# Antecedentes

## La promiscuidad puede abordarse a distintos niveles incluyendo enzima, familia y ruta de metabólica.

Si se entiende a la promiscuidad como funciones alternativas de alguna unidad molecular, puede observarse promiscuidad a distintos niveles: desde un mismo gen que presenta splicing alternativo, una misma enzima con funciones alternativas, o bien una familia enzimática donde al menos algunos homologos codifican enzimas promiscuas (Nobeli, Favia, and Thornton 2009). Existen también rutas que generan productos metabólicos alternativos(Lamble et al. 2003,Weng and Noel (2012)), si se considera como la unidad de estudio a una ruta biosintética podemos generalizar la noción de promiscuidad al concepto de rutas promiscuas. La *Figura 1* muestra tres niveles en los que se puede estudiar la promiscuidad: i) Distinguiendo familias de enzimas promiscuas de familias especialistas, ii) Distinguiendo enzimas específicas de enzimas especialistas en una familia enzimática promiscua y finalmente iii) encontrando promiscuidad en rutas de metabolismo especializado.

PriA en Actinobacteria y HisA en enterobacteria son familias que ambas isomerizan proFAR en la ruta de síntesis de histidina pero solo la familia PriA es promiscua pues puede además isomerizar el sustrato PRA durante la síntesis de triptofano (Barona Gómez and Hodgson 2003). Sin embargo dentro de Actinobacteria, existen miembros no promiscuos de PriA, en algunas especies de Actinomyces los homólogos de *priA* codifican para enzimas monofuncionales en alguno de los dos sustratos, al menos en análisis in vitro (A. L. Juárez-Vázquez et al. 2018). Así pues dentro de una familia promiscua no todos los miembros tienen esta propiedad.

En este trabajo se buscará encontrar marcas de promiscuidad a nivel familia, enzima y ruta biosintética en el metabolismo especializado.

## Antecedentes conceptuales

Se ha intentado identificar enzimas promiscuas mediante aprendizaje máquina utilizando únicamente la secuencia de aminoácidos. Estos enfoques no han distiguido entre identificación de familias promiscuas e identificación a nivel de enzima. (Carbonell and Faulon 2010) Hasta ahora al utilizar únicamente la información de la secuencia no ha sido posible identificar una familia promiscua sin conocer previamente al menos un miembro promiscuo de ella. Por otra parte, diferenciar la promiscuidad a nivel de enzima se dificulta cuando la identidad de secuencia es alta, como en el caso de las PriA que han perdido la promiscuidad en Actinomyces (A. L. Juárez-Vázquez et al. 2018)

Para mejorar nuestro entendimiento del fenomeno además de la comparación de secuencias es necesario integrar otros elementos al análisis, Figura 2. Es difícil medir la promiscuidad en términos absolutos, por ejemplo, no se puede aseverar que una enzima es no promiscua sin haber previamente descartado todos los posibles sustratos del universo químico. Además incluso enzimas que resultan no promiscuas en análisis in vivo si son promiscuas al examinarlas in vivo (Noda-Garcia 2012). Sin embargo debido a que al adquirir una nueva función existe un umbral donde la función ancestral es conservada, es plausible estudiar transformar el problema de encontrar promiscuidad al de encontrar cambios en promiscuidad al relacionar estos últimos con las huellas que dejan los cambios funcionales(Soskine and Tawfik 2010,Aharoni et al. (2005),Bloom et al. (2007)). Entre los elementos relevantes que correlacionan con la adquisición de una función alternativa se encuentran además de divergencia de secuencia, diversidad en vecindad genomica(Zhao et al. 2014), la perdida o ganancia de genes(A. L. Juárez-Vázquez et al. 2018), las expansiones genicas, i.e. el crecimiento del pangenoma endtro de un grupo taxonómico (Martínez-Núñez, Rodríguez-Vázquez, and Pérez-Rueda 2015) y finalmente cambios estructurales o de flexibilidad durante la dinámica molecular(Zou et al. 2015, Gatti-Lafranconi and Hollfelder (2013)). Estos elementos tienen en común que reflejan un cambio en alguna propiedad genómica o biofísica observable en el registro evolutivo, de lo que se deriva que el buscar cambios en la promiscuidad de una enzima, familia o ruta, resulta mas factible por ahora que la búsqueda intrínseca de promiscuidad.

Debido a la abundancia de datos genómicos los primeros capítulos de este trabajo se centran en encontrar variaciones en secuencia, distribución del pangenoma, y vecindades genómicas para encontrar candidatos de familias, enzimas y rutas promiscuas.

## El establecimiento de un marco de conservación permite distinguir cambios

La función de una enzima es un concepto jerárquico, dependiente de la filogenia de un organismo (Szklarczyk et al. 2015). Por ello, para poder encontrar marcas de cambio funcional, por ejemplo diferencias en número de copias de una familia, primero es importante trabajar en la construcción de un marco filogenético consistente que permita ordenar inclusive organismos de la misma especie. La dificultad de esta tarea consiste en que si los organismos que se desea ordenar son muy cercanos, marcadores clásicos como el 16s son también muy parecidos en secuencia y no permiten resolver las relaciones entre ellos. Este caso dificultó la construcción de un árbol filogenético de *Actinomyces* y con ello se imposibilitaba encontrar patrones en la matriz de presencia / ausencia de genes (A. L. Juárez-Vázquez et al. 2018).

Para solventar la falta de resolución de genes individuales en organismos cercanos puede utilizarse el conjunto de todos los genes comunes en un linaje. Este conjunto es conocido como el core genome. Se han desarrollado herramientas bioinformáticas para este problema, por ejemplo phyloPhlan fija 400 genes comunes en bacteria y trata de localizarlos dado un conjunto de genomas sin importar su linaje (Segata et al. 2013). Sin embargo el contenido de genomas procariontes suele ser muy variable debido a mecanismos como transferencia horizontal y duplicación génica (Land et al. 2015, Koonin (2015)). Esta variabilidad puede dificultar encontrar muchos de estos 400 genes o bien puede ser que en cierto linaje no sean tan informativos. Entre organismos de la misma especie pueden suceder fenómenos como que el core siempre se reduzca al aumentar un nuevo genoma y que el conjunto total de familias geńicas (pangenoma) siempre aumente. Aunado a esta observación biológica están también las limitaciones técnicas, hay genes que no aparecen en un genóma porque este fue mal secuenciado o ensamblado y por lo tanto estos genes disminuyen el tamaño del core.

Asi pues para distinguir cambios genómicos ayuda establecer primero un orden entre los genomas del linaje a analizar. Para ello, un camino es localizar los genes del core exclusivos de cada grupo de genomas y libres de parálogos. Aunque al momento existe publicada *metaphor*, una herramienta de selección de ortólogos (Veen et al. 2014), al comenzar este trabajo no existía un método disponible para ello. Finalmente una vez obtenidos los genes del core es deseable contar con un algoritmo que además los concatene y entregue un árbol filogenético.

## La genómica comparativa como herramienta en la distinción de familias y enzimas promiscuas que participan en el metabolismo especializado.

Una parte del metabolismo especializado está compuesta por familias enzimáticas que evolucionaron de rutas de metabolismo central (G. Caetano-Anollés et al. 2009). En las familias expandidas, ya sea por duplicación o por transferencia horizontal, las expansiones pueden retener la función química de las rutas centrales (Schniete et al. 2018,Verdel-Aranda et al. (2015)), así como también la función alternativa suele estár presente presente aún a bajos niveles antes de la divergencia o duplicación (Soskine and Tawfik 2010, Aharoni et al. (2005),(**???**)). Por tanto las familias con expansiones en un linaje son candidatas a ser familias promiscuas en él. Se ha notado que en el linaje en que una familia enzimática es promiscua hay una zona de cambio en promiscuidad (Noda-García et al. 2015), Las expansiones de rutas centrales que participan en la síntesis de productos naturales son candidatos a presentar cambios en promiscuidad tanto a nivel familia como a nivel enzima.

por ello la observación de la retención de función ancestral al aparecer una función alternativa proporciona una zona favorable para la búsqueda de promiscuidad a nivel de enzima. En la familia existirá un gradiente de promiscuidad, las más cercanas a la zona de duplicación o divergencia tienen más posibilidades de tener un cambio en promiscuidad que las más conservadas y cercanas al metabolismo central.

Después de que en la sección anterior se estableció la posibilidad de mejorar las relaciones filogenéticas de un linaje, se abre la posibilidad de buscar en él expansiones de familias génicas de metabolismo central. Estas copias extra son candidatas a pertenecer a rutas de metabolismo especializado. Como prueba de concepto esta idea de minar genomas incorporando información evolutiva permitió la identificacion de la biosintesis de arsenolipidos (Cruz-Morales et al. 2016). La busqueda de productos naturales cuenta entre sus premisas que estos se producen en vecindades genomicas llamadas clusters y que ademas clusters cercanos (ya sea en contenido genico o en la secuencia de sus componentes), exploran variaciones metabolicas, es decir sus enzimas catalizan reacciones sobre sustratos parecidos aunque no identicos (Cruz-Morales et al. 2016,Medema and Fischbach (2015)).

La primera versión de evomining cuenta con 200 genomas de Actinobacteria, una base de datos de secuencias de enzimas de productos naturales y otra base de datos de secuencias de enzimas de rutas centrales curada a mano.

1 ARTS (Alanjary et al. 2017) 2 EvoMining (Cruz-Morales et al. 2016) Se ha avanzado en archaea (Martínez-Núñez et al. 2017)

Desarrollarla en combinacion con algoritmos de busqueda de cambios en la vecindad genomica la haran una plataforma ideal para abordar el problema de las familias, proporcionando una solucion a la dificultad de no tener conocimiento previo de un miembro promiscuo en la familia investigada.

Respecto al problema de los miembros, se propone explorar variaciones en vecindad genomica, flujo genico y dinamica molecular, como candidatos a reflejar la variacion en promiscuidad. Finalmente,

tomando como modelo biologico el phylum Actinobacteria, un grupo de bacterias reconocido por su diversidad metabolica donde se ha probado la existencia de promiscuidad enzimatica.

Evomining es una plataforma bioinformatica pensada para la identificacion de productos naturales  
Si se combinara evomining con la premisa de que vecindades distintas son marcadoras de funciones quimicas distintas, al encontrar una familia expandida con vecindades genomicas diferentes se podria solventar la deficiencia de otros metodos bioinformaticos consistente en que para identificar familias promiscuas se debe conocer previamente un miembro promiscuo de la misma. (Fig 4) Asi pues al combinar evomining con herramientas de vecindad genomica tanto de comparacion como de visualizacion estaremos mejorando su funcionalidad en la identificacion de familias promiscuas. En la siguiente sección BLA BKA

la variación es la materia prima de la evolucion.

## La genómica comparativa como herramienta en la priorización de clusters promiscuos

La promiscuidad nos interesa por su produccion de variantes. Si bien en rutas centrales rescata la función en metaboismo secundario crea nuevas variantes moleculares que permiten adaptación, de hecho pangenomas grandes correlacioan con aparición de nuevas funciones enzimáticas. Considero que el concepto de promiscuidad puede ser extendido a un nuevo niivel. Promiscuidad enzimatica, promiscuida de familia, promiscuidad de cluster. En rutas centrales robustez, pero en metabolismo secundario platicidad. Pangenoma abierto cerrado, Aqui proponemos organizar los clusters y finalmente la medida de su apertura. variantes de clusters producen nuevos compuestos, ya sea por promiscuidad enzimática en un core conservado o por variación en la presencia / ausencia de genes. Un cluster recibe los mismos sustratos y los transforma en diferentes productos.

### Expansion y contextos genomicos como herramienta de anotacion funcional

Al evaluar la herramienta de análisis de promiscuidad PROMISE (Carbonell and Faulon 2010) en un set de datos de la familia HisA/PriA (Noda-García et al. 2015) obtuve que en su mejor desempeño es (huella molecular de tamaño 6) clasifica correctamente casi todas las no promiscuas, (HisA) pero no sucede lo mismo con la familia PriA donde tiene exito en 16 de 45 casos. Al aplicar el mismo tamaño de huella a 9 miembros promiscuos de la familia IlvC no consigue predecir correctamente ninguno de ellos reflejando tal vez que en su conjunto de entrenamiento no había miembros promiscuos ilvC. Por lo menos para estas familias el conjunto de entrenamiento o los descriptores no son suficientes para la anotacion de promiscuidad.

La diversidad enzimática existente es el resultado de un proceso de expansion, mutacion y seleccion que se ha desarrollado durante el transcurso de la historia evolutiva (Khersonsky and Tawfik 2010, Pearson (2012)). Existe evidencia de que cierto grado de promiscuidad o divergencia funcional precede a la duplicacion genica (Hughes 1994). Por este motivo detectar expansiones ya sea duplicaciones o transferencias horizontales (Treangen and Rocha 2011), puede ser un buen punto de partida para determinar divergencia funcional y promiscuidad. No todas las expansiones denotan cambio de funcion enzimatica, algunas pueden ser meros accidentes, sin embargo dado que la funcion de una enzima suele estar relacionada con sus vecinos (R. Overbeek et al. 1999,(**???**),Zhao et al. (2013), Verdel-Aranda et al. (2015)), una expansion en una vecindad genomica diferente de la tradicional sera un referente de adquisicion de una nueva funcion y entonces un indicador de existencia previa de promiscuidad.

Para sistematizar el estudio de contextos y vecindades genomicas se desarrollo Search Tool for the Retrieval of Interacting Genes/Proteins STRING (Snel et al. 2000), que cuenta con una anotacion de ortologia jerarquica y consistente, realizada en 2000 organismos en cuyo marco interacciones de proteinas con implicaciones funcionales son predichas tanto de novo por informacion genomica de co-ocurrencia como por mineria de datos en articulos publicados. STRING es una base de datos, y como tal no permite agregar nuevos genomas para su analisis. Sus 2000 organismos incluyen especies tanto bacterianas como eucariotas. Al existir tanta diversidad, los genomas disponibles para un genero o clase especificos son escasos, p. g. de los mas de 300 genomas disponibles de Streptomyces solo 24 estan incluidos.

Para resolver la baja cobertura de STRING hacia ciertos grupos taxonomicos se pueden desarrollar scripts de vecindad genomica utilizando RAST (Rapid Annotation using Subsystem Technology); un servicio interactivo de anotacion automatica de genomas de bacterias y arqueas (Aziz et al. 2008,R. Overbeek et al. (2014)) donde la funcion de cada gen se asigna de acuerdo a conocimiento previo de subsecuencias de organismos cercanos filogeneticamente, cuando es posible se incluye en un subsistema metabolico. Estamos en una era de explosion de datos genomicos, proximamente se espera contar con millones de genomas bacterianos incluso provenientes de bacterias no cultivables, por ello los algoritmos deben ser constantemente optimizados a los nuevos volumenes de datos (Medema and Fischbach 2015). Ante esta expectativa seria muy util desarrollar algoritmos de analisis genomico que sean de codigo libre o al menos interactivos para que cada laboratorio pueda personalizarlos para sus propios genomas.

Finalmente, no solo la vecindad genomica inmediata puede ser utilizada como distintivo en la busqueda de promiscuidad, diferencias en el contexto genomico en genes relacionados con una enzima promiscua, sin importar su ubicacion dentro del genoma tambien pueden ser relevantes para la perdida o ganancia de funcion quimica (Noda-García et al. 2013), (Juarez Vazquez et al 2015).

### Contexto y vecindades genomicas

En 2012 fueron analizados 102 genomas de 29 familias de Actinobacteria (Noda-Garcia 2012). sugiriendo que al menos en *Corynebacteria* el contexto y la vecindad genomica incidian en la sub-funcionalizacion de PriA en subHisA (Noda-García et al. 2013). Respecto a IlvC, otra familia involucrada en la sintesis de aminoacidos fue estudiada y caracterizada bioquimicamente en 1 Corynebacterium y 8 Streptomyces (Verdel-Aranda et al. 2015). Para ampliar estos resultados, utilizando la anotacion de RAST y una generalizacion de la definicion de vecindad de STRING, se diseño un algoritmo para identificar vecindades similares asi como uno de visualizacion de contexto, ambos disponibles como software libre en github [nselem/perlas](https://github.com/nselem/perlas) .

El algoritmo de clasificacion de vecindades permite agruparlas en clusters y calificar estos clusters segun su conservacion dado un grupo de bacterias. La definicion de vecindad y similitud de vecindad esta descrita posteriormente en los metodos. El algoritmo fue aplicado a la familia IlvC en 290 Streptomyces resultando 9 clusters [Datos](http://148.247.230.43/nselem/CONTEXTS/REL_St275/ilvC/Contextos.php) entre los mas poblados el primero cuenta con 279 elementos, otro con 9 elemento y dos mas con 7 miembros (Fig 3), resultados experimentals son congruentes con que existe divergencia funcional entre miembros de clusters distintos (Verdel-Aranda et al. 2015)

Natural products genomic era(Harvey, Edrada-Ebel, and Quinn 2015)

## Estudio de una familia PriA

### Caracterizacion in vivo

No todas las familias promiscuas provienen de expansiones, tal es el caso de PriA en Actinobacteria, donde no tiene expansiones y hasta el momento no se le conoce participación en rutas de metabolismo especializado.

cuya promiscuidad es debida a la pérdida de TrpF Algunas enzimas PriA no han mostrado promiscuidad in vitro pero si in vivo ya que sobreviven en un medio sin triptofano, es decir in vivo complementan la funcion trpF. Para la construccion de cepas de Streptomyces con variantes no nativas de priA minimizando la modificacion genomica y el efecto de sobreexpresion, se planea utilizar E. coli como intermediario para realizar seleccion por auxotrofia. Se cuenta con un conjunto de plasmidos para transformar a E. coli asi como con las mutantes sencillas de E. coli para trpF y hisA que permiten realizar seleccion por auxotrofias. Ademas tenemos una coleccion de cepas nativas de Streptomyces asi como un mutante de PriA de S. coelicolor. Se optimizo una reaccion de PCR para la amplificacion de un segmento de DNA de S. coelicolor que contiene a priA.

### Caracterizacion bioquimica in vitro.

De la familia PriA y sus subfamilias se han caracterizado bioquimicamente miembros selectos de Actinomycetaceae, Bifidobacteriaceae, Micrococcaceae, Acidimicrobiaceae, Corynebacterium, Mycobacteriaceae, Streptomycetaceae, Camera (provenientes de metagenoma), reconstrucciones ancestrales, 80 mutantes de Corynebacterium, y 2 mutantes de Camera mediante cineticas enzimaticas para calcular las constantes Kcat,Km. El genero Streptomyces, el que cuenta con mayor cantidad de genomas disponibles representa una oportunidad muy poco explotada de explorar la influencia del contexto y la vecindad genomicas en secuencias de PriA (Tabla 3, Figura 5).

### Modelado de dinamica molecular

La dinamica es un metodo que permite hacer simulaciones de particulas que sirve para obtener informacion de propiedades macroscopicas de un conjunto de atomos (Petrenko and Meller 2001,(*Molecular Modeling of Proteins Andreas Kukol Springer* 2017)). Es util en el marco de mi proyecto porque permite la exploracion del espacio conformacional, y se ha visto que este esta relacionado con la actividad de la enzima (Sikosek and Chan 2014), ademas dado un conformero permite verificar su estabilidad. Resuelve la ecuacion de movimiento de Newton con base a una configuracion inicial, las fuerzas interatomicas como los enlaces covalentes, las fuerzas de Van der Waals y la carga de las particulas(Campbell 2012). Entonces para generar una simulacion de dinamica molecular, debe contarse con una estructura como punto de partida, ya sea esta cristalografica o modelada de novo o por homologia. El laboratorio de bioinformatica y biofisica computacional ha desarrollado un protocolo de generacion de modelos homologos estructurales y dinamicas moleculares (Carrillo-Tripp et al 2015 in prep); con este pipeline se han generado dos estructuras de Camera (Noda-García et al. 2015), 30 estructuras y dinamicas de miembros de Actinobacteriaceae y Bifidobacteriaceae (Vazquez-Juarez et al in prep.) y finalmente una estructura de subHisA de Corynebacterium diphteriae. En la familia Streptomyces, interesante debido a su variacion en contexto genomico y en mediciones in vitro aun no se modelan dinamicas moleculares aunque 40 estructuras por homologia estan en proceso.

En un estudio de subHisA (Noda-Garcia 2012) se utilizo el método de dinámica molecular y se comparó el número de confórmeros entre miembros de subHisA y PriA, resultando mayor el de PriA como corresponde a una enzima promiscua. El estudio sobre la relación dinámica-flexibilidad de $\noindent\beta$-lactamasas utiliza replica exchange, una variacion de dinamica molecular que corre replicas en paralelo a distintas temperaturas (Zhou 2006). Una desventaja de este metodo es que por el costo computacional de las replicas agregar explicitamente otras moleculas a la simulacion como el solvente no es posible en tiempo razonable. Una vez generadas las dinamicas moleculares se procedera a calcular tanto el numero de conformeros como el indice de flexibilidad dsi (Zou et al. 2015). Se esta desarrollando PEDB, promiscuous enzyme database, una base datos genomicos, evolutivos, bioquimicos y estructurales y de metabolismo de PriA en Actinobacteria donde se procedera al analisis de los mismos (<http://148.247.230.43/nselem/PHP/queries.html>).

En conclusion la promiscuidad enzimatica es un fenomeno complejo debido a multiples causas. Existe una gran variedad de estudios con enfoques puntuales sobre aspectos estructurales, dinamicos y evolutivos sin embargo hasta ahora no se han reportado trabajos multidisciplinarios que involucren a todas las partes involucradas

Aharoni, Amir, Leonid Gaidukov, Olga Khersonsky, Stephen McQ Gould, Cintia Roodveldt, and Dan S. Tawfik. 2005. “The ’Evolvability’ of Promiscuous Protein Functions.” *Nature Genetics* 37 (1): 73–76. doi:[10.1038/ng1482](https://doi.org/10.1038/ng1482).

Alanjary, Mohammad, Brent Kronmiller, Martina Adamek, Kai Blin, Tilmann Weber, Daniel Huson, Benjamin Philmus, and Nadine Ziemert. 2017. “The Antibiotic Resistant Target Seeker (ARTS), an Exploration Engine for Antibiotic Cluster Prioritization and Novel Drug Target Discovery.” *Nucleic Acids Research* 45 (W1): W42–W48. doi:[10.1093/nar/gkx360](https://doi.org/10.1093/nar/gkx360).

Aziz, Ramy K, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, et al. 2008. “The RAST Server: Rapid Annotations Using Subsystems Technology.” *BMC Genomics* 9 (February): 75. doi:[10.1186/1471-2164-9-75](https://doi.org/10.1186/1471-2164-9-75).

Barona Gómez, Francisco, and David A. Hodgson. 2003. “Occurrence of a Putative Ancient Like Isomerase Involved in Histidine and Tryptophan Biosynthesis.” *EMBO Reports* 4 (3): 296–300. doi:[10.1038/sj.embor.embor771](https://doi.org/10.1038/sj.embor.embor771).

Bloom, Jesse D., Philip A. Romero, Zhongyi Lu, and Frances H. Arnold. 2007. “Neutral Genetic Drift Can Alter Promiscuous Protein Functions, Potentially Aiding Functional Evolution.” *Biology Direct* 2: 17. doi:[10.1186/1745-6150-2-17](https://doi.org/10.1186/1745-6150-2-17).

Caetano-Anollés, Gustavo, Liudmila S. Yafremava, Hannah Gee, Derek Caetano-Anollés, Hee Shin Kim, and Jay E. Mittenthal. 2009. “The Origin and Evolution of Modern Metabolism.” *The International Journal of Biochemistry & Cell Biology*, Molecular and Cellular Evolution: A Celebration of the 200th Anniversary of the Birth of Charles Darwin, 41 (2): 285–97. doi:[10.1016/j.biocel.2008.08.022](https://doi.org/10.1016/j.biocel.2008.08.022).

Campbell, Ian. 2012. *Biophysical Techniques - Paperback - Iain D. Campbell - Oxford University Press*.

Carbonell, Pablo, and Jean-Loup Faulon. 2010. “Molecular Signatures-Based Prediction of Enzyme Promiscuity.” *Bioinformatics* 26 (16): 2012–9. doi:[10.1093/bioinformatics/btq317](https://doi.org/10.1093/bioinformatics/btq317).

Cruz-Morales, Pablo, Johannes Florian Kopp, Christian Martínez-Guerrero, Luis Alfonso Yáñez-Guerra, Nelly Selem-Mojica, Hilda Ramos-Aboites, Jörg Feldmann, and Francisco Barona-Gómez. 2016. “Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes.” *Genome Biology and Evolution* 8 (6): 1906–16. doi:[10.1093/gbe/evw125](https://doi.org/10.1093/gbe/evw125).

Gatti-Lafranconi, Pietro, and Florian Hollfelder. 2013. “Flexibility and Reactivity in Promiscuous Enzymes.” *ChemBioChem* 14 (3): 285–92. doi:[10.1002/cbic.201200628](https://doi.org/10.1002/cbic.201200628).

Harvey, Alan L., RuAngelie Edrada-Ebel, and Ronald J. Quinn. 2015. “The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era.” *Nature Reviews Drug Discovery* 14 (2): 111–29. doi:[10.1038/nrd4510](https://doi.org/10.1038/nrd4510).

Hughes, Austin L. 1994. “The Evolution of Functionally Novel Proteins After Gene Duplication.” *Proceedings of the Royal Society of London B: Biological Sciences* 256 (1346): 119–24. doi:[10.1098/rspb.1994.0058](https://doi.org/10.1098/rspb.1994.0058).

Juárez-Vázquez, Ana Lilia, Janaka N Edirisinghe, Ernesto A Verduzco-Castro, Karolina Michalska, Chenggang Wu, Lianet Noda-García, Gyorgy Babnigg, et al. 2018. “Evolution of Substrate Specificity in a Retained Enzyme Driven by Gene Loss.” *eLife* 6. Accessed January 16. doi:[10.7554/eLife.22679](https://doi.org/10.7554/eLife.22679).

Khersonsky, Olga, and Dan S. Tawfik. 2010. “Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective.” *Annual Review of Biochemistry* 79: 471–505. doi:[10.1146/annurev-biochem-030409-143718](https://doi.org/10.1146/annurev-biochem-030409-143718).

Koonin, Eugene V. 2015. “The Turbulent Network Dynamics of Microbial Evolution and the Statistical Tree of Life.” *Journal of Molecular Evolution* 80 (5-6): 244–50. doi:[10.1007/s00239-015-9679-7](https://doi.org/10.1007/s00239-015-9679-7).

Lamble, Henry J., Narinder I. Heyer, Steven D. Bull, David W. Hough, and Michael J. Danson. 2003. “Metabolic Pathway Promiscuity in the Archaeon Sulfolobus Solfataricus Revealed by Studies on Glucose Dehydrogenase and 2-Keto-3-Deoxygluconate Aldolase.” *Journal of Biological Chemistry* 278 (36): 34066–72. doi:[10.1074/jbc.M305818200](https://doi.org/10.1074/jbc.M305818200).

Land, Miriam, Loren Hauser, Se-Ran Jun, Intawat Nookaew, Michael R. Leuze, Tae-Hyuk Ahn, Tatiana Karpinets, et al. 2015. “Insights from 20 Years of Bacterial Genome Sequencing.” *Functional & Integrative Genomics* 15 (2): 141–61. doi:[10.1007/s10142-015-0433-4](https://doi.org/10.1007/s10142-015-0433-4).

Martínez-Núñez, Mario Alberto, Zuemy Rodríguez-Escamilla, Katya Rodríguez-Vázquez, and Ernesto Pérez-Rueda. 2017. “Tracing the Repertoire of Promiscuous Enzymes Along the Metabolic Pathways in Archaeal Organisms.” *Life* 7 (3). doi:[10.3390/life7030030](https://doi.org/10.3390/life7030030).

Martínez-Núñez, Mario Alberto, Katya Rodríguez-Vázquez, and Ernesto Pérez-Rueda. 2015. “The Lifestyle of Prokaryotic Organisms Influences the Repertoire of Promiscuous Enzymes.” *Proteins: Structure, Function, and Bioinformatics* 83 (9): 1625–31. doi:[10.1002/prot.24847](https://doi.org/10.1002/prot.24847).

Medema, Marnix H., and Michael A. Fischbach. 2015. “Computational Approaches to Natural Product Discovery.” *Nature Chemical Biology* 11 (9): 639–48. doi:[10.1038/nchembio.1884](https://doi.org/10.1038/nchembio.1884).

*Molecular Modeling of Proteins Andreas Kukol Springer*. 2017. Accessed February 8. <http://www.springer.com/us/book/9781588298645>.

Nobeli, Irene, Angelo D. Favia, and Janet M. Thornton. 2009. “Protein Promiscuity and Its Implications for Biotechnology.” *Nature Biotechnology* 27 (2): 157–67. doi:[10.1038/nbt1519](https://doi.org/10.1038/nbt1519).

Noda-Garcia, Lianet. 2012. “Estudio de La Evolución Molecular de La Función Enzimática Susando Como Modelo Una Enzima Con Características Ancestrales.” PhD thesis, Irapuato, GTO: Langebio, CINVESTAV.

Noda-García, Lianet, Aldo R. Camacho-Zarco, Sofía Medina-Ruíz, Paul Gaytán, Mauricio Carrillo-Tripp, Vilmos Fülöp, and Francisco Barona-Gómez. 2013. “Evolution of Substrate Specificity in a Recipient’s Enzyme Following Horizontal Gene Transfer.” *Molecular Biology and Evolution* 30 (9): 2024–34. doi:[10.1093/molbev/mst115](https://doi.org/10.1093/molbev/mst115).

Noda-García, Lianet, Ana L. Juárez-Vázquez, María C. Ávila-Arcos, Ernesto A. Verduzco-Castro, Gabriela Montero-Morán, Paul Gaytán, Mauricio Carrillo-Tripp, and Francisco Barona-Gómez. 2015. “Insights into the Evolution of Enzyme Substrate Promiscuity After the Discovery of Isomerase Evolutionary Intermediates from a Diverse Metagenome.” *BMC Evolutionary Biology* 15 (June). doi:[10.1186/s12862-015-0378-1](https://doi.org/10.1186/s12862-015-0378-1).

Overbeek, Ross, Michael Fonstein, Mark D’Souza, Gordon D. Pusch, and Natalia Maltsev. 1999. “The Use of Gene Clusters to Infer Functional Coupling.” *Proceedings of the National Academy of Sciences* 96 (6): 2896–2901. doi:[10.1073/pnas.96.6.2896](https://doi.org/10.1073/pnas.96.6.2896).

Overbeek, Ross, Robert Olson, Gordon D. Pusch, Gary J. Olsen, James J. Davis, Terry Disz, Robert A. Edwards, et al. 2014. “The SEED and the Rapid Annotation of Microbial Genomes Using Subsystems Technology (RAST).” *Nucleic Acids Research* 42 (Database issue): D206–D214. doi:[10.1093/nar/gkt1226](https://doi.org/10.1093/nar/gkt1226).

Pearson, Helen. 2012. “Prehistoric Proteins: Raising the Dead.” *Nature News* 483 (7390): 390. doi:[10.1038/483390a](https://doi.org/10.1038/483390a).

Petrenko, Roman, and Jarosław Meller. 2001. “Molecular Dynamics.” In *eLS*. John Wiley & Sons, Ltd. <http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0003048.pub2/abstract>.

Schniete, Jana K., Pablo Cruz-Morales, Nelly Selem-Mojica, Lorena T. Fernández-Martínez, Iain S. Hunter, Francisco Barona-Gómez, and Paul A. Hoskisson. 2018. “Expanding Primary Metabolism Helps Generate the Metabolic Robustness to Facilitate Antibiotic Biosynthesis in Streptomyces.” *mBio* 9 (1): e02283–17. doi:[10.1128/mBio.02283-17](https://doi.org/10.1128/mBio.02283-17).

Segata, Nicola, Daniela Börnigen, Xochitl C. Morgan, and Curtis Huttenhower. 2013. “PhyloPhlAn Is a New Method for Improved Phylogenetic and Taxonomic Placement of Microbes.” *Nature Communications* 4 (August): 2304. doi:[10.1038/ncomms3304](https://doi.org/10.1038/ncomms3304).

Sikosek, Tobias, and Hue Sun Chan. 2014. “Biophysics of Protein Evolution and Evolutionary Protein Biophysics.” *Journal of the Royal Society Interface* 11 (100): 20140419. doi:[10.1098/rsif.2014.0419](https://doi.org/10.1098/rsif.2014.0419).

Snel, B., G. Lehmann, P. Bork, and M. A. Huynen. 2000. “STRING: A Web-Server to Retrieve and Display the Repeatedly Occurring Neighbourhood of a Gene.” *Nucleic Acids Research* 28 (18): 3442–4. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC110752/>.

Soskine, Misha, and Dan S. Tawfik. 2010. “Mutational Effects and the Evolution of New Protein Functions.” *Nature Reviews Genetics* 11 (8): 572–82. doi:[10.1038/nrg2808](https://doi.org/10.1038/nrg2808).

Szklarczyk, Damian, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, et al. 2015. “STRING V10: Protein–protein Interaction Networks, Integrated over the Tree of Life.” *Nucleic Acids Research* 43 (D1): D447–D452. doi:[10.1093/nar/gku1003](https://doi.org/10.1093/nar/gku1003).

Treangen, Todd J., and Eduardo P. C. Rocha. 2011. “Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes.” *PLOS Genetics* 7 (1): e1001284. doi:[10.1371/journal.pgen.1001284](https://doi.org/10.1371/journal.pgen.1001284).

Veen, Bernd E. van der, Hugh M. Harris, Paul W. O´Toole, and Marcus J. Claesson. 2014. “Metaphor: Finding Bi-Directional Best Hit Homology Relationships in (Meta)genomic Datasets.” *Genomics* 104 (6, Part B): 459–63. doi:[10.1016/j.ygeno.2014.10.008](https://doi.org/10.1016/j.ygeno.2014.10.008).

Verdel-Aranda, Karina, Susana T. López-Cortina, David A. Hodgson, and Francisco Barona-Gómez. 2015. “Molecular Annotation of Ketol-Acid Reductoisomerases from Streptomyces Reveals a Novel Amino Acid Biosynthesis Interlock Mediated by Enzyme Promiscuity.” *Microbial Biotechnology* 8 (2): 239–52. doi:[10.1111/1751-7915.12175](https://doi.org/10.1111/1751-7915.12175).

Weng, J.-K., and J. P. Noel. 2012. “The Remarkable Pliability and Promiscuity of Specialized Metabolism.” *Cold Spring Harbor Symposia on Quantitative Biology* 77 (January): 309–20. doi:[10.1101/sqb.2012.77.014787](https://doi.org/10.1101/sqb.2012.77.014787).

Zhao, Suwen, Ritesh Kumar, Ayano Sakai, Matthew W. Vetting, B. McKay Wood, Shoshana Brown, Jeffery B. Bonanno, et al. 2013. “Discovery of New Enzymes and Metabolic Pathways by Using Structure and Genome Context.” *Nature* 502 (7473): 698–702. doi:[10.1038/nature12576](https://doi.org/10.1038/nature12576).

Zhao, Suwen, Ayano Sakai, Xinshuai Zhang, Matthew W. Vetting, Ritesh Kumar, Brandan Hillerich, Brian San Francisco, et al. 2014. “Prediction and Characterization of Enzymatic Activities Guided by Sequence Similarity and Genome Neighborhood Networks.” *eLife* 3 (June): e03275. doi:[10.7554/eLife.03275](https://doi.org/10.7554/eLife.03275).

Zhou, Ruhong. 2006. “Replica Exchange Molecular Dynamics Method for Protein Folding Simulation.” In *Protein Folding Protocols*, edited by Yawen Bai and Ruth Nussinov, 205–23. Methods in Molecular Biology™ 350. Humana Press. <http://dx.doi.org/10.1385/1-59745-189-4%3A205>.

Zou, Taisong, Valeria A. Risso, Jose A. Gavira, Jose M. Sanchez-Ruiz, and S. Banu Ozkan. 2015. “Evolution of Conformational Dynamics Determines the Conversion of a Promiscuous Generalist into a Specialist Enzyme.” *Molecular Biology and Evolution* 32 (1): 132–43. doi:[10.1093/molbev/msu281](https://doi.org/10.1093/molbev/msu281).