para modelar el flujo de Actividades, Eventos y Pasarelas, y cómo se secuencian dentro de un proceso.

Los procesos pueden definirse a cualquier nivel, desde procesos de toda la empresa a procesos realizados por una sola persona. Los procesos de bajo nivel pueden agruparse para lograr un objetivo común.

5. Atributos y asociaciones de modelos

Esta especificación define una serie de atributos y propiedades de los elementos semánticos representados por los elementos gráficos.

Algunos atributos se especifican como obligatorios, pero no tienen representación o sólo tienen representación opcional, y algunos atributos se especifican como opcionales.

6. Elementos ampliados y opcionales

En este punto explica que una implementación no está obligada a soportar ningún elemento o atributo que se especifique como no normativo o informativo. Esto aplica en los casos de elementos ampliados y opcionales. Al no tener ningún elemento de este tipo, este punto no se toma en consideración al modelo.

7. Intercambio visual

Permite la portabilidad de los diagramas de proceso de forma que los usuarios puedan tomar definiciones de procesos de negocio creadas en el entorno de un proveedor y utilizarlas en el entorno de otro proveedor.

B) Conforme al programa Bizagi Modeler

Otra forma de validar que el modelo está correctamente construido fue por medio del programa en el que se realizó dicho modelo.

El proceso de anotación del gen V se diseñó con el programa Bizagi Modeler, una herramienta para la creación de modelos BPMN. El programa contiene una sección de “validación” que verifica que no contenga algún error en el modelo. Esta herramienta se utilizó como otro filtro para validar el modelo.

En caso de haber empleado mal algún elemento, el programa arrojará el siguiente mensaje (fig. 21), y posteriormente mostrará en dónde se encuentra el error y la ubicación exacta dentro del diagrama.

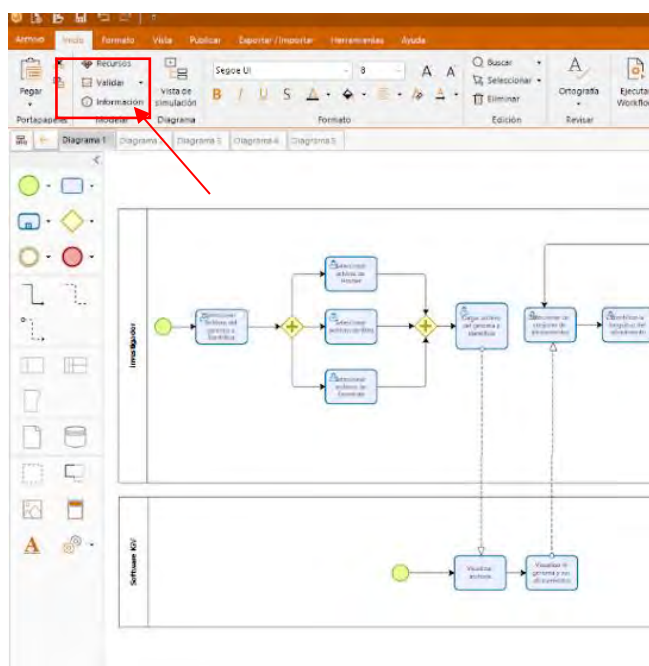


Figura 21 Programa Bizagi Modeler

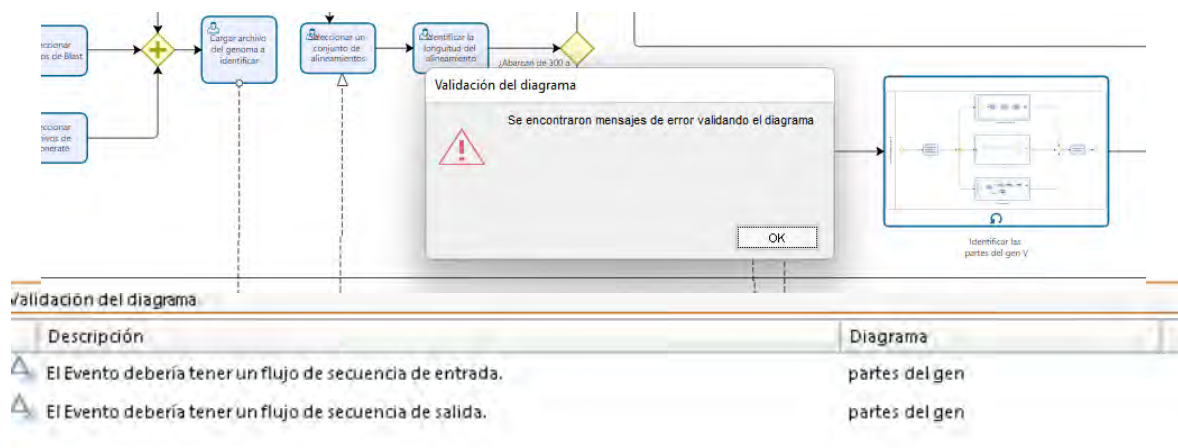


Figura 21.1 Programa Bizagi Modeler

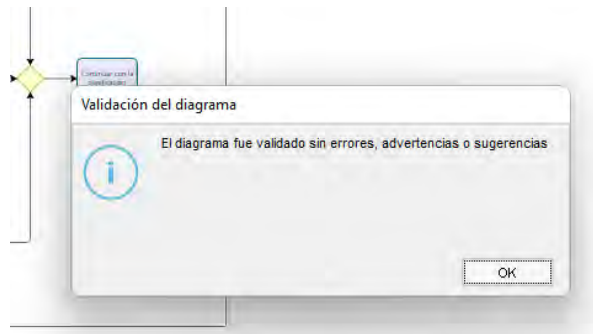


Figura 21.2 Programa Bizagi Modeler

Una vez corregido el error, aparece un mensaje (fig. 21.2) validando que el modelo ya se encuentra sin errores, advertencias o sugerencias.

Conclusión de la evaluación:

Esta fase se evaluó de dos formas, 1) la conformidad del modelo respecto a la ISO 19510 y 2) La validación con el programa Bizagi Modeler.

Una vez que se verificó el modelo conforme a lo establecido en los pasos anteriores de la conformidad de modelado de procesos, se puede afirmar que: 1) el modelo BPMN está completo y se aplicó correctamente con base en la ISO 19510, validando que se encuentra correctamente construido, 2) conforme al programa Bizagi Modeler se validó que el modelo conceptual del proceso de anotación del gen V, no marca ningún error de conexión o elemento mal empleado, por lo tanto, podemos decir que se encuentra correctamente construido.

5.3 Fase 3 Correctes del modelo por parte de expertos de biología molecular

Otro punto importante para validar el modelo fue la correcta representación del conocimiento de los biólogos moleculares en el proceso de anotación del gen V.

Esta fase fue evaluada por tres expertos, un inmunólogo, un bioinformático y un biólogo, siendo expertos en el proceso de anotación de genes. El Dr. Jesús Barnetche es el principal encargado de la curación manual del gen V y se encuentra trabajando con una serie de especies de quirópteros. Hasta el momento se han logrado clasificar 9 de 13 especies a lo largo de un año y medio. El objetivo del Dr. Barnetche es sistematizar el proceso de anotación para así identificar los genes funcionales y no funcionales con la menor intervención humana posible. El modelo conceptual busca la estandarización para el aprendizaje del proceso de anotación de genes y a futuro es un paso para sistematizar el proceso.

Se presentó la versión final del modelo conceptual a los expertos para validar que se encuentra correctamente construido conforme al proceso de anotación del gen V y sus reglas, para posteriormente evaluar el modelo.

La fase 3 se dividió en dos secciones:

A) Puntos importantes para considerar en un modelo conceptual. El autor (Stephen A. White et al., 2009) menciona que un modelo BPMN debe contener los siguientes puntos para considerarse un buen modelo, los cuales fueron considerados para el desarrollo del modelo y la aplicación del cuestionario a los expertos:

- **Selectivo:** Ningún modelo puede representar todo, debe representar selectivamente los aspectos que son más relevantes de la tarea en cuestión.
- **Exacto:** El modelo debe codificar exactamente el estado actual del negocio y no una noción parcial o errónea.
- **Cuidadosamente completo:** El modelo debe ser lo más simple posible, pero no más simple que eso.
- **Comprensible:** Una vez que se percibe el modelo, debe estar en condiciones de encontrarle sentido, no debe ser muy complicado o resultar poco familiar para comprender.

B) Conformidad del modelo. Una vez validando que el modelo se encuentra correctamente descrito conforme al conocimiento de los expertos, se diseñó y se aplicó un cuestionario utilizando un cuestionario conforme a los puntos de la sección A.

Para aplicar el cuestionario se establecieron 8 pasos:

1. **Elaborar los enunciados:** En este caso se realizaron 10 preguntas con referencia a la utilidad del modelo conceptual.
2. **Establecer las opciones de respuesta:** Se midió el grado en que el encuestado está de acuerdo o en desacuerdo con cada consulta. En este caso se eligió pésimo, malo, regular, bueno y excelente de acuerdo con una encuesta de satisfacción.

3. **Aplicar la escala:** En este punto se debe aplicar a las personas correctas para tener una aplicación exitosa. En este caso se aplicó a los expertos del proceso de anotación.
4. **Asignar puntos a los enunciados:** Se decidió el número de puntos en la escala. En este caso se eligió del 1 al 5 donde indica “1- Pésimo, 2-Malo, 3- Regular, 4-Bueno, 5-Excelente”.

Preguntas y resultados obtenidos de la encuesta se muestran a continuación:

1. ¿El modelo representa los aspectos más importantes del proceso de anotación?
3 respuestas



Figura 22 Pregunta uno del cuestionario de los expertos

La pregunta 1 ¿El modelo representa los aspectos más importantes del proceso de anotación? Se plantearon dos opciones, donde se obtuvo el 100% de si, considerando que se representaron los aspectos más importantes.

2. Del 1 al 5 ¿Qué tan comprensible es el modelo
3 respuestas

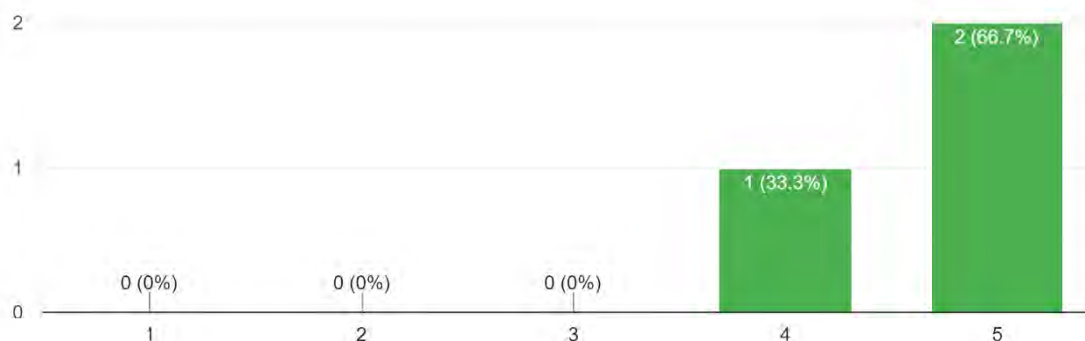


Figura 23 Pregunta dos del cuestionario de los expertos

La pregunta 2 ¿Qué tan comprensible es el modelo? Se obtuvo el 66.7% con la puntuación de excelente que se interpreta como es muy comprensible el modelo, y un 33.3 % con la puntuación de bueno que se interpreta como es comprensible el modelo.

3. Del 1 al 5 ¿Qué tan fácil es seguir el modelo para anotar el gen V ?

3 respuestas

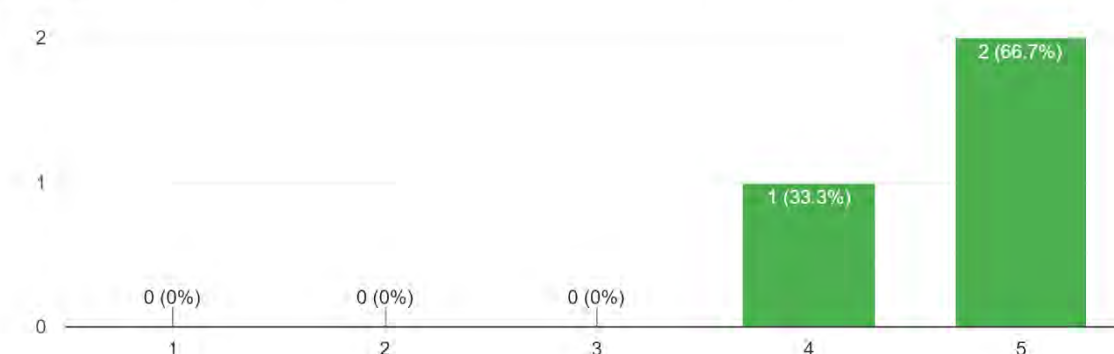


Figura 24 Pregunta tres del cuestionario de los expertos

La pregunta 3 ¿Qué tan fácil es seguir el modelo para anotar el gen V? Se obtuvo el 66.7% con la puntuación de excelente que se interpreta como es muy fácil de seguir el modelo, y un 33.3 % con la puntuación de bueno, que se interpreta como es fácil de seguir el modelo.

4. Del 1 al 5 ¿Qué tan correcto se representa el proceso de anotación del Gen V?

3 respuestas

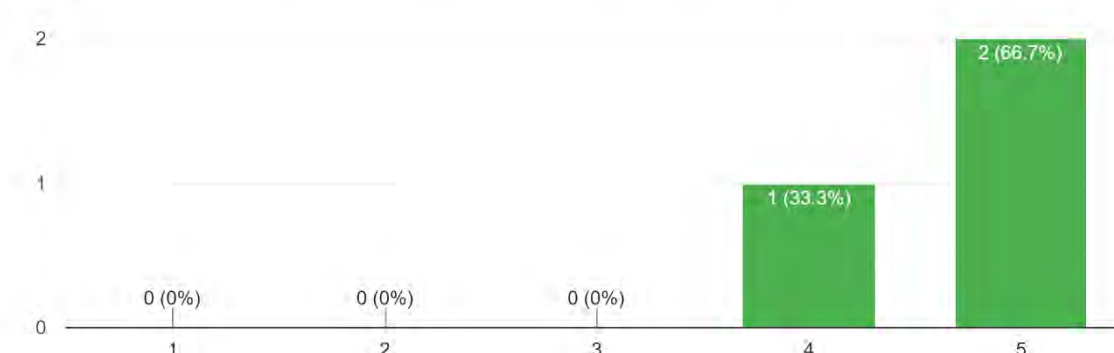


Figura 25 Pregunta cuatro del cuestionario de los expertos

La pregunta 4 ¿Qué tan correcto se representa el proceso de anotación del Gen V? Se obtuvo el 66.7% con la puntuación de excelente que se interpreta como es muy correcto el modelo, y un 33.3 % con la puntuación de bueno, que se interpreta como es correcto el modelo.

5. Del 1 al 5 ¿Qué tan correcto se representan las reglas para identificar las partes del Gen V?

3 respuestas

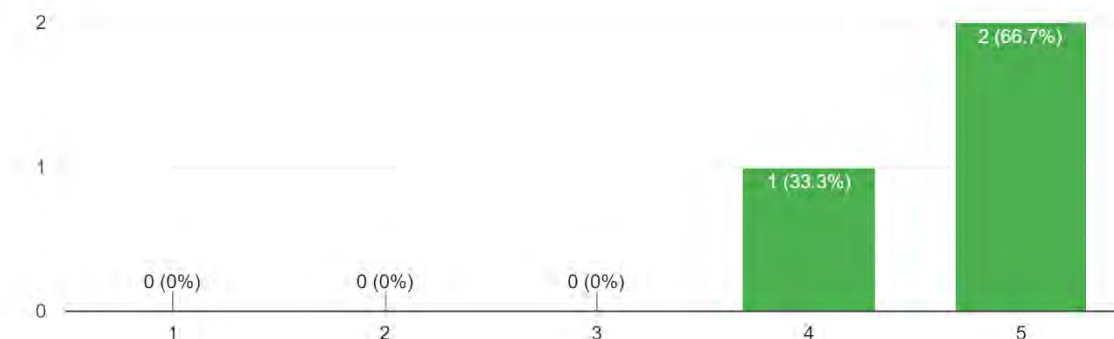


Figura 26 Pregunta cinco del cuestionario de los expertos

La pregunta 5 ¿Qué tan correcto se representan las reglas para identificar las partes del Gen V? Se obtuvo el 66.7% con la puntuación de excelente que se interpreta como son muy correctas las reglas del modelo, y un 33.3 % con la puntuación de bueno, que se interpreta como son correctas las reglas del modelo.

6. ¿Tenían anteriormente alguna representación gráfica de cómo realizar el proceso de anotación de los genes V?

3 respuestas

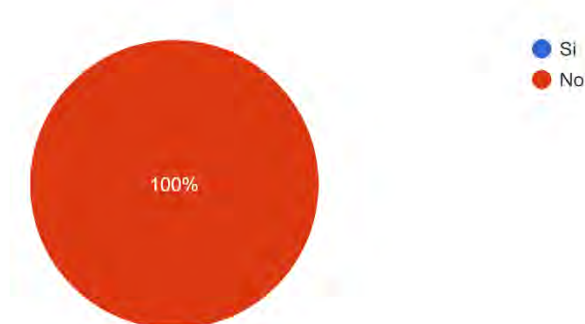


Figura 27 Pregunta seis del cuestionario de los expertos

La pregunta 6 ¿Tenían anteriormente alguna representación gráfica de cómo realizar el proceso de anotación de los genes V? Se obtuvo el 100 % afirmando que anteriormente no se tenía alguna representación gráfica. Por lo cual deduce su utilidad.

7. Del 1 al 5 ¿Qué tan exacto es el modelo?

3 respuestas

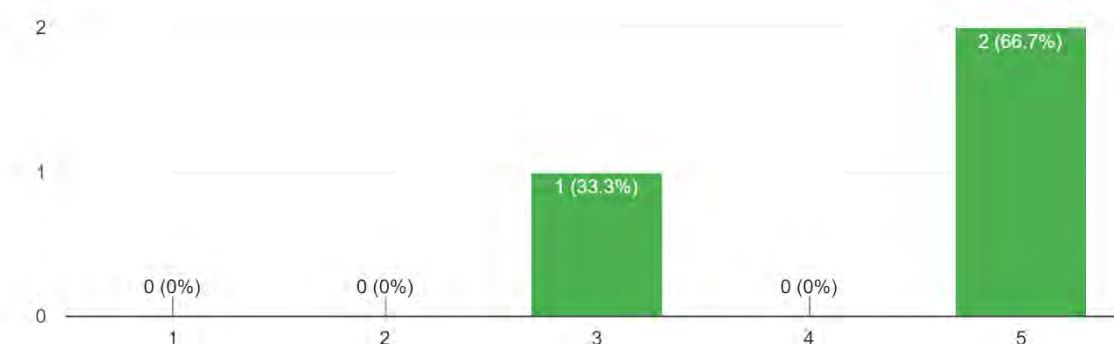


Figura 28 Pregunta siete del cuestionario de los expertos

La pregunta 7 ¿Qué tan exacto es el modelo? Se obtuvo el 66.7% con la puntuación de excelente que se interpreta como es muy exacto el modelo, y un 33.3 % con la puntuación de regular, que se interpreta como es regularmente exacto.

8. Del 1 al 5 ¿Que tan completo se encuentra el modelo?

3 respuestas ¿Qué

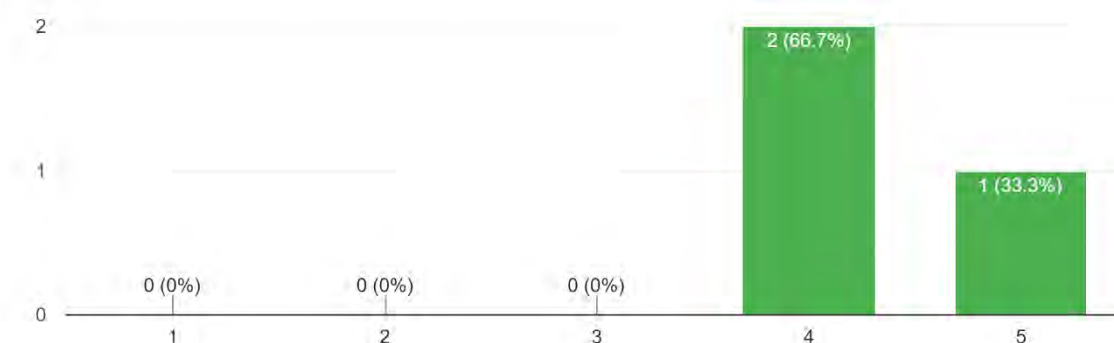


Figura 29 Pregunta ocho del cuestionario de los expertos

La pregunta 8 ¿Qué tan completo se encuentra el modelo? Se obtuvo el 66.7% con la puntuación de bueno que se interpreta como es completo el modelo, y un 33.3 % con la puntuación de excelente que se interpreta como es muy completo el modelo.

9. ¿Utilizaría el modelo para enseñarle a nuevos curadores el proceso de anotación del gen V?
3 respuestas



Figura 30 Pregunta nueve del cuestionario de los expertos

La pregunta 9 ¿Utilizaría el modelo para enseñarle a nuevos curadores el proceso de anotación del gen V? Se obtuvo el 100 % afirmando que sí se utilizaría el modelo para la enseñanza a principiantes.

10. ¿El modelo presenta el estado actual de la curación del proceso de anotación del gen V?
3 respuestas



Figura 31 Pregunta diez del cuestionario de los expertos

La pregunta 10 ¿El modelo presenta el estado actual de la curación del proceso de anotación del gen V? Se obtuvo el 100 %, afirmando que sí se ha representado el estado actual del proceso de anotación en el modelo.

Como último, se brindó una sección para comentarios / sugerencias. Estos son algunos de los comentarios de los expertos:

Faltan algunos aspectos sobre las secuencias canónicas tanto en el péptico señal como en las regiones V y señales de re combinación.

Felicitaciones por el esfuerzo y la dedicación

Esta completo, una buena representación del proceso

Figura 32 Comentarios del cuestionario de los expertos

El primer comentario de la figura 32 fue atendido por el Dr. Jesús Barnetche, encargado principal de determinar el alcance del modelo. Nos indica que las secuencias canónicas no se requieren incluir en el modelo.

5. **Generar una tabla de datos:** Una vez que se recopiló la opinión de los encuestados, se creó una tabla en Excel con una primera columna con el nombre de la persona que respondió la encuesta. Después, se añadieron los valores que corresponden a las respuestas (tabla 12).
6. **Calcular los puntajes:** Se realizó una sumatoria final con las respuestas de cada uno de los ítems para conocer la información de los expertos.
 - Suma los valores de cada respuesta para conocer el grado de aceptación de cada ítem.
 - Marca los rangos óptimos de respuesta para saber, en un simple vistazo, cuál es el estatus de los expertos.

En este caso, por cada sujeto el puntaje máximo es 30 y el mínimo es de 23. Conforme a la pregunta, el valor máximo debería ser 15, en este caso se tuvo un 14 como máximo y un 13 como mínimo. Donde se encuentra en un rango aceptable.

Tabla 11 Base de datos de los encuestados del cuestionario de los expertos

No.	Encuestado	Especialidad	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	Pregunta 7	Pregunta 8	Valor por sujeto	Promedio
1	Sujeto A	Inmunología	4	4	4	4	3	4	23	3.8
2	Sujeto B	Bioinformático	5	5	5	5	5	4	29	4.8
3	Sujeto C	Doctorado en Ciencias	5	5	5	5	5	5	30	5.0
		Valor por elemento	14	14	14	14	13	13		
		Promedio	4.7	4.7	4.7	4.7	4.3	4.3		

7. **Calcular la frecuencia de cada elemento:** Se inserta una tabla que incluye las alternativas de opción múltiple y se anota el valor de cada una de estas respuestas. Posteriormente, se calcula la frecuencia de los elementos, la cual indica el número de veces que un factor se repite dentro de una serie (tabla 12).

Tabla 12 Frecuencia de elementos del cuestionario de los expertos

Frecuencia							
No.	Encuestado	pregunta 2	pregunta 3	pregunta 5	pregunta 6	pregunta 7	pregunta 8
1	Pésimo	0	0	0	0	0	0
2	Malo	0	0	0	0	0	0
3	Regular	0	0	0	0	1	0
4	Bueno	1	1	1	1	0	2
5	Excelente	2	2	2	2	2	1

- 8. Elaborar gráficos:** Primero, se suman los valores individuales de respuesta de acuerdo con cada ítem, lo cual da un valor absoluto en la encuesta. Después, se divide el valor de cada ítem entre el total, obteniendo el porcentaje.

Tabla 13 Porcentaje de la encuesta del cuestionario de los expertos

Porcentaje			
No.	Encuestado	Valor por sujeto	Porcentaje
1	Pésimo	0	0%
2	Malo	0	0%
3	Regular	1	6%
4	Bueno	6	33%
5	Excelente	11	61%
	TOTAL	18	100%

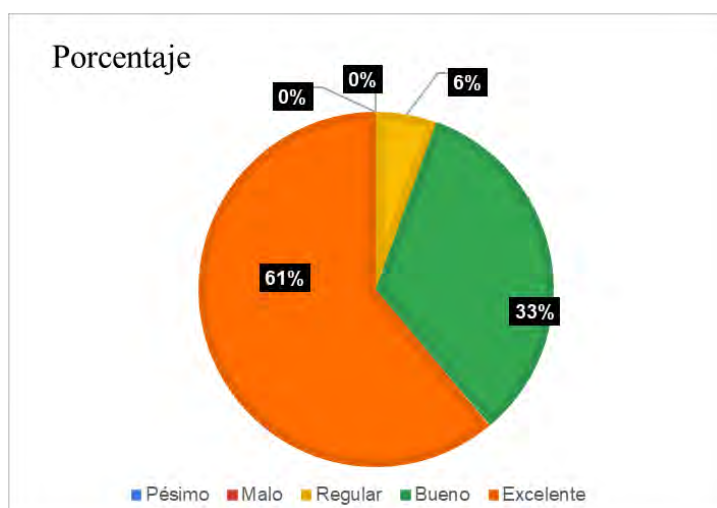


Figura 33 Gráfica del porcentaje de correctes del modelo conceptual conforme a los expertos

Para finalizar, se realizó una gráfica (fig.33) que muestra el porcentaje de acuerdo con los encuestados, obteniendo un 61% de excelencia, un 33% de Bueno y 6% de Regular en el modelo desarrollado. Estos resultados nos muestran la satisfacción de los expertos conforme a la correctes del modelo, recordando que se encuestaron a 3 expertos siendo toda nuestra población.

Conclusión: En esta fase obtuvimos excelente a bueno, conforme a la evaluación del cuestionario. Los resultados nos dicen que se encuentra correctamente representado el conocimiento de los expertos en el modelo conceptual.

De acuerdo con la aprobación de los expertos y las correcciones de la fase 4, se entregó el modelo conceptual del proceso de anotación del gen V, para usos prácticos en el CISEI/INSP.

5.4 Fase 4 Aprendizaje del proceso de anotación con el modelo conceptual

En esta fase 4 se evaluó la utilidad del modelo para el aprendizaje del proceso de anotación en un caso real. En el CISEI/INSP tienen comprometido un proyecto de investigación en el que realizan el proceso de anotación de genes de 13 especies de quirópteros. Para ello reclutaron estudiantes de diversas instituciones, entre ellas de la Universidad Autónoma del Estado de Morelos y del propio INSP. En total se reclutaron 12 aprendices para realizar el proceso de anotación, algunos sin experiencia y otros con un poco de conocimiento.

Esta fase se dividió en 2 secciones:

A) Aprendizaje del proceso de anotación para los aprendices

Para enseñar el proceso de anotación a los aprendices, se impartió un curso de anotación por parte de expertos del CISEI/INSP. A continuación, se explica a grandes rasgos por medio de los siguientes 6 pasos:

1. Conocimientos generales de la composición de los genes V, D, J. Estructura biológica, reglas de composición y para identificar los genes, así como para diferenciarlos unos de otros. Se hace énfasis en los conceptos claves a desarrollar a lo largo del curso.
2. Utilización de Linux para abrir el programa de software IGV que permite visualizar, analizar y anotar el genoma. Siendo la mayoría biólogos con pocos conocimientos acerca de computación, fue necesario enseñarles paso a paso a levantar el ambiente en el que se estará trabajando. Se incluye una sesión para el aprendizaje de los comandos para realizar esta actividad.
3. Posteriormente se mostró cómo se utiliza el software IGV, cómo se navega, se edita, se guarda, se carga, se expande y se anota, entre otras tareas esenciales para la anotación de los genes.
4. Una vez ya conociendo cómo funciona el software, se procede a enseñar cómo realizar la anotación de los genes. Desde cómo cargar los archivos, ubicar un gen, identificar un gen, y cómo clasificar el gen. También se enseña a identificar las partes del gen y cómo leer las reglas, así como la funcionalidad del gen y cómo etiquetar un gen correctamente en el software IGV.
5. Después de entender cómo hacer la anotación conforme al paso 4, se practica anotando una especie la cual contenga muy pocos genes para clasificar, todo se realizaba bajo supervisión de un experto, el cual se encarga de resolver dudas y recordar las reglas.
6. Terminando de anotar la especie de práctica, se le asigna a cada aprendiz una especie para trabajar y clasificarla individualmente.

Como material de apoyo y guía para realizar estas actividades se cuenta únicamente con algunas presentaciones en PowerPoint, muy poco ilustrativas, que llegaban a causar confusión. La anotación no se encuentra representada dentro de este material de apoyo, solo

contiene las reglas para encontrar un gen. Al no tener representado el proceso, ocasiona muy poca retención que provoca olvidar la anotación.

B) Aplicación a los aprendices de una encuesta para la evaluación de la utilidad del modelo

Para poder evaluar qué tan útil fue el modelo para el aprendizaje se realizó en dos etapas.

-La primera fue meramente intuitiva. Los 12 aprendices participaron del proceso de aprendizaje como se mencionó en el inciso A) anterior. Realizaron cada uno de los pasos, y se se dejó que identificaran un gen V sin ayuda. Se observó que la mayoría se encontraban confundidos al no recordar cómo identificar un gen, sus reglas y cómo clasificarlos, por lo finalmente para concluir su tarea tuvieron que pedir ayuda.

-En la segunda etapa se mostró, se explicó a los 12 aprendices el modelo conceptual con la representación del proceso de anotación del gen V. Se dejó que anotaran un gen V solo utilizando el modelo, sin ningún tipo de ayuda. Se observó que, al tener una representación gráfica del proceso, no requirieron ayuda, debido a que el modelo representa todo el proceso, cómo identificar un gen V, sus reglas y cómo clasificarlo.

Una vez que se realizó esta experimentación, se continuó con la aplicación de un cuestionario para evaluar la utilidad del modelo en el aprendizaje del proceso de anotación del gen V.

En estas pruebas se presenta una declaración y la persona debe indicar en qué grado se siente de acuerdo o en desacuerdo con su contenido, se estableció 8 pasos:

- 1. Elabora los enunciados:** En este caso se realizaron 10 preguntas con referencia a la utilidad del modelo conceptual.
- 2. Establece las opciones de respuesta:** Se eligió pésimo, malo, regular, bueno y excelente de acuerdo con una encuesta de satisfacción.
- 3. Aplica la escala:** Se aplicó a los principiantes del proceso de anotación.
- 4. Asigna puntos a los enunciados:** Se decidió el número de puntos en la escala. En este caso se eligió del 1 al 5 donde indica “1- Pésimo, 2-Malo, 3- Regular, 4-Bueno, 5- Excelente”.

A continuación, se muestra el cuestionario que se aplicó al grupo de principiantes:

1. ¿Ha realizado el proceso de anotación anteriormente?

12 respuestas

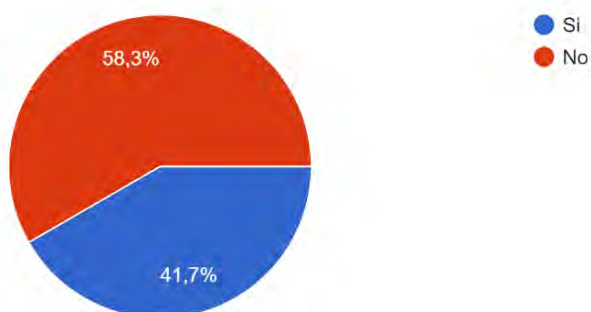


Figura 34 Pregunta uno del cuestionario de los aprendices

En la pregunta 1 se obtuvo de las 12 respuestas un 58.3% de personas que no habían realizado anteriormente el proceso de anotación y el otro 41.7% ya había realizado el proceso; sin embargo, algunos no dominaban aún el proceso.

2. Del 1 al 5 ¿Qué tan bien representa el proceso de anotación el modelo?

12 respuestas

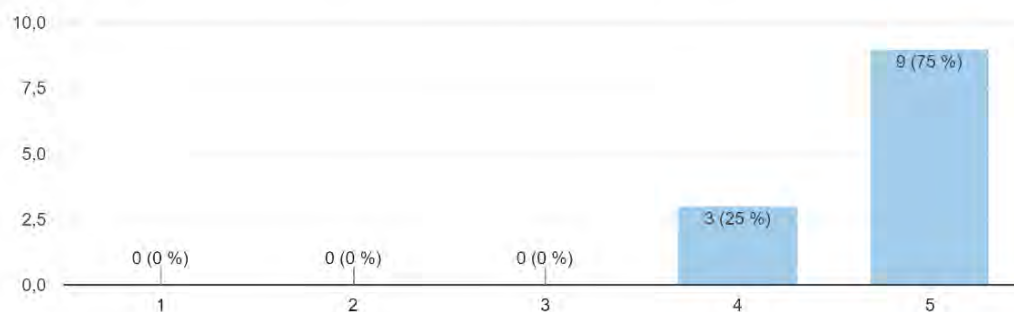


Figura 35 Pregunta dos del cuestionario de los aprendices

En la pregunta 2 ¿Qué tan bien se representa el proceso de anotación del modelo? Se obtuvo el 75% con la puntuación de excelente que se interpreta como se ha representado muy bien. Y un 25% respondió con la puntuación de bueno que se interpreta a que se representa bien.

3. Del 1 al 5 ¿Qué tan claras son las reglas para identificar las partes del gen V?

12 respuestas

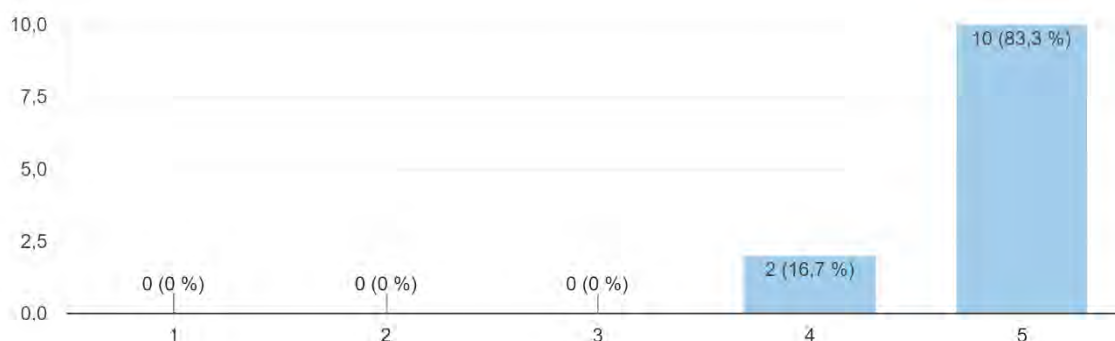


Figura 36 Pregunta tres del cuestionario de los aprendices

En la pregunta 3 ¿Qué tan claras son las reglas para identificar las partes del gen V? Se obtuvo un 83.3% que respondió excelente, que se interpreta que son muy claras las reglas y un 16.7% con un bueno que se interpreta cómo son claras las reglas.

4. ¿Te gustaría que se representaran los genes D y J en el modelo?

12 respuestas

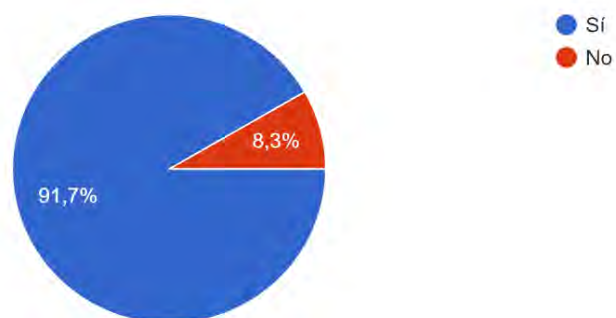


Figura 37 Pregunta cuatro del cuestionario de los aprendices

En la pregunta 4 ¿Te gustaría que se representaran los genes D y J en el modelo? Se dio dos opciones sí o no, en la cual se tuvo el 91.7% con un sí y el 8.3% con un no.

5. Del 1 al 5 ¿En qué medida te sirvió el modelo para entender el proceso de anotación?

12 respuestas

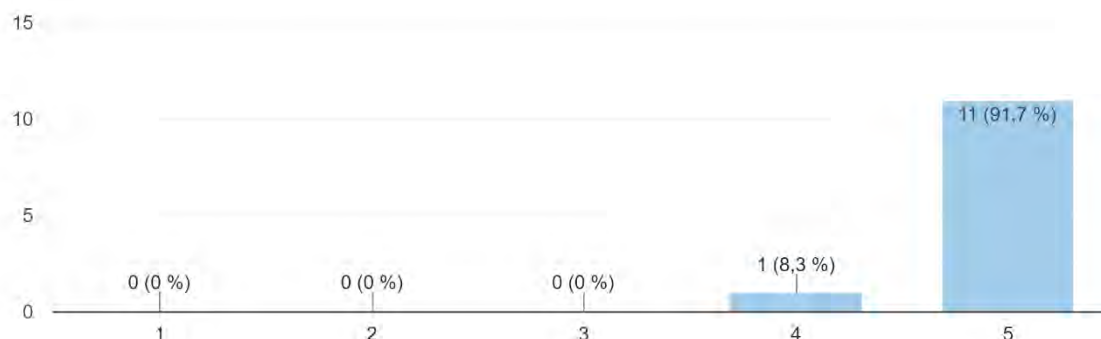


Figura 38 Pregunta cinco del cuestionario de los aprendices

En la pregunta 5 ¿En qué medida te sirvió el modelo para entender el proceso de anotación? Respondieron el 91.7% con la puntuación de excelente, que se interpreta con un muy bien y el 8.3% con la puntuación de bueno que se interpreta con un bien.

6. Del 1 al 5 ¿Qué tan fácil es seguir el modelo para anotar el gen V ?

12 respuestas

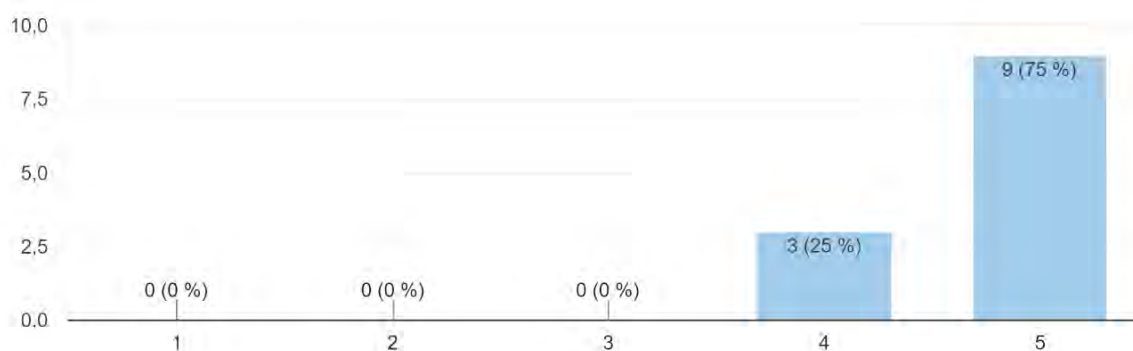


Figura 39 Pregunta seis del cuestionario de los aprendices

En la pregunta 6 ¿Qué tan fácil es seguir el modelo para anotar el gen V? Se obtuvo un 75% que se interpreta con un muy fácil y un 25% que se interpreta como fácil.

7. Del 1 al 5 ¿ En qué medida te sirvió el modelo para anotar el gen V?

12 respuestas

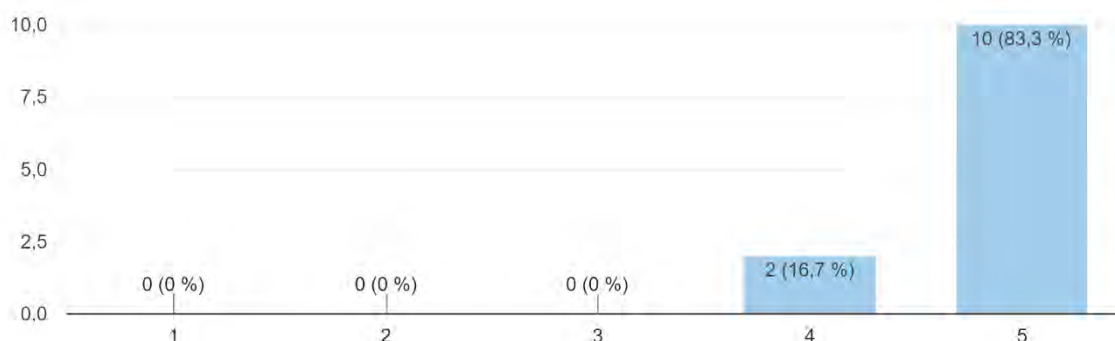


Figura 40 Pregunta siete del cuestionario de los aprendices

En la pregunta 7 ¿En qué medida te sirvió el modelo para anotar el gen V? Respondieron un 83.3% excelente y un 16.7% bien.

8. Del 1 al 5 ¿Qué tan práctico te resulta el modelo?

12 respuestas

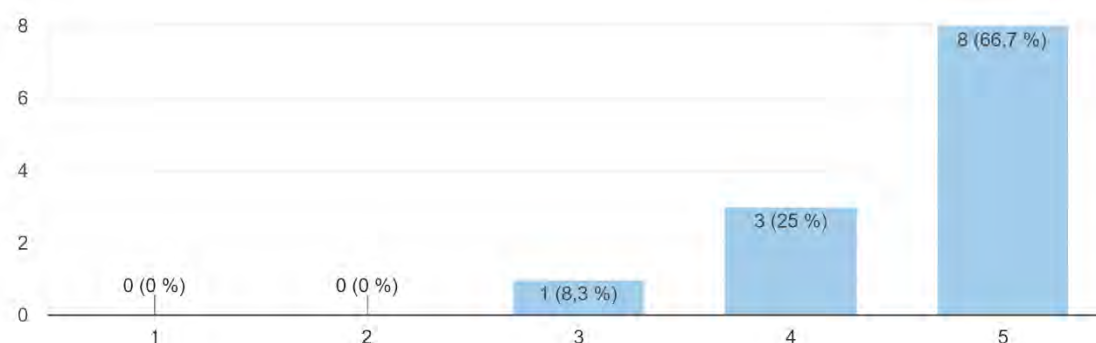


Figura 41 Pregunta ocho del cuestionario de los aprendices

En la pregunta 8 ¿Qué tan práctico te resulta el modelo? Se obtuvo un 66.7% que se interpreta con un muy bueno, un 25% que se interpreta como bueno y un 8.3% que se interpreta como regular.

9. Si tuviera que enseñarle a alguien el proceso de anotación del gen V ¿Utilizarías el modelo?

12 respuestas



Figura 42 Pregunta nueve del cuestionario de los aprendices

En la pregunta 9 Si tuviera que enseñarle a alguien el proceso de anotación del gen V ¿Utilizarías el modelo? Se obtuvo el 100 % afirmando que si enseñaran el proceso por medio del modelo.

10. En caso de haber realizado anteriormente el proceso de anotación ¿Te hubiera gustado aprender con el modelo?

6 respuestas

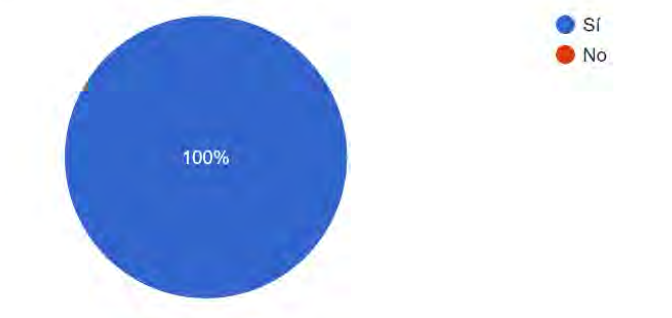


Figura 43 Pregunta diez del cuestionario de los aprendices

En la pregunta 10 en caso de haber realizado anteriormente el proceso de anotación ¿Te hubiera gustado aprender con el modelo? Se obtuvo el 100 % afirmando que si les hubiera gustado aprender con el modelo.

Como último se brindó una sección para comentario/sugerencias. Estos son algunos de los comentarios obtenidos, los cuales fueron revisados y se atendieron conforme a la decisión del experto.

Una buena herramienta de apoyo para aprender el proceso, ayuda bastante
Es un buen modelo para empezar a enseñar a hacer anotación de genes de inmunoglobulinas
Solo como observación, estaría bien conectar el proceso de RSS's y SP con la clasificación de la funcionalidad del gen VH. Ya que si hay un SP o una RSS no funcional en automático el gen V, ya no es funcional. En general muy bien
Un agradecimiento por el diagrama de flujo que nos permite identificar facilmente los criterios de interese !!
Es la primera vez que hago este tipo de anotaciones. El modelo es bueno para realizar las anotaciones del gen en cuestión.
Buen modelo, ayuda a entender mejor el proceso que solo explicaso
Se puede considerar utilizar un espacio en el apartado numero 3 para extender los componentes posibles de la etiqueta.

Figura 44 Comentarios del cuestionario de los aprendices

- 5. Generar una base de datos:** Una vez que se recopiló la opinión de tus encuestados, se crea una tabla en Excel (tabla 14).
- 6. Calcular los puntajes:** Se realiza una sumatoria final con las respuestas de cada uno de los ítems para conocer la información de los aprendices.
En este caso por cada sujeto el puntaje máximo es 30 y el mínimo es de 25. Y conforme a la pregunta, el valor máximo debería ser 60, en este caso se tuvo un 59 como máximo y un 55 como mínimo. Donde se encuentra en un rango aceptable.

Tabla 14 Base de datos de los encuestados del cuestionario de los aprendices

Encuestado	Carrera/Especialidad	Pregunta 2	Pregunta 3	Pregunta 5	Pregunta 6	Pregunta 7	Pregunta 8	Valor por sujeto	Promedio
Sujeto A	Biología	5	5	5	5	5	5	30	5.0
Sujeto B	Doctorado en Ciencias	4	4	4	4	4	4	24	4.0
Sujeto C	Maestría en Ciencias	5	5	5	5	5	5	30	5.0
Sujeto D	Biología	4	5	5	5	5	3	27	4.5
Sujeto E	Biología	5	5	5	5	5	5	30	5.0
Sujeto F	Biología	5	5	5	5	5	5	30	5.0
Sujeto G	Dr. en Ciencias	5	5	5	5	5	5	30	5.0
Sujeto H	Medicina	5	5	5	5	5	4	29	4.8
Sujeto I	Químico	4	4	5	4	4	4	25	4.2
Sujeto J	Biología	5	5	5	5	5	5	30	5.0
Sujeto K	Biología	5	5	5	4	5	5	29	4.8
Sujeto L	Biología	5	5	5	4	5	5	29	4.8
	Valor por elemento	57	58	59	56	58	55		
	Promedio	4.8	4.8	4.9	4.7	4.8	4.6		

7. **Calcular la frecuencia de cada elemento:** Inserta una tabla que incluya las alternativas de opción múltiple y anota el valor de cada una de estas respuestas. Posteriormente, se calcula la frecuencia de los elementos. La cual indica el número de veces que un factor se repite dentro de una serie (tabla 15).

Tabla 15 Frecuencia de elementos del cuestionario de los aprendices

Frecuencia							
No.	Encuestado	pregunta 2	pregunta 3	pregunta 5	pregunta 6	pregunta 7	pregunta 8
1	Pésimo	0	0	0	0	0	0
2	Malo	0	0	0	0	0	0
3	Regular	0	0	0	0	0	1
4	Bueno	3	2	1	3	2	3
5	Excelente	9	10	11	9	10	8

8. **Elaborar gráficos:** Primero, se suman los valores individuales de respuesta de acuerdo con cada ítem. Esto te dará un valor absoluto en tu cuestionario. Después, divide el valor de cada ítem entre el total obteniendo el porcentaje.

Tabla 16 Porcentaje de la encuesta del cuestionario de los aprendices

Porcentaje			
No.	Encuestado	Valor por sujeto	Porcentaje
1	Pésimo	0	0%
2	Malo	0	0%
3	Regular	1	1%
4	Bueno	14	19%
5	Excelente	57	79%
	TOTAL	72	100%

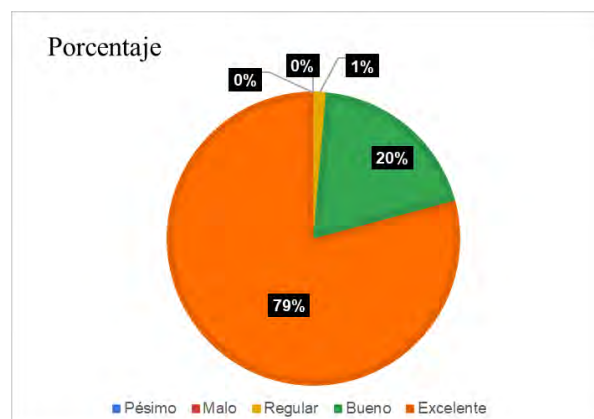


Figura 45 Gráfica del porcentaje de satisfacción con la utilidad del modelo conceptual para los aprendices

Para finalizar se realizó una gráfica (fig.45) en la cual muestra el porcentaje de acuerdo con los encuestados, obteniendo un 79% de excelencia, un 20% de Bueno y 1% de Regular en el modelo desarrollado.

Conclusión: Estos resultados nos muestran la satisfacción conforme a la utilidad que tiene el modelo para los principiantes en la curación, siendo una buena ponderación para el modelo recordando que tenemos 12 encuestados siendo nuestra población total.

En esta fase 4 se logró enseñar a los principiantes con el modelo conceptual, teniendo una favorable puntuación del modelo. El modelo logró estandarizar el proceso de anotación y será una herramienta de gran utilidad para futuros principiantes en el proceso de anotación y actuales expertos.

Capítulo 6 Conclusiones y recomendaciones

El proceso de anotación es un proceso con una alta complejidad debido a la gran cantidad de pasos estructurados, y la incorporación de múltiples fuentes de datos. En particular, la anotación del gen V, es un proceso con una gran dificultad por la gran cantidad de elementos que participan en la anotación.

Entender el proceso de anotación y representarlo en un modelo conceptual fue un gran reto que se llevó a cabo siguiendo los pasos de la metodología Ciencia del Diseño. Aunque la Ciencia del Diseño no solo permitió representar el proceso de anotación, también permitió un amplio entendimiento del dominio, emprender la búsqueda de posibles soluciones y lograr su representación por medio de un artefacto, en este caso un modelo conceptual BPMN.

En la revisión del estado del arte no se encontraron referencias acerca de modelos conceptuales para representar el proceso de anotación del genoma. El modelo BPMN obtuvo una puntuación de bueno a excelente en las fases de evaluación consideradas dentro de la metodología de la Ciencia del Diseño. En particular en la fase 4 Aprendizaje del proceso de anotación con el modelo conceptual, obtuvimos un 79% de aceptación del modelo, lo cual corresponde a que 11 de 12 personas consideran el modelo excelente-bueno y de gran utilidad para el aprendizaje del proceso de anotación.

El modelo logró estandarizar el proceso de anotación del gen V y se estará utilizando para el aprendizaje de principiantes en el proceso de anotación del gen V para diferentes especies de quirópteros dentro del CISEI/INSP. El modelo cumplió con todos los objetivos planteados al inicio de la investigación.

Trabajo futuro

El modelo propuesto del proceso de anotación del gen V es una base sólida para realizar un posterior diseño algorítmico que permita informatizar todo o parte de este proceso, con la finalidad de reducir tiempo y obtener mejores resultados. Dado que en el alcance solo se consideró al gen V, consideramos importante incluir algunos pasos faltantes descritos de la fase 4, y que esta misma metodología se puede aplicar para los genes D y J.

Referencias

- Bernasconi, A., García S, A., Ceri, S., & Pastor, O. (2022). A Comprehensive Approach for the Conceptual Modeling of Genomic Data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13607 LNCS, 194–208. https://doi.org/10.1007/978-3-031-17995-2_14
- Birta, L. G., & Arbez, G. (2013). *Modelling and Simulation*. <https://doi.org/10.1007/978-1-4471-2783-3>
- Brandan, N., Aquino Esperanza, J., & Codutti, A. (2007). *Respuesta Inmunitaria* [Facultad de medicina UNNE]. <https://med.unne.edu.ar/sitio/multimedia/imagenes/ckfinder/files/files/Carrera-Medicina/BIOQUIMICA/inmunitaria.pdf>
- Crovani, P., Pidò, S., & Garzotto, F. (2020a). Towards an Ontology for Tertiary Bioinformatics Research Process. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12584 LNCS, 82–91. https://doi.org/10.1007/978-3-030-65847-2_8/COVER
- Crovani, P., Pidò, S., & Garzotto, F. (2020b). Towards an Ontology for Tertiary Bioinformatics Research Process. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12584 LNCS, 82–91. https://doi.org/10.1007/978-3-030-65847-2_8/COVER
- Davis, M. M. (2004). The evolutionary and structural ‘logic’ of antigen receptor diversity. *Seminars in Immunology*, 16(4), 239–243. <https://doi.org/10.1016/j.smim.2004.08.003>
- De Sordi, J. O. (2021). Design science research methodology: theory development from artifacts. *Springer Nature*. <https://doi.org/978-3-030-82155-5>
- Dresch, A. , L. D. P. , A. J. J. A. V. , D. A. , L. D. P. , & A. J. A. V. (2015). *Design science research: A Method for Science and Technology Advancement* (Springer, Ed.).
- Ejigu, G. F., & Jung, J. (2020). Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology*, 9, 295. <https://doi.org/10.3390/biology9090295>
- Eraso Schattka, L. E. (2011). *Del Esquema Conceptual al Mundo Real, una Base de Datos Genómica*. <https://riunet.upv.es/handle/10251/11365>
- Evermann, J., & Wand, Y. (2001). Towards ontologically based semantics for UML constructs. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2224, 354–367. https://doi.org/10.1007/3-540-45581-7_27/COVER
- Fahim Ahmed, Stewart Robinson, & Antuela A. Tako. (2014). *Using the structured analysis and design technique (SADT) in simulation conceptual modeling*. <https://doi.org/10.1109/WSC.2014.7019963>
- Fotini, P., Casellas, R., Suh, H., Qin, X. F., Besmer, E., Pelanda, R., Nemazee, D., Rajewsky, K., & Nussenzweig, M. C. (1997). V(D)J Recombination in Mature B Cells: A Mechanism for

- Altering Antibody Responses. *Science*, 278(5336), 298–301.
<https://doi.org/10.1126/SCIENCE.278.5336.298>
- Frederick W, A., Eugene M, O., Faith, Y., Gorman, J., Taccioli, G., & Chen, J. (1992). VDJ recombination. *Immunology Today*, 13(8), 306–314. [https://doi.org/10.1016/0167-5699\(92\)90043-7](https://doi.org/10.1016/0167-5699(92)90043-7)
- Fuentes, L., & Vallecillo, A. (2004). *Una Introducción a los Perfiles UML Behavioral types for concurrent objects View project Requirements Engineering conference View project*.
<https://www.researchgate.net/publication/266276241>
- Garcia Mariscal, A., Del Blanco, B., & Hernandez Munain, C. (2013). Generación de diversidad de los receptores de antígeno en linfocitos: validación del «modelo de accesibilidad» en el control de la recombinación V(D)J. *Inmunología*, 32(2), 57–69.
<https://doi.org/10.1016/J.INMUNO.2012.07.003>
- García Merino, A. (2011). Anticuerpos monoclonales. Aspectos básicos. *Neurología*, 26(5), 301–306. <https://doi.org/10.1016/J.NRL.2010.10.005>
- García S, A., & Casamayor, J. C. (2020). Towards the Generation of a Species-Independent Conceptual Schema of the Genome. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12584 LNCS, 61–70. https://doi.org/10.1007/978-3-030-65847-2_6/COVER
- García S, A., & Casamayor, J. C. (2021). On how to generalize species-specific conceptual schemes to generate a species-independent Conceptual Schema of the Genome. *BMC Bioinformatics*, 22. <https://doi.org/10.1186/s12859-021-04237>
- García S, A., Palacio, A. L., Román, J. F. R., Casamayor, J. C., & Pastor, O. (2020). Towards the understanding of the human genome: A holistic conceptual modeling approach. *IEEE Access*, 8, 197111–197123. <https://doi.org/10.1109/ACCESS.2020.3034793>
- Garcia S, A., Palacio, A. L., Roman, J. F. R., Casamayor, J. C., & Pastor, O. (2021). A Conceptual Model-Based Approach to Improve the Representation and Management of Omics Data in Precision Medicine. *IEEE Access*, 9, 154071–154085.
<https://doi.org/10.1109/ACCESS.2021.3128757>
- García Simón, A., Casamayor Ródenas, J. C., & Pastor López, O. (2018). *Gestión de datos genómicos basada en Modelos Conceptuales*. <https://riunet.upv.es/handle/10251/111666>
- Gea, M., Gutiérrez, L. F., Garrido, J. L., Cañas, J. J., & Aranda, S. (2003). *Teorías y Modelos Conceptuales para un Diseño basado en Grupos*.
https://lsi2.ugr.es/~mgea/invest/articulos/gea_interacc03.pdf
- Godoy Lozano, E. E. (2015). *Análisis estructural y funcional del repertorio de anticuerpos en pacientes con dengue clásico y dengue hemorrágico* [Instituto Nacional de Salud Pública].
<https://catalogoinsp.mx/bib/18073%7D>
- Grady Booch, I. J., & James Rumbaugh. (2011). Unified Modeling Language Reference Manual, The (2nd Edition). Available on: <Http://Www.Omg.Org/Spec/UML/2.1>.
https://books.google.com/books/about/The_Unified_Modeling_Language_Reference.html?hl=es&id=0a5QAAAAMAAJ

- Guarino, N., Oberle, D., & Staab, S. (2009). *What Is an Ontology?* <https://doi.org/10.1007/978-3-540-92673-3>
- Guerola Martínez, J. (2013). *Sistema de información genómico: Extensión del módulo de carga e integración de información genómica*. <http://hdl.handle.net/10251/44661>
- Gupta, A., Poels, G., & Bera, P. (2022). Using Conceptual Models in Agile Software Development: A Possible Solution to Requirements Engineering Challenges in Agile Projects. *IEEE Access*, 10, 119745–119766. <https://doi.org/10.1109/ACCESS.2022.3221428>
- Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19. : <http://aisel.aisnet.org/sjis/vol19/iss2/4>
- Insfrán, E., Pastor, O., & Wieringa, R. (2014). Requirements Engineering-Based Conceptual Modelling. *Requirements Engineering*. <https://doi.org/10.1007/s007660200005>
- Jacobson, Ivar., Booch, G., & Rumbaugh, J. (2000). *El proceso unificado de desarrollo de software*. Addison Wesley. https://www.researchgate.net/profile/Anaisa-Hernandez-Gonzalez/publication/312656269_Aplicacion_del_Proceso_Unificado_de_Desarrollo_a_proyectos_de_software/links/58878a87a6fdcc6b791ec281/Aplicacion-del-Proceso-Unificado-de-Desarrollo-a-proyectos-de-software.pdf
- Jaqu e, R., Laura, L., & Oscar, P. (2022). *Dise o de un Modelo Conceptual para caracterizar proteoma y pathways humanos: aplicaciones al metabolismo de la CoVid-19*. <https://riunet.upv.es/handle/10251/188337>
- Johannesson, P., & Perjons, E. (2014). An introduction to design science. *An Introduction to Design Science*, 9783319106328, 1–197. <https://doi.org/10.1007/978-3-319-10632-8/COVER>
- Julia Mar a Lesh er Gordillo, & Carlos Alfonso Tovilla Z arate. (2013). *Introducci n a la Gen mica*.
- Leon, A., & Pastor, O. (2021). Towards a Shared, Conceptual Model-Based Understanding of Proteins and Their Interactions. *IEEE Access*, 9, 73608–73623. <https://doi.org/10.1109/ACCESS.2021.3080040>
- Le n Palacio, A., & Pastor L pez, O. (2020). *Conceptual Modeling of Proteins Based on UniProt*. <https://riunet.upv.es/handle/10251/145619>
- Lewin, B. (2008). *Genes IX* (novena edici n). McGraw-Hill.
- Lic. Victoria Mar a Posner. (2022). *Uso de datos de transcript mica para la anotaci n de genes en el genoma de pac  (Piaractus mesopotamicus)* [Universidad Nacional de Rosario]. [oai:rehip.unr.edu.ar/2133/26076](https://oai.rehip.unr.edu.ar/2133/26076)
- Liddle, S. W. (2011). Model-Driven Software Development. *Handbook of Conceptual Modeling*, 17–54. https://doi.org/10.1007/978-3-642-15865-0_2
- Lukyanenko, R., Storey, V. C., & Pastor, O. (2022). Conceptual modelling for life sciences based on systemist foundations. *BMC Bioinformatics*, 23. <https://doi.org/10.1186/S12859-023-05287-Z>
- Mar a Jes s Lamarca Lapuente. (2013). *Ontolog as*. [Http://www.hipertexto.info/documentos/ontologias.htm](http://www.hipertexto.info/documentos/ontologias.htm).

- Martín Mayordomo, A. (2011). *Integración de bases de datos genómicas: una aproximación basada en modelado conceptual* [Universidad Politécnica de Valencia]. https://riunet.upv.es/bitstream/handle/10251/15897/TesinaMaster_AinohaMartin.pdf?sequence=1&isAllowed=y
- Martínez-Frías, M. L. (2010). Estructura y función del ADN y de los genes. Tipos de alteraciones de la función del gen por mutaciones. *Semergen*, 36(5), 273–277. <https://doi.org/10.1016/j.semereg.2009.12.014>
- Moncada C. (2019). *Análisis de datos genómicos para el diagnóstico temprano de osteosarcoma* [Centro Nacional de Investigación y Desarrollo Tecnológico]. https://rinacional.tecnm.mx/bitstream/TecNM/1365/6/MC_Carlos_Alberto_Moncada_Vazquez_2019.pdf
- Object Management Group / ISO/IEC 19510:2013. (2013). *Information technology-object Management Group Business Process Model and Notation formal*.
- Offermann, P., & Levina, O. (2009). Outline of a design science research process. *DESRIST '09: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, 1–11.
- Olivé, A. (2007). *Conceptual modeling of information systems*. 455. https://books.google.com/books/about/Conceptual_Modeling_of_Information_Systems.html?hl=es&id=RoLZWcdA4VkC
- Palacio, A. L., López, Ó. P., & Ródenas, J. C. C. (2018). A method to identify relevant genome data: Conceptual modeling for the medicine of precision. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11157 LNCS, 597–609. https://doi.org/10.1007/978-3-030-00847-5_44
- Pastor, Ó., León, A. P., Reyes, J. F. R., García, A. S., & Casamayor, J. C. R. (2020). Using conceptual modeling to improve genome data management. *Briefings in Bioinformatics*, 22(1), 45–54. <https://doi.org/10.1093/BIB/BBAA100>
- Pastor, O., López, M., & Giménez, C. (2012). *Esquema Conceptual del Genoma Humano: una herramienta para la integración y gestión de su información*. <https://riunet.upv.es/handle/10251/14390>
- Percy Edwin De la Cruz Vélez de Villa, Loayza, D. E. B., & Huamán, M. R. (2010). Documentación de Proyectos Con UML 2. *Revista de Investigación de Sistemas e Informática*, 7(1), 27–36. <https://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/3263>
- Pliego, L. R. (2009). *Recuperación y Procesamiento de Datos Genómicos para la carga de Información Clasificada en una Base de Datos Basada en un Modelo Conceptual del Genoma Humano* [Centro Nacional de Investigación y Desarrollo Tecnológico]. <https://www.cenidet.edu.mx/subplan/biblio/seleccion/Tesis/MC%20Jose%20Luis%20Rodriguez%20Pliego%202009.pdf>
- Reyes Román, J. F. (2016). *Applying Conceptual Modeling to Better Understand the Human Genome. Lecture Notes in Computer Science*, 404–412 | 10.1007/978-3-319-46397-1_31. 404–412. https://sci-hub.se/10.1007/978-3-319-46397-1_31

- Reyes Román, J. F., & Pastor López, Ó. (2018). *Diseño y Desarrollo de un Sistema de Información Genómica Basada en un Modelo Conceptual Holístico del Genoma Humano*. <https://doi.org/10.4995/THESIS/10251/99565>
- Richard Barker. (1990). *El modelo entidad-relación CASE* methodm*. https://books.google.com.mx/books?hl=es&lr=&id=hbOTo05ddxAC&oi=fnd&pg=PR11&dq=modelo+entidad+relaci%C3%B3n&ots=F8pWl42tDF&sig=JRZPmJgFVeK2mBgWWc1r_HbSw4c&redir_esc=y#v=onepage&q=modelo%20entidad%20relaci%C3%B3n&f=true
- Robinson, S., Arbez, G., Birta, L. G., Tolk, A., & Wagner, G. (2016). Conceptual modeling: Definition, purpose and benefits. *Proceedings - Winter Simulation Conference, 2016-February*, 2812–2826. <https://doi.org/10.1109/WSC.2015.7408386>
- Schroeder, H. W., & Cavacini, L. (2010). Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*, 125(2), S41–S52. <https://doi.org/10.1016/J.JACI.2009.09.046>
- Sirupurapu, V., Safonova, Y., & Pevzner, P. A. (2022). Gene prediction in the immunoglobulin loci. *Genome Research*, 32(6), 1152–1169. <https://doi.org/10.1101/gr.276676.122>
- Stephen A. White, PhD, & Derek Miers. (2009). *Guía de Referencia y Modelado BPMN*. www.integradoc.com
- Suarez Mendez, L. L. (2018). *Desarrollo del modelado de un almacén de datos como parte integral de un sistema de inteligencia de negocios*. [Tecnológico Nacional de México]. <https://rinacional.tecnm.mx/jspui/handle/TecNM/729>
- Tebes, G., Peppino, D., Rivera, M. B., & Becker, P. (2019). *Especificación del Proceso de Design Science Research: Caso Aplicado a una Ontología de Testing de Software*. <https://www.researchgate.net/publication/336562926>