

Intro. a la Bioinformatica

Nombre: Santiago Vietto

Docente: Danilo Ceschin

Institucion: UCC

Año: 2020

Introducción a la Biología (Unidad 1)

_ Toda la materia se organiza en diferentes estructuras que van desde lo más pequeño hasta lo más grande, y en ese sentido también van de lo más simple a lo más complejo.

_ Toda esta materia se organiza en diferentes niveles para poder comprender mejor la vida. Cada nivel de organización incluye niveles inferiores y van conformando a su vez niveles superiores, pero lo más importante es que todos estos niveles se caracterizan por tener propiedades que van emergiendo de un nivel a otro, esto lo llamamos **propiedades emergentes**.

Organización biológica: del átomo a la biosfera

_ Toda esta organización biológica la podemos clasificar que va desde el átomo o partículas subatómicas hasta lo más complejo como la ecosfera o biosfera. Donde las propiedades emergentes se van a ir sumando a medida que se va complejizando.

Propiedades emergentes

_ Cuando hablamos a nivel celular tenemos la síntesis de macromoléculas, división nuclear (mitosis y meiosis), donde están involucradas esas macromoléculas.

_ Si hablamos a nivel organismo tenemos estructura, función y coordinación de tejidos donde estos forman órganos y luego estos forman sistemas, donde cada uno va teniendo propiedades nuevas.

_ A nivel de población tenemos estructuras sociales, sistemas de emparejamiento, distribución por edades.

Niveles de organización de la materia

_ Podemos mencionar que hay dos grandes clasificaciones:

Nivel abiótico: significa que no tiene vida propia, donde podemos nombrar a el nivel subatómico, atómico, molecular y macromolecular.

Nivel bióticos: son los que consideramos ya que tienen vida, donde están los componentes celular, orgánico, poblacional y ecosistema.

_ Como humanos necesitamos clasificar y dividir para que se nos haga más fácil la vida, si no podemos hacer esto todo se complica.

Características de los diferentes niveles de organización

Nivel Sub-atómico y Atómico: tenemos lo que conocemos como protones, neutrones, y electrones, acompañadas de otras moléculas más pequeñas. Se da el nombre de átomo por que cuando se descubrió de qué era una partícula que no se podía dividir

(tomo: dividir,A: no se puede). El nivel atomico es formado por atomos que son la materia fundamental y estos pueden interaccionar entre ellos en reacciones quimicas y formar de esa manera moleculas.

Nivel Molecular: las moleculas son la agrupacion de dos o mas atomos, y si estas moleculas forman parte de la materia viva o son componentes celulares de un organismo se las puede denominar biomoleculas.

Nivel Macromoleculas: son las componentes claves de cualquier organismo vivo. En esas macromoleculas hay 4 principales, hidratos de carbono, lipidos, proteinas y acidos laicos. Cuando las macromoleculas se unen forman la unidad viva que es la celula.

Nivel Celula: es la parte mas pequena de vida que puede existir en el medio ambiente, en esta existe una primera clasificacion que son las eucariotas y procariotas, dentro de las eucariotas podemos mencionar las celulas animales y vegetales, ambos con ciertas carcteristicas, y en el caso de las procariotas generalmente son bacterias u organismos similares.

_ En las celulas bacterianas, estas no poseen nucleo por lo tanto el material genetico esta suelto, tienen una pared celular mas gruesa lo que les permite tener vida libre.

_ En el caso de las celulas animal y vegetal, estas si tienen nucleo donde en este esta el mayor contenido de info. genetica, y tambien hay un poco en las mitocondrias. Ambas celulas son parecidas, pero la pared celular de la vegetal es mucho mas rigida que la animal, tambien la vegetal posee una vacuola central llamada cloroplasto donde se produce todo el sistema metabolico. Estas celulas se pueden unir e interaccionar entre ellas, y formar un nivel pluricelular organico que son los tejidos.

_ Unidad estructural y funcional de toda forma de vida.

Nivel Tejidos: a los tejidos los definimos como un conjunto de celulas que tienen la misma funcion y que tienen el mismo origen embriologico.

- Tejido muscular: existen a su vez 4 tipos de tejidos que son el tejido conectivo, epitelial, muscular y nervioso.

_ Cuando estos tejidos se juntan forman los organos.

Nivel Organos: son tejidos que actuan conjuntamente, y cuando estos organos se agrupan y empiezan a cumplir funciones similares dentro de un mismo grupo de funciones, forman lo que conocemos como sistemas y/o aparatos.

Nivel Sistemas y/o aparatos: el conjunto de organos que pueden ser muy diferentes entre si pero que pueden tener funciones estrechamente coordinadas, a eso le llamamos aparatos, y los sistemas clasifica a los organos mas o menos similares que tienen la misma funcion. Todo esto es lo que llamamos organismo.

Nivel Organismo: donde definimos el árbol de la vida donde se supone que hubo un primera y única célula y por diferentes cambios evolutivos se generaron todos los organismos que conocemos y no en la actualidad.

Nivel Poblacional: cuando tenemos un grupo de organismos de una misma especie que se juntan en un mismo lugar.

Nivel Comunidad: la comunidad básicamente es un conjunto de poblaciones distintas, ya no es la misma especie sino que son especies diferentes que están en un mismo espacio y muchas veces establecen relaciones entre ellas.

Nivel ecosistemas: es en el que se incluyen tanto el conjunto de poblaciones que viven interrelacionadas a lo cual llamamos comunidad e interactúan con el lugar en el cual terminan viviendo, lo conocemos también como biotopo.

Nivel biosfera: conjunto de ecosistemas tanto marinos como terrestres que están integrados en la superficie del planeta.

Clasificación de los seres vivos

Biodiversidad: es enorme, se cree que en nuestro planeta presenta gran variedad de seres vivos, alrededor de 5 a 10 millones de especies. A esta biodiversidad necesitamos clasificarla por eso se necesita de un sistema de clasificación ya que este provee una forma conveniente de no perder de vista a todas las formas de vida conocidas, a este sistema lo llamamos **sistémica** y es el estudio científico de la diversidad de organismos y de sus interrelaciones, y trata de interpretar la diversidad orgánica.

Sistémica

_ Consta de 3 divisiones:

Taxonomía: es el estudio teórico de la clasificación, tiene como objetivo descubrir y describir a las diferentes especies o grupos de especies (=taxa).

Filogenia: es el área que trata de establecer que relaciones o parentesco hay entre las diferentes especies.

Clasificación: lo que busca es ordenar las especies de acuerdo a su filogenia.

Objetivos:

1. Reconocer, describir y dar nombres a las especies y taxones.
2. Reconstruir la historia evolutiva de los grupos de organismos.
3. Plantear hipótesis sobre el origen y evolución de los grupos de organismos.
4. Construir clasificaciones de los taxones con alto valor explicativo.

5. Proporcionar información para desarrollar investigaciones en otras áreas de la biología comparada.

Historia:

Aristoteles: fue quien primero empezó a dar una clasificación que se dividió entre el reino mineral, animal y vegetal, y además subdividió a los animales según el hábitat. Fue el primero que introdujo el término de especie.

Teofrasto (discípulo de aristoteles): amplió el concepto y desarrolló un sistema para clasificar las plantas según como crecían (hierbas, arbustos, árboles). Y además introdujo la idea de la clasificación basada en similitud de estructuras, en algunos casos esta es válida pero en otras no.

Carl von Linné: fundador de la taxonomía moderna, basada en 7 categorías, y asignó cada organismo al reino animal o al reino vegetal, donde también subdividió cada categoría en categorías más pequeñas. Y en ese sentido la especie era o es la unidad básica del sistema de clasificación.

_ El se basaba en las similitudes de la estructura del cuerpo.

Especie:

_ Grupo de organismos de una clase en particular, estrechamente relacionados, que comparten características fenotípicas (lo que uno ve) y genotípicas (info. genética) y que pueden entrecruzarse y producir crías fértiles.

Sistema Binomial

_ Desarrollado por von Linné para darle nombre a los seres vivos, utilizando dos palabras en latín, ej: Homo sapiens para el humano. Esta posee ciertas reglas y ventajas que se describen en el pdf.

Taxonomía

_ Ciencia que estudia los principios, métodos y fines de la clasificación de los seres vivos. Reglas y leyes que sirven para ordenar de manera sistemática y poder darle un nombre científico a todos los seres vivos.

Categorías Taxonómicas

_ Esas categorías taxonómicas de los seres vivos se basa en la constitución de un sistema jerárquico de grupos dentro de grupos, donde el grupo jerárquico más grande o importante sería el dominio, luego el reino, filo/división, clase, orden, familia, género, a especie.

Características empleadas para clasificar organismos

_ Cuadros pdf.

_ Estas fueron cambiando a través del tiempo.

Cromista: algas que pueden hacer fotosíntesis.

Tipo de información

_ Dependiendo del abordaje de estos niveles es el tipo de información que voy a tener:

Nivel ecosistemas: se estudian comportamientos y dinámica de las especies o poblaciones, etc.

Nivel población: aprender cómo esa población se va a colonizar, depreda, reclutar o reproducir, etc.

Nivel individuo: podemos estudiar crecimiento, desarrollo, movimiento, etc.

Nivel funciones fisiológicas: estudiar si esos individuos van a tener adaptación, cuál es su metabolismo endocrino, qué respuesta inmune tienen, su función reproductiva, etc.

Nivel parámetros moleculares: podemos estudiar genes, transcritos, proteínas y metabolitos.

_ Dependiendo en el lugar que me encuentre dentro de estos niveles cuanto más arriba estoy tengo una visión o perspectiva ecológica, y cuanto más abajo estoy voy a tener una perspectiva mecanística, pero poca de las contrarias.

¿Cuáles son organismos modelos?

_ Se les llama organismos modelos generalmente a los organismos que se utilizan en laboratorios para estudiar diferentes cosas desde fisiología, funciones a nivel celular, nivel mecanístico, como se presentan los genes, ante un determinado estímulo como se comporta ese organismo. Para todos estos hay herramientas como la disponibilidad del genoma secuenciado está completo.

_ Los organismos no modelos, son organismos que muchas veces se utilizan para algunos experimentos, pero en general la cantidad de info. o herramientas que hay respecto a esos organismos es menor.

Macromoléculas biológicas (biomoléculas)

_ Este nombre se les da por que también tenemos macromoléculas orgánicas que no son consideradas biológicas.

_ Estas son como los ladrillos que forman un organismo vivo, como mencionamos antes de ellas las cuatro principales son los hidratos de carbono, los lípidos, las proteínas y ácidos nucleicos.

Carbohidratos (hidratos de carbono)

_ Estas moléculas o compuestos están formadas principalmente por átomos de carbono, hidrógeno y oxígeno. Son los que también conocemos como glucidos y son la primera fuente de energía para la célula.

Clasificación:

Simples:

- Monosacáridos:
 - Glucosa: principal fuente de energía a nivel eucariota.
 - Fructuosa
 - Galactosa
- Disacáridos:
 - Sacarosa
 - Maltosa
 - Lactosa

Complejos:

- Polisacáridos:
 - Glucógeno: (en los animales) fuente de reserva de glucosa.
 - Celulosa: (en plantas).

Función:

_ Aportan energía al organismo, y tienen la característica de que el organismo de estos no genera residuos tóxicos como cuando hay un catabolismo de proteínas.

Metabolismo:

_ Hablamos de metabolismo cuando hablamos de funciones que están implicadas en degradar la glucosa para posteriormente formar dióxido de carbono y agua. El metabolismoposee dos procesos:

Catabolismo: cuando estamos degradamos alguna molécula para obtener o moléculas más simples y energía de las mismas.

Anabolismo: cuando estamos creando esas moléculas. Los humanos no creamos glucosa (no hay anabolismo de glucosa), pero sí hay anabolismo del glucógeno.

Glucógeno:

_ Es la sustancia de reserva en el cuerpo. Se genera en el hígado a partir de la glucosa; cuando esta glucosa se empieza a acumular se produce el anabolismo a glucógeno y de ahí puede saltar al anabolismo de lípidos.

Lípidos

_ Podemos decir que son muy importantes en la alimentación. También son una fuente importante de reserva energética y además es componente de muchas estructuras celulares principalmente las membranas biológicas, y además de tener una participación como precursores de hormonas.

_ Estos también tienen carbono e hidrógeno en su composición, y tienen además otros elementos en menor grado como el nitrógeno, fósforo y azufre.

Clasificación:

Ácidos grasos: su subclasificación tiene que ver con la presencia o no de dobles enlaces

- Saturados: no hay dobles enlaces en el lípido de la molécula, y estas son flexibles, pueden rotar y generalmente son sólidos a temperatura ambiente (esta depende del largo).
- Insaturados: hay presencia de dobles enlaces y estos hacen que los lípidos sean rígidos y no se puedan rotar ni mover libremente en el lugar, y a temperatura ambiente son aceitosos.

Lípidos complejos: aquellos que van a interactuar con otras macromoléculas del organismo

- Lipoproteínas
- Glucolípidos

Glicéridos:

- Glicéridos neutros:
 - Triglicéridos
- Fosfoglicéridos: (fosfolípidos), son lípidos que están interactuando con un grupo fosfato, y que están presentes principalmente en las membranas y esto tiene que ver con las propiedades de los lípidos.

Lípidos sin glicerol:

- Esfingolípidos

- Esteroides: el precursor de estos es el colesterol, y tiene una función vital en lo que es la composición de las membranas y la rigidez de las paredes celulares. Si hay poca concentración de colesterol en la membrana, esta es muy fluida, y si hay altas concentraciones, también. Este es precursor de la bilis (a nivel de intestino nos ayuda a digerir las grasas), y es precursor de hormonas esteroideas.
- Ceras
- Terpenos

Característica:

_ Los lípidos poseen una característica que es la hidrofobicidad, es decir que tiene rechazo a lo acuoso, entonces por más que se mezclen con agua no van a formar una solución acuosa. Cuando se unen los lípidos pueden formar micelas, monocapas y bicapas, que son uno de los componentes principales de las membranas celulares.

Funciones:

- _ Son una gran reserva energética, por gramo son 9kcal (kilocalorías).
- _ A nivel celular forman las membranas celulares y las vainas de mielina en los nervios.
- _ Además transportan proteínas que son liposolubles.
- _ Esto son los que le dan sabor y textura a los alimentos.

Proteínas

_ Las proteínas están compuestas por aminoácidos que son la unidad más simple que las constituyen, de las cuales no solo en el humano sino que para la mayoría de los seres vivos son 20, de los cuales 9 de ellos son esenciales aportados por la dieta en el caso del humano.

_ Dentro de los aminoácidos tenemos los proteicos, canónicos o naturales que son los que están codificados en el genoma.

Aminoácidos:

_ Estos tienen una característica en común en cuanto a su estructura esta formada por monómeros que forman polímeros, donde las proteínas directamente es un polímero de aminoácidos, que por un lado poseen un grupo amino y por el otro un grupocarboxilo, además de tener un radical/residuo R, dependiendo del residuo que tenga va a ser el aminoácido que vamos a tener.

Enlace peptídico: enlace característico de las proteínas que se da entre el grupo carboxilo de un aminoácido con el grupo amino de otro aminoácido. Este enlace no es espontáneo sino que tiene que ser catabolizado por una enzima. Entonces dependiendo de cómo realice el enlace obtengo los diferentes tipos de proteínas.

Estructura:

_ Constituida en cuatro niveles:

Primaria: es la secuencia lineal de los monómeros, de los aminoácidos de las proteínas.

Secundaria: es la disposición de la secuencia de aminoácidos en el espacio.

Terciaria: forma tridimensional que va a tomar esa proteína o cadena polipeptídica.

Cuaternaria: cuando se pueden unir varias estructuras terciarias en una proteína más compleja. Complejo proteico.

Funciones:

Anticuerpos: proteínas que genera el cuerpo como defensa.

Proteínas de transporte: de transporte de iones u otras moléculas como lípidos, hormonas, citoquina es un ejemplo.

Proteínas estructurales: principalmente las queratinas y colágenos (la matriz extracelular está formada por todo tipo de colágenos), ejemplos: uña, piel.

Proteínas del movimiento coordinado: el músculo.

Ácidos nucleicos

_ Son aquellos que desde el primer ácido nucleico formado (ARN), este se fue uniendo formando estructuras más complejas, donde luego fue el ADN y se fueron generando los organismos unicelulares y así siguiendo.

_ Los ácidos nucleicos son macromoléculas, que también son polímeros pero a diferencia de las proteínas los monómeros que forman los ácidos nucleicos son los nucleótidos, y la unión por la cual se unen es fosfodiéster.

_ Estas macromoléculas pueden alcanzar tamaños enormes.

_ Los dos que vemos son ADN y ARN.

Nucleótidos: están formados por un monosacárido de 5 carbonos y un grupo fosfato. Ese monosacárido en el caso del ADN es una desoxiribosa (sin oxígeno) y en el caso del ARN es una ribosa (hexitol). Esta pentosa (monosacárido+grupo fosfato) está interactuando con una base nitrogenada que se clasifica en:

- Purinica:
 - Adenina
 - Guanina

- Pirimidinica:
 - Timina
 - Citocina
 - Uracilo(ARN)

_ Estos pueden interaccionar entre si, y siempre interaccionan guanina con citocina y timina con adenina.

Dogma central

_ Es el pasaje del ADN a traves de la transcripcion a las moleculas de ARN mensajero, y de esa la traduccion a proteinas. La palabra dogma implica algo que es asi, no se puede modificar. Hoy en dia este cambio con otras funcionalidades.

Transcripción del ADN:

_ Es una expresion genica, es decir, leer esa secuencia de ADN, interpretarlo y generar el ARN mensajero. Esto va a depender de que tan empaquetada este la cromatina ya que esta va a estar modulando la transcripcion.

_ Siempre la transcripcion es de 5' a 3'.

_ En este proceso y en el de replicacion se transcriben las dos hebras pero dependiendo del sentido que tienen es como van a estar regulados los genes.

Proceso post-transcripcional:

_ Tenemos lo que se llama pre-ARN mensajero o Transcripto primario, y necesitamos procesarlo. El ADN tiene una region que se llama 5' UTR, y el pre mRNA va a estar formado de exones e intrones.

Intrones: no tienen info. para codificar una proteina. E interrumpen la secuencia que es valida para el ARN mensajero.

Exones: son los que codifican las proteinas.

Traduccion de proteinas:

_ Estan involucrados en el ARN mensajero. Ocurre en el citoplasma y en el reticulo endoplasmatico rugoso.

Regulacion de la expresion genica:

_ Puede ser dada a nivel del ADN o del ARN mensajero. A nivel del ADN lo que puede regular la replicacion es si esta disponible o menos disponible ese ADN y eso va a estar dado por que tan compacto o no este ese ADN. A nivel de ARN mensajero lo que puede regular la transcripcion son otros ARM reguladores.

Epigenetica: por fuera de lo que es la secuencia genética. No tiene que ver con la secuencia del ADN.

ADN

_ Descripto en 1953 por Watson y Crick, pero Rosalind Franklin interpreto otro sistema mediante rayos X.

_ Este es la informacion genetica donde se guarda toda la info. que esta basicamente en el nucleo, y esta se va transmitiend de generacion en generacion.

_ La estructura secundaria del ADN en doble elices, esta estabilizada por un monton de interacciones entre el fosfato, la pentosa y los nucleotidos. Hay muchas fuerzas que los unen:

Puentes de hidrogeno: uniones que se dan entre un hidrogeno y un oxigeno por ejemplo, interacciones de tipo electristaticas por las cargas.

Interacciones hidrofobicas: es una de las fuerzas mas potentes dentro de la estructura del ADN.

Interacciones electrostaticas: que tienen que ver con los fosfatos y con iones que esten

_ Tenemos la estructura de doble elice y ademas tenemos porciones del genoma que toman una estructura cuádruple (doble doble elice) que no es muy comun.

Organización del ADN:

_ El ser humano tiene 3000 millones de nucleotidos que estan distribuidos entre los 23 pares de cromosomas. El ADN forma estructuras que va desde la doble elice, se va enrollando en rol de proteínas llamadas histonas donde estas se van enrollando en una cromatina mas condensada, se van empaquetando hasta que forman los cromosomas. Dependiendo del grado de empaquetamiento esa cromatina va a estar disponible para que se pueda o no leer el codigo genetico.

ARN

_ Este es un intermediario entre el ADN y las proteínas que va a generar. Es una secuencia complementaria al ADN, es decir, que tiene una secuencia que complementa al ADN.

_ Leer cuadro comparativo

_ El ARN como es un monomero, tiene una estructura:

Primaria: secuencia lineal de esos nucleotidos.

Secundaria: pliegues que forma el ARN que pueden ser elices o bucles, y eso es por que la cadena se puede plegar y encontrar bases complementarias y toman estructuras tridimensionales.

Terciaria: diferentes apilamientos o pliegues en el espacio.

_ ADN hay uno solo pero moleculas de ARN hay muchas, las primeras que se conocieron fueron tres, el ARN ribosomal, de transferencia, y mensajero, pero al pasar el tiempo se fueron descubriendo otros como ARN pequeños nucleares, interferencia, micro ARN, los largos no codificantes, y estos van a cumplir diferentes funciones en la celula en la que estan.

ARN implicados en la sintesis de proteinas: mensajero, ribosomico y de transferencia.

- ARN Mensajero: su estructura es lineal donde puede formar horquillas, se sintetiza en el nucleo (transcripcion), donde esta secuencia es complementaria a la del ADN, y muchas veces pasa al citoplasma donde se produce la sintesis de proteinas y a ese procesos se le llama traduccion.
- ARN Ribosomico: es el mas abundante (80%), y tiene una conformacion globular. A partir del precursor de este voy a tener diferentes subunidades de este. Las bases nitrogenadas estan heladas y entonces puede formar estructuras mas complejas. Este se une con otros complejos de proteinas y forma subunidades.
- ARN Transferencia: formado alrededor de 80 nucleotidos teniendo una estructura secundaria de tipo hoja de trebol, la particularidad de esta estructura es que posee un anticodon que son 3 nucleotidos complementarios a 3 nucleotidos del ARN mensajero y cada uno va a codificar para un aminoacido diferente. Hay 50 tipos diferentes pero tienen características similares, tienen un brazo aceptor (brazo unido a un determinado aminoacido y este depende del anticodon), tambien un extremo (5') que es un triplete de guanina y citosina que se unen a un fosforo.

_ Entre los tres van a estar vinculados al mecanismo de traduccion que es la lectura del RNA mensajero y la sintesis de proteinas.

ARN precursores: (nucleolar y heterogeneo-nuclear), ARNs que se originan en el ADN, y se agrupan en una region que se llama region organizadora nuclear que son basicamente los nucleolos y a partir de esto obtenemos los diferentes tipos de ARN ribosomicos.

ARN reguladores: se llaman asi por que van a estar involucrados a regular la expresion de proteinas

- ARM Antisentido: se descubrieron primero en las plantas y luego en los animales y en la actualidad se hacen terapias antisentido para utilizar estos ARM para regular la expresion de determinados genes.

- ARM Interferencia: van de 19 a 23 nucleotidos, y hay 3 tipos, Micro ARN, ARN interferente pequeño y ARN asociados de Piwi (proteína). También hay algunos que van de 20 a 30 nucleotidos que suprimen la expresión de genes específicos (controlan los niveles proteicos).

GEN

_ Se considera gen a la molécula de ADN que me va a dar una proteína finalmente. Este gen está formado por una secuencia de nucleótidos y hay diferentes variaciones en las secuencias, que son las formas alélicas y puede estar sujeto a mutaciones.

_ En la actualidad la definición de gen está más relacionada a cualquier porción del genoma que se expresa y que se transmite de manera hereditaria.

GEN

¿Qué es un gen?

Tramos activos de la molécula de ADN ordenados a lo largo de los cromosomas.

- Unidad de estructura (no divisible por entrecruzamiento);
- Unidad fundamental de cambio (entre formas alélicas y sujeto a mutaciones y recombinación);
- Un par de nucleótidos ---
- **Unidad básica de función** (un gen – una proteína).

“Todo segmento de DNA que se encuentra luego de un promotor y que puede ser transcrito por una ARN polimerasa y originar un ARN funcional (ARNm, ARNr, ARNt, snARN, ribozina, etc”.

Variación genética: Variabilidad en el material genético de una población o especie. La presencia de varios alelos para un gen permite el proceso de selección natural y la evolución.

Fuentes de variación: mutaciones y combinaciones de genes (meiosis y fecundación).

Código genético:

_ Definiciones de la filmin y tabla.

Código Genético

“Conjunto de reglas que define la traducción de una secuencia de nucleótidos en el ARN a una secuencia de aminoácidos en una proteína”.

- ❖ Consta de **64 codones** o tripletes de bases, 61 codifican aminoácidos y 3 funcionan como **señales de terminación**.
- ◆ **No es ambiguo**, cada codón especifica a un solo aminoácido.
- ◆ **No es solapado**, un nucleótido pertenece a un único triplete.
- ❖ Su lectura es **“sin comas”**, no existen espacios en blanco.
- ◆ **Es degenerado**, un aminoácido puede estar codificado por diferentes codones.
- ❖ Es universal, sus mensajes son interpretados de la misma forma por todos los organismos.
- ❖ Utiliza un marco de lectura establecido al inicio de la traducción y no lo modifica.

		Segunda base					
		U	C	A	G		
Primera base	U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U	Termino
		Phe UUC	Ser UCC	Tyr UAC	Cys UGC	C	
		Leu UUA	Ser UCA	Stop UAA	Stop UGA	A	
		Leu UUG	Ser UCG	Stop UAG	Trp UGG	G	
	C	Leu CUU	Pro CCU	His CAU	Arg CGU	U	
		Leu CUC	Pro CCC	His CAC	Arg CGC	C	
		Leu CUA	Pro CCA	Gln CAA	Arg CGA	A	
		Leu CUG	Pro CCG	Gln CAG	Arg CGG	G	
	A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U	
		Ile AUC	Thr ACC	Asn AAC	Ser AGC	C	
		Ile AUA	Thr ACA	Lys AAA	Arg AGA	A	
		Met AUG	Thr ACG	Lys AAG	Arg AGG	G	
	G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U	
		Val GUC	Ala GCC	Asp GAC	Gly GGC	C	
		Val GUA	Ala GCA	Glu GAA	Gly GGA	A	
		Val GUG	Ala GCG	Glu GAG	Gly GGG	G	

Introduccion a la Bioinformatica (Unidad 2)

Definiciones de Bioinformatica

_ Esta fue cambiando a lo largo del tiempo, y ademas fue cambiando dependiendo de la disciplina por la cual uno ingresa a hacer bioinformatica.

_ Biblio: "Instant Notes in Bioinformatics" – T. Charlie Hodgman, Andrew French y David R. Westhead.

Definicion Antidiluviana: (termino frances), estaba basada en términos de datos biológicos, a menudo secuencias de macromoléculas (ADN, ARN, Proteínas), y computadoras. Ryback en su libro utiliza 3 capítulos para describir:

- Forma quiral de aa (amino acidos), azúcares y metabolitos. Forma quiral es cuando tengo una molecula que puede polarizar una luz, dicho de otra manera, una molecula tiene un reflejo de si misma y son las formas quirales.
- Orden en la secuencia de bases, aa y azúcares.
- Describia algunos tipos de moléculas, organelas celulares, tipos celulares y en organismos, tipos de tejidos y órganos.
- Dio definiciones de densidad y distribución de organismos.
- Y se hablaba de señalización dentro de los organismos: hormonas, impulsos nervios, color epidérmico, sonidos, tacto.

_ Con estos tres capitulos, en 1978 Ryback diseño el primer curso de "bioinformática" que fue brindado a lo largo de una década.

Definicion Real: por una rivalidad entre ingleses y franceses

Definicion canonica: es decir, todo lo que es canonico es aceptado. Donde la Bioinformática es la disciplina con interface entre Biología, ciencias de la información y matemática, entonces con la interseccion de estas tres diciplinas podemos decir que se forma la bioinformatica. Esta es una disciplina intrínsecamente multidisciplinaria. Y los bioinformáticos evalúan y producen nuevos conocimientos sobre los procesos biológicos a través del análisis digital de la información biológica.

_ En esta area se debe saber lo siguiente:

- Nadie puede ser competente en todos los aspectos. Esto es algo que se repite en otras diciplinas.
- Los bioinformáticos pretenden mantenerse al día en una disciplina, aunque están lo suficientemente familiarizados con los demás para actuar como intérpretes entre especialistas en un dominio y otro. Debemos estar familiarizados con los temas.

- Idealmente se debe trabajar como parte de un equipo que aporta diferentes conocimientos para abordar un problema determinado.
- Las novedades en bioinformática provienen de tres fuentes:
 - Matemáticas y algoritmos bien establecidos para ser aplicados en una nueva categoría de datos biológicos.
 - Desarrollo de nuevo “hardware” o interfase para la investigación en las Ciencias Biológicas.
 - Nuevas técnicas matemáticas que resuelvan “cuellos de botellas” en las investigaciones existentes.
- Biólogos van detrás del “Por qué” un proyecto es importante.
- Informáticos son más dirigidos a “Cómo” un algoritmo o pieza de software puede ser aplicada.
- Matemáticos definen mejor “Cuál” es la mejor algoritmo o técnica para analizar datos.
- De esta manera el balance entre las diferentes disciplinas debe mantenerse.

Definición funcional: es una definición que explica que la bioinformática busca generar conocimiento de las propiedades, poblaciones y procesos de entidades (biológicas).

Definición para el público: es una definición que explica que la es la aplicación de la computación y las matemáticas a la gestión, el análisis y la comprensión de los datos para resolver preguntas o cuestiones biológicas relacionadas con la medicina, quemo-, neuro-, etc, informática. Esto quiere decir que:

- “La bioinformática es la aplicación de la computación y las matemáticas...”, tiene que ver la definición canónica.
- “...a la gestión, el análisis y la comprensión de datos...”, se refiere a como la bioinformática se ocupa de todo el proceso desde la captura inicial de los datos, su administración en bases de datos, el análisis de estos datos y la formulación de los resultados en un contexto que da como resultado una nueva comprensión genuina. El objetivo está impulsado por la cantidad y calidad de los datos.
- “...para resolver preguntas/cuestiones biológicas...”, es decir, que la bioinformática se aplica a problemas biológicos y no puramente cuestiones de informática, ya que lo que hagan los bioinformáticos va a estar relacionado con las ciencias de la biología.
- “...relacionadas con la medicina, quemo-, neuro-, etc, informática.”, donde los desarrollos matemáticos e informáticos que se han aplicado a datos en otras áreas del conocimiento también pueden ser útiles en bioinformática.

_ Nosotros nos quedamos con esta definición, la bioinformática es una disciplina, no es un apéndice de otras de otras disciplinas, sino que es por sí misma, que si se alimenta de otras áreas para poder llevar adelante los objetivos.

_ Estudiar guía 2.1 y 2.2.

Base de datos y Fuentes de Datos (Unidad 3)

Biología, Bioinformática y Biología Computacional

_ La biología computacional es a veces definida como sinónimo de Bioinformática y a veces como una disciplina emparentada, pero distinta, de esta. El NIH (National Institutes of Health) define a ambas disciplinas como distintas aunque con cierto grado de solapamiento, según esta definición la bioinformática está más relacionada con el desarrollo de herramientas computacionales con el fin de analizar y procesar datos, y la biología computacional con el estudio por medios computacionales de sistemas biológicos, la biología computacional está más relacionada a la simulación o modelización de procesos biológicos más que el análisis de datos generados por experimentos a gran escala.

BIG DATA

_ La cantidad de datos que se pueden generar teniendo en cuenta el dogma central de la biología y los diferentes aspectos que tiene este, hay algo a lo que llamamos big data por que tenemos una gran cantidad de datos, esta tiene un punto de vista distinto a lo que se ve en programación.

_ Llamamos big data por que tiene en cuenta ciertas características de lo que es el big data que es totalmente aplicable a las bases de datos, donde estas bases, que tienen que ver con la biología o las ciencias de la vida (todo lo que tenga que ver con cuestiones biológicas), hay de diferentes tipos de análisis.

_ El PDB fue la primera base de datos de bioinformática donde uno puede encontrar estructuras de proteínas y modelos tridimensionales de proteínas.

_ De las bases de datos uno puede sacar información, trabajarla y puede relacionarla entre sí. Por ejemplo en el esquema de la filmina tenemos una línea celular de cáncer y podemos buscar en diferentes bases de datos relaciones de información y protocolos para tratar de construir conocimiento.

_ Uno puede hacer síntesis de lo mencionado y tener por ejemplo bases de datos donde tenemos drogas y como actúan en organismos que están genéticamente modificados, bases de datos donde tenemos factores de transcripción y que genes están regulando esos factores, bases de datos de modificaciones de histonas (que regulan la expresión de genes), etc, y así podemos integrar donde podemos tener Gene-set library que son listas de genes que se van a representar algún proceso biológico que puede ser fisiológico o patológico, también tenemos los Bi-partite graphs que son gráficos de relaciones entre dos entidades y sirven para modelar procesos y construir Networks donde tenemos nodos

que van a ser entidades y las líneas que los conectan representan relaciones o información, haciendo un análisis topológico de las redes uno puede detectar por ejemplo que nodos son más importantes que pueden servir para construir hipótesis que luego las probamos en el laboratorio.

_ Como decíamos todo esto forma parte de lo que llamamos big data básicamente por que tomamos del big data lo que llamamos 4V, donde esta tiene que ver con características que van a tener los datos:

Volumen: es la escala de los datos que yo voy a estar guardando en la base de datos. Vamos a tener bases de datos que van a tener archivos pequeños con coordenadas por ejemplo de alguna estructura de una proteína, o pueden ser archivos enormes como los archivos de secuenciación que son archivos planos de texto que tienen gigabytes de tamaño y más grandes si son archivos de imágenes.

Variedad: tenemos la variedad de los datos. Tenemos bases de datos donde vamos a tener imágenes, secuencias. O sea va a ser los diferentes formatos de datos que voy a utilizar en las bases de datos.

Veracidad: acá tiene que ver mucho el tipo de tecnología que voy a emplear para generar los datos y tiene que ver muchas veces en como se procesan esos datos. En bioinformática hablamos de datos curados (verificados) y no curados.

Velocidad: con la cual voy a poder acceder a los datos y el análisis que voy a poder hacer de los mismos.

Generación de datos

_ Metodología experimental para el estudio de las diferentes moléculas en experimentos clásicos y a gran escala. Volviendo al dogma central de la biología, los tipos de metodologías que se utilizan para estudiar el ADN son Southern blot (electroforesis) y Secuenciamiento de ADN (método de Sanger). En cuanto al ARN las metodologías son similares a las de ADN y son Northern blot donde en vez de detectar secuencias de ADN podemos detectar secuencias de ARN, y también tenemos PCR (reacción en cadena de polimerasa), qPCR, donde estas amplifican el material de algo. En cuanto a las proteínas tenemos los métodos Western blot y sec. de Edman, donde la secuencia es la misma pero el revelado de las proteínas las hacemos mediante anticuerpos específicos que interactúan con las proteínas que quiero estudiar.

_ Estas metodologías fueron diseñadas para estudiar una sola entidad, en el caso de Southern blot estudio un locus del genoma, en el caso de Northern blot y los PCR estudiamos un determinado gen, y con Western blot estudiamos una proteína. Estas no son a gran escala.

La era “-ómica”/“-oma”

_ A partir de lo anterior viene esto, que tiene que ver con el estudio a gran escala de los experimentos y los procesos biológicos. Entonces si es a nivel de la información genética del ADN y a nivel global vamos a hablar de genoma, si queremos estudiar los transcriptos o los genes a gran escala voy a hablar del transcriptoma, si queremos estudiar las proteínas a gran escala vamos a hablar del proteoma y así con cualquier entidad que yo quiera estudiar de un sistema biológico le agrego “oma” y ya nos damos la idea de que es a gran escala, otros casos son lipidoma (lípidos), metaboloma (metabolitos), degradoma (encimas que degradan matriz extracelular), y así sucesivamente. Cuando yo quiero aprender de todas estas metodologías es cuando surge el concepto de system biology.

System biology

_ Esta es el estudio de manera olistica de los métodos anteriores para tratar de entender un proceso biológico. Este sistema para poder analizar la cantidad de datos obtenidos por técnicas moleculares a gran escala debemos tener un enfoque multi o interdisciplinario que involucran diferentes disciplinas (informática, matemáticas, estadística, química, física, ingeniería, lingüística) donde la bioinformática y biología computacional van a estar involucradas y se van a encargar de analizar todos los datos y experimentos a gran escala.

_ En cuanto al dogma central de la biología, existen metodologías que son los Microarreglos (MicroArrays) que actualmente están en desuso y son chips que tienen secuencias de todos los genes y al poner mi muestra en los chips, podemos detectar la expresión de los diferentes genes, pero ahora se utiliza la Nueva Generación de Secuenciación masiva que es secuenciar directamente todo el transcriptoma por ejemplo, los genes tienen exones e intrones, y el exoma sería secuenciar todos los exones de un individuo. Los microarreglos y la secuenciación masiva también es aplicable a ARN o a todos sus tipos, y en el caso de proteínas también hay microarreglos pero caso no se utiliza, entonces lo que se usa es una metodología que se llama espectrometría de masas que lo que hace es poder estudiar, identificar y hasta cuantificar las diferentes proteínas en una mezcla de proteínas.

_ Todas estas metodologías son el caballo de batalla de la era ómica y son metodologías a gran escala que me van a generar datos que luego voy a poder utilizar en el laboratorio.

_ Visto desde el punto de vista de para que sirven y para que no las metodologías:

ADN: las metodologías Southernblot, Secuenciación de Sanger, NGS (nueva generación de secuenciación) y Microarrays pueden ser aplicadas en organismos modelos, pero tanto Southernblot como Microarrays en organismos no modelos no podemos hacer estudios a gran escala, pero con las otras dos sí.

ARNm: para este caso es casi lo mismo, las metodologías Northern blot, PCR, qPCR, Microarrays y NGS pueden ser aplicadas a organismos modelos, pero Northern blot, PCR, qPCR y Microarrays no pueden ser usadas a gran escala en organismos no modelos, a excepción de NGS.

Proteínas: con respecto a las metodologías Western blot, secuenciación de Edman, Microarrays y 2D-espect. de masas pueden ser aplicadas en organismos modelos, pero tanto Western blot como Microarrays no pueden ser aplicados a gran escala en organismos no modelos, a diferencia de los otros dos.

_ Entonces tenemos ciertas metodologías que me van a servir para estudiar sabiendo o no la secuencia del genoma o de un determinado gen, y hay otras metodologías que me van a estar limitadas a el conocimiento que yo tenga en ese momento sobre una determinada entidad.

Historia de la Secuenciación de ADN

_ La secuenciación de ADN es una de las herramientas que hizo que explotara la bioinformática. Analizamos el esquema de la filminá.

_ En 1965 se pudo secuenciar el tRNA para alanina y la eficiencia de este era de 1 (bp/person/year), o sea un nucleótido por persona por año. Luego esta metodología fue mejorando y se podían estudiar cadenas más largas, pero un gran salto significativo con el desarrollo de Sanger en 1977. A partir de 1980 estaba mucho más desarrollado además de la polimerasa lo que eran las enzimas de restricción que son enzimas con la capacidad de cortar la secuencia de ADN (en cualquier lugar o en secuencias específicas). En 1990 donde ya había una gran eficiencia, se propone secuenciar el genoma humano a un costo económico muy alto, y en 2002 se publicó el primer borrador del genoma humano (borrador por que si bien había sido secuenciado mucho genoma humano, había otras partes donde habían huecos donde no se conocía la secuencia). En el 2005 y 2006 aparece una empresa que revolucionó con lo que se llama secuenciación masiva con una eficiencia de 100,000,000,000 de bases, al medida que pasó el tiempo se fue mejorando y en 2013 aparece una nueva generación de secuenciación que se denomina la 3ra generación que si bien la secuenciación de Sanger había que clonar en vectores, en la 2da generación no se clonaba en vectores, se amplificaba el material, pero en la 3ra generación llamada SMS por single molecule sequencing, puede secuenciar una sola molécula de ADN o ARN sin la necesidad de amplificarla pudiendo saltar a 500,000,000,000 de bases, y eso hace que cuando el primer genoma humano se tardó 12 años, en la actualidad se puede secuenciar el genoma humano en 8 horas a un costo de \$1000 USD o mucho menos.

Microarreglo

_ Nos imaginamos una sola posición y esta tiene una identidad. Tenemos un chip con oligonucleótidos que tienen una secuencia de ADN o ARNm, y esto tiene una identidad A1, A2, etc (matriz).

_ Entonces esta metodología consiste en, tenemos las células, purificamos el ARNm si es un microarray para ver la expresión de genes donde el ARNm es más inestable que el ADN entonces para evitar que se degrade hacemos una transcripción reversa donde transformamos el ARNm en ADN copia (ADNc problema), este ADNc lo marcamos con un color (rojo o verde), luego ponemos a hibridizar en el chip (encubar mi chip con la muestra problema), lo dejamos por un tiempo y al lavar vamos a ver unas secuencias que se van a pegar a los oligonucleótidos que tengo en el chip porque son complementarios, luego irradiamos esto con un láser y vamos a tener una mancha (captura de imágenes).

_ La manera de construir el microarray puede ser de diversas maneras, que se fue mejorando con el tiempo.

Secuenciación de Sanger

_ Esta se basa en, tenemos un nucleótido que tiene la ribosa (azúcar) que es una deoxiribosa por que en la posición 3 es oxidrilo, es trifosfato y tenemos la base nitrogenada N (que puede ser citosina, timina, guanina o adenina), esto se escribe como dNTP. La elongación de una cadena siempre se da de 5' a 3', y esto es por que en los 5' tengo los fosfatos y en 3' tengo el oxidrilo que se va a unir a otro nucleótido. Si no tenemos el oxidrilo y tenemos un hidrógeno entonces estamos en presencia del dideoxi nucleótido donde la polimerasa no puede adicionar otra base nitrogenada, por que para agregar otra base nitrogenada necesariamente debe haber un oxidrilo, entonces no voy a poder seguir elongando con los nucleótidos.

_ Con un ejemplo, lo que hacemos son cuatro mezclas, en un tubo ponemos los 4 nucleótidos y el ddATP pero en menor proporción, en el otro tubo pongo los 4 nucleótidos y el ddTTP y así con el ddCTP y el ddGTP. Entonces la polimerasa va a empezar a copiar la hebra pero a veces en vez de tomar un deoxi nucleótido va a tomar un dideoxi nucleótido y cuando toma este último se corta y no puede seguir copiando.

_ Al final de todo estos ciclos tenemos una muestra con pedazos de secuencia, y para poder identificar necesitamos un método de separación que sería la electroforesis en una matriz para cada uno de los tubos, y con los resultados podemos reconstruir la secuencia. Esto se hace con un equipo que se llama electroforesis capilar.

Repaso parcial: (guia 3)

Pregunta 1

Biología computacional y bioinformática: si bien estas se solapan, pero se dedican a abordar de diferentes maneras preguntas biológicas. La biología computacional trata de modelar un proceso biológico (biología sintética), ahora bien, para poder modelar o tratar de reproducir in silico un determinado metabolismo por ejemplo, necesariamente tengo que alimentar mis componentes con información o datos que vienen de experimentos biológicos y de información que va a estar guardada en las que nosotros generemos con por ejemplo los análisis bioinformáticos. Es incorrecto decir que la bioinformática no hace análisis porque esta interdisciplina que tiene la bioinformática justamente se nutre de diferentes disciplinas para abordar el análisis de experimentos a gran escala, y esos experimentos por un lado pueden ser para estudiar ácidos nucleicos, proteínas o metabolitos, donde dependiendo del tipo de experimento que se haga van a ser el tipo de dato que voy a tener. En bioinformática, bio es por la parte de biología, y la informática que es información pero esta gestionada a través de la computación, y a su vez se alimenta también de la matemática, y dependiendo si uno aborda la bioinformática como biólogo, matemático, diseñador de software, va a ser el tipo de análisis o desarrollo que se va a realizar.

_ La biología computacional se sirve también de las estadísticas y matemáticas pero no para gestionar la información solamente, sino para tratar de modelar algún proceso biológico. Esta necesariamente también va a utilizar métodos y algoritmos matemáticos pero la bioinformática también.

_ La bioinformática como disciplina podría existir sin la biología computacional pero no al revés, porque si no gestionamos los datos primero y no podemos analizar los experimentos a gran escala no podría alimentar a la biología computacional para poder resolver o modelizar un proceso biológico. Es decir, la bioinformática puede existir por sí sola, pero la biología computacional como necesita de los datos de experimentos previos o de análisis bioinformáticos.

Pregunta 2

_ Existen metodologías de separación (casi todas se basan en electroforesis), y de identificación (PCR, sonda complementaria para los ácidos nucleicos, western blot = anticuerpos para proteínas, hibridación, y para los metabolitos la espectrometría de masas que es una electroforesis) que en el caso del ADN y ARNm se basan en las bases nitrogenadas y su capacidad de interaccionar la adenina con la timina y la guanina con la citosina, y los puentes de hidrógeno.

_ La hebra de ADN es una doble hélice que se mantiene de esa manera porque existen las interacciones entre las purinas y pirimidinas, entonces si tenemos una hebra simple

podemos identificarla si ponemos otra hebra simple que sea complementaria a esa. A partir de esto se desarrollan todas las demás técnicas.

_ En las hebras, la PCR consiste en 3 etapas, la primera es aumentar la temperatura a 95° haciendo que quede una cadena lineal simple, el segundo paso es bajar la temperatura, porque con temperatura damos energía para romper los enlaces de hidrógeno y por eso las hebras se separan y quedan en hebras simples, hasta una temperatura que permita que los cebadores (entre 19 y 21 nucleótidos) se reconozcan la hebra simple y se formen una hebra doble nueva, y el tercer paso es aumentar la temperatura a 72° que es la temperatura donde hay una enzima que se llama polimerasa que incorpora los nucleótidos de las muestras y va elongando nucleótidos que sean complementarios a la hebra de ADN molde.

_ Con un ejemplo, lo que hacemos son cuatro mezclas, en un tubo ponemos los 4 nucleótidos y el ddATP pero en menor proporción, en el otro tubo pongo los 4 nucleótidos y el ddTTP y así con el ddCTP y el ddGTP. Entonces la polimerasa va a empezar a copiar la hebra pero a veces en vez de tomar un deoxi nucleótido va a tomar un dideoxi nucleótido y cuando toma este último se corta y no puede seguir copiando.