

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
DISSERTAÇÃO DE MESTRADO

ORIGEM DOS GENES DAS VIAS DE SECREÇÃO
DO SISTEMA DIGESTIVO HUMANO

FENÍCIA BRITO SANTOS

BELO HORIZONTE

2017

FENÍCIA BRITO SANTOS

ORIGEM DOS GENES DAS VIAS DE SECREÇÃO
DO SISTEMA DIGESTIVO HUMANO

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Mestre em Bioinformática

Orientador: Dr. José Miguel Ortega

Co-orientador: Dr. Tetsu Sakamoto

BELO HORIZONTE

2017

043

Santos, Fenícia Brito.

Origem dos genes das vias de secreção do sistema digestivo humano
[manuscrito] / Fenícia Brito Santos. - 2017.

67 f. : il. ; 29,5 cm.

Orientador: Dr. José Miguel Ortega. Co-orientador: Dr. Tetsu Sakamoto.
Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de
Ciências Biológicas.

1. Genômica. 2. Evolução. 3. Digestão. 4. Bioinformática - Teses. I. Ortega,
José Miguel. II. Sakamoto, Tetsu. III. Universidade Federal de Minas Gerais.
Instituto de Ciências Biológicas. IV. Título.

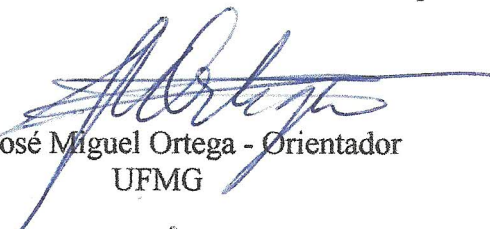
CDU: 573:004



"Origem Dos Genes Das Vias De Secreção Do Sistema Digestivo Humano"

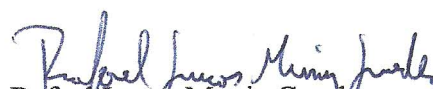
Fenícia Brito Santos

Dissertação aprovada pela banca examinadora constituída pelos Professores:


Prof. José Miguel Ortega - Orientador
UFMG


Prof. Tetsu Sakamoto - Co-Orientador
UFMG


Prof. Francisco Pereira Lobo
UFMG


Rafael Lucas Muniz Guedes
Hermes Pardini

Belo Horizonte, 31 de julho de 2017.

*The show must go on
Inside my heart is breaking
My make-up may be flaking
But my smile still stays on
- Queen*

AGRADECIMENTOS

Esse trabalho representa um recomeço para mim e sem o apoio incondicional da minha família não teria forças para realiza-lo. Agradeço imensamente aos meus Pais Agnaldo e Oderiana Santos pelos sacrifícios que fizeram para que eu realizasse o meu sonho e por me apoiarem e incentivarem a ser sempre melhor. À minha irmã pelo incentivo, apoio e conselhos que são tão essenciais na minha vida.

Aos meus amigos Núbia Prates, Thais Laurano, Stellamares Soares, Marcele Laux, Gustavo Santos, Vinícios Rodovalho, Alessandra Lima, Andersson Coqueiro e Rodrigo Profeta por tornarem meus dias em BH mais felizes e por sempre torcerem pelo meu sucesso. Muito obrigada.

À família Biodados: Lissur Orsine, Fernanda Stussi, Élisson Nogueira, Rayson Barbosa, Carlos Gonçalves, Beatriz Moura e Tetsu Sakamoto que me receberam tão bem desde o primeiro dia. Sou muito feliz por ter vocês como colegas de trabalho e amigos.

À Carlos Xavier, por sempre estar disponível para ajudar e resolver problemas, assim como pela generosa ajuda dada com o Arcobaleno. E à Lissur Orsine pela grande ajuda durante a realização desse trabalho, pelos ensinamentos e conselhos.

Ao meu co-orientador Tetsu Sakamoto, sempre muito gentil e paciente, pela ajuda com as análises com o ELDOgraph e por toda a atenção dada para esse projeto, sem você esse trabalho não seria tão rico.

Agradeço especialmente ao meu orientador J. Miguel Ortega, por ter aceitado me orientar de forma tão generosa e pelo cuidado que tem, não só comigo, mas com todos os seus alunos.

Agradeço ao programa de Pós-graduação em Bioinformática por terem me recebido tão bem e, sempre muito generosos, ajudarem nos momentos mais difíceis.

RESUMO

Os processos que garantem a manutenção da vida requerem energia e para todos os organismos heterótrofos a fonte dessa energia são os nutrientes disponíveis no ambiente ao seu redor. Desde a origem das primeiras formas de vida os processos e sistemas envolvidos na obtenção e metabolismo dos alimentos tem sido determinantes no sucesso evolutivo das espécies. O surgimento de organismos multicelulares com sistemas compartimentalizados permitiu aos metazoários acumular energia e realocá-la para o desenvolvimento de outras funções. Em humanos, as principais etapas da digestão dependem das secreções salivar, gástrica, pancreática e biliar. Apesar dos estudos de anatomia comparada em metazoários serem bem documentados e o conhecimento sistêmico de diversos processos seja reportado em muitas bases de dados, trabalhos que abordam a genômica comparativa e evolução dos componentes desse sistema são escassos. Assim, neste trabalho buscamos estudar a origem sequencial dos genes envolvidos na heterotrofia usando como modelo o sistema digestivo humano. Para isso, foram criados diagramas reconhecíveis por software para as vias de secreção do sistema digestivo com base em modelos disponíveis no KEGG Pathway. A origem de cada componente foi estimada através de ferramentas de agrupamento de homólogos e inferência do ancestral comum mais recente. Assim, foi possível retratar a origem do sistema com base na origem dos seus genes. Nossos resultados mostram que os genes mais ancestrais encontrados nas vias atuam nos processos de sinalização celular e surgiram antes do sistema ser originado. Já os componentes mais recentes, como os receptores e alguns transportadores, essenciais para a função de secreção, surgiram a partir do clado Metazoa, principalmente em Bilateria. Além disso, alguns componentes com função acessória, como a secreção de bicarbonato no pâncreas e na bile tem origem mais recente, indicando que esse processo aparece como um refinamento nas vias de secreção. A secreção salivar é a que possui o maior número de componentes recentes, sendo muitas das proteínas secretadas exclusivas de mamíferos. Além da datação dos genes, verificamos quais organismos ou OTU tinham maior número de proteínas próximas às de humano, utilizando o programa ELDOgraph. Os resultados mostram que a espécie que possui mais ELDO com humano são as do gênero Pan (chimpanzés e bonobos); já analisando em nível de ordem as espécies com mais ELDO com humano são das ordens Dermoptera e Rodentia, e dentre a superordem Laurasiateria, Perissodactyla e Carnivora. Em conjunto as análises mostram que as vias do sistema digestivo em mamíferos compartilham muitas similaridades, embora algumas proteínas sejam mais distantes. Os dados apresentados nesse trabalho permitiram delinear um panorama a respeito da evolução do processo digestivo, contribuindo assim para a história da evolução desse sistema.

Palavra-chave: heterotrofia, evolução, digestão, vias metabólicas.

ABSTRACT

The processes that ensure the maintenance of life require energy and for all heterotrophic organisms the source of energy are the nutrients available in the environment. Since the origin of the first forms of life the processes and systems involved in obtaining and metabolizing food have been determinant in the evolutionary success of the species. The emergence of multicellular organisms with compartmentalized systems allowed metazoans to accumulate energy and reallocate it for the development of other functions. In humans the main stages of digestion depends on salivary, gastric, pancreatic and biliary secretions. Although studies of comparative anatomy in metazoa are well documented and the systemic knowledge of several processes is reported in many databases, studies addressing the comparative genomics and the evolution of the components of this system are scarce. Here our goal was to study the sequential origin of the genes involved in heterotrophy using the human digestive system as model. For this, software-recognizable diagrams for the digestive system secretion pathways were created based on models available in KEGG Pathway database. The origin of each component was estimated using tools for homologous clustering and for lowest common ancestor inference. This allowed us to infer the origin of the system based on the origin of its genes. Our results show that the most ancestral genes found in the pathways act on cell signaling processes and arose before the systems have been originated. The most recent components, such as receptors and some transporters, which are essential for secretion function, appeared from Metazoa. In addition, some components with auxiliary function, such as bicarbonate secretion in the pancreas and bile, have a more recent origin, indicating that this process appears as a refinement in these secretory pathways. Salivary secretion has the highest number of recent components and many of the proteins secreted are exclusive in mammals. In addition, we performed an analysis using the ELDOgraph program to identify which organisms or OTUs have more proteins close to human. The results show that the species that has more ELDO with human belong to the genus Pan (chimpanzees and bonobos). Analyzing at the order level the species with more ELDO with human belong to the orders Dermoptera and Rodentia. Our results show that the secretion pathways of the digestive system in mammals share many similarities, although some proteins are more distant. The data presented here allowed us to draw a scenario about the evolution of the digestive process, contributing to the evolutionary history of this system.

Keywords: heterotrophy, evolution, digestion, metabolic pathways.

LISTA DE ABREVIATURAS

ARCoBALeno	Application for Rapidly COloring Biological pathways by Ancestrality or gene ontoLogy-Enriched iNformatiOn
BA	Ácido biliar
BBH	Bidirectional Best Hits
BLAST	Basic Local Alignment Search Tool
COG	Clusters of Orthologous Groups
eggNOG	evolutionary genealogy of genes: Non-supervised Orthologous Groups
ELDO	External Least Divergent Ortholog
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LBP	Last Biological Process
LCA	Lowest Common Ancestor (menor ancestral comum)
LDO	Least Divergent Ortholog
LMF	Last Molecular Function
NCBI	National Center for Biotechnology Information
OMA	Orthologous Matrix
OTU	Unidade Taxonômica Operacional
PANTHER	Protein ANalysis THrough Evolutionary Relationships
RefSeq	Reference Sequence
Uniprot	Universal Protein Resource
BOWS	Bioinformatics Open Web Services
GPML	Graphical Pathway Markup Language

LISTA DE FIGURAS

- Figura 1 – Anatomia das glândulas salivares. (A) Localização das glândulas; (B) representação geral da histologia da glândula salivar. Adaptado de Mescher AL. Junqueira's Basic Histology, 13e; 2013..... 3
- Figura 2 – Divisão fisiológica do estômago. Representação anatômica das regiões do estômago indicando os principais grupos de células presentes. As áreas denominadas Fundo e Corpo estão coloridas de vermelho, e o antro de laranja. Reproduzido de (SCHUBERT, 2014). 4
- Figura 3 – Anatomia das glândulas gástricas. As glândulas oxíntica e pilórica possuem diferentes grupos de células, como indicado na figura. Reproduzido de (SCHUBERT; PEURA, 2008). 5
- Figura 4 – Anatomia do pâncreas. (A) Representação anatômica do pâncreas; (B) Estrutura das glândulas exócrinas pancreáticas; (C) Representação do ácino pancreático, onde é formado o suco pancreático; (D) Anatomia de uma ilhota de Langerhans que secretam insulina e glucagon. Essa ilhota encontra-se rodeada por ácinos pancreáticos. Reproduzido de (BARDEESY; DEPINHO, 2002). 6
- Figura 5 – Sistema de canais hepáticos. Os hepatócitos produzem o suco biliar que é transportado para os canálculos biliares e fluem para o duto biliar. 7
- Figura 6 – Exemplo de LCA. Reproduzido da tese de doutorado de Henrique Velloso Ferreira Melo (2014). 10
- Figura 7 – Saída do programa ELDOgraph. Reproduzida da tese de doutorado de Tetsu Sakamoto (2016). A porcentagem de genes da OTU (neste caso ordens de mamíferos) que está em análise (ponta da aresta) é representada pela cor da aresta como indicado na legenda..... 12
- Figura 8. Representação simplificada da linhagem evolutiva humana. Reproduzido da dissertação de mestrado de Carlos Alberto Xavier Gonçalves (2016). 14
- Figura 9 – Modelos de arquivos de entrada (A) e saída (B) do SeedServer. 17
- Figura 10 - Esquema da via de secreção salivar. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); verde: genes originados entre Metazoa (n4) e Chordata (n8); rosa: genes originados entre Vertebrata (n10) e *Homo sapiens* (n31). Os números dentro dos círculos correspondem ao clado no qual os genes se originaram. Fonte: adaptado do KEGG, desenhado com PathVisio. 22
- Figura 11. Análise filogenética de AQP5. Resultado de BLAST (base RefSeq sem fragmentos) e de árvore filogenética construída com TaxOnTree, mostrando a proteína AQP5 humana em vermelho e a mais distante taxonomicamente no mesmo ramo em azul. LCAs indicados na legenda e nos ramos. Ao final da notação do ramo, entre parênteses, estão elencadas as classes, ordens e famílias distintas encontradas na análise: (1,12,28) significa a 28ª família, a 12ª ordem e a mesma classe da proteína humana (1). Ramos contendo (n21) em azul, (n22) em verde e (n17) em lilás foram colapsados para clareza. AQP5 parece restrita a Amniota (n17). 24

Figura 12. Análise filogenética de LPO. Resultado de BLAST (base RefSeq sem fragmentos) e de árvore filogenética construída com TaxOnTree, mostrando a proteína LPO humana em vermelho. Notações similares às da figura anterior. Ramos contendo (n11) em lilás e (n21) em azul foram colapsados para clareza. LPO parece restrita a Gnathostomata (n11). 26

Figura 13 - Esquema da via de secreção gástrica representando uma célula parietal. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); verde: genes originados entre Metazoa (n4) e Chordata (n8); rosa: genes originados entre Vertebrata (n10) e *Homo sapiens* (n31). Os números dentro dos círculos correspondem à linhagem na qual os genes se originaram. Fonte: Adaptado do KEGG..... 29

Figura 14 - Esquema da via de secreção pancreática. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); Verde: genes originados entre Metazoa (n4) e Chordata (n8); Rosa: genes originados entre Vertebrata (n10) e *Homo sapiens* (n31). Os números dentro dos círculos correspondem à linhagem na qual os genes se originaram. Fonte: Adaptado do KEGG..... 31

Figura 15 - Esquema da via de secreção biliar. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); Verde: genes originados entre Metazoa (n4) e Chordata (n8); Rosa: genes originados entre Vertebrata (n10) e *Homo sapiens*. Os números dentro dos círculos correspondem à linhagem na qual os genes se originaram. Fonte: Adaptado do KEGG. 37

Figura 16 – Diagrama de Venn das proteínas encontradas nas vias da digestão. Os números entre parênteses indicam o número total de genes em cada via. 41

Figura 17 - Distribuição acumulada dos genes para as vias de secreção salivar (verde), gástrica (azul), pancreática (roxo) e biliar (vermelho). Valores em porcentagem do total de genes em cada via. 42

Figura 18 – Distribuição absoluta dos genes na linhagem evolutiva humana. (A) Secreção salivar; (B) Secreção gástrica; (C) Secreção biliar; (D) Secreção pancreática. A área cinza representa o surgimento de todos os agrupamentos Kegg Orthology contendo genes humanos (base KEGG). 42

Figura 19 – Grafo gerado pelo programa ELDOgraph. O grafo mostra a proporção dos genes compartilhados entre diferentes espécies de organismos. (A) Genoma completo de humano; (B) Secreção salivar; (C) Secreção gástrica; (D) Secreção biliar; (E) Secreção pancreática. Os números sobre as arestas indicam a proporção dos genes humanos analisados que cada espécie são ELDO..... 44

Figura 20 – Grafo gerado pelo programa ELDOgraph. O grafo mostra a proporção dos genes compartilhados entre diferentes organismos. (A) Genoma completo de humano; (B) Secreção salivar; (C) Secreção gástrica; (D) Secreção biliar; (E) Secreção pancreática. Os números sobre as arestas indicam a proporção dos genes humanos analisados que cada espécie são ELDO. ... 46

Figura 21 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção salivar em espécies de mamíferos..... 49

Figura 22 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção gástrica em espécies de mamíferos.	50
Figura 23 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção pancreática em espécies de mamíferos.	51
Figura 24 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção biliar em espécies de mamíferos.	52
Figura 25. Gráfico das médias das distâncias filogenéticas das proteínas de mamífero. As barras pretas representam os primatas. As barras cinzas representam espécies cuja média das distâncias é menor que o primata com a maior média. O eixo da esquerda indica as distâncias filogenéticas dos genes em relação aos genes de humano. O eixo da direita indica a quantidade de proteínas encontradas em cada espécie. (A) Secreção salivar; (B) Secreção gástrica; (C) Secreção pancreática; (D) Secreção biliar.	54

LISTA DE TABELAS

Tabela 1 – Proteínas compartilhados pelas vias se secreção salivar, gástrica, pancreática e biliar	40
Tabela 2. Proteínas deletadas ou mantidas nas espécies de mamíferos	47

Sumário

1. INTRODUÇÃO	1
1.1. SISTEMA DIGESTIVO HUMANO.....	2
1.1.1. SALIVA	2
1.1.2. ESTÔMAGO.....	3
1.1.3. PÂNCREAS	6
1.1.4. BILE.....	7
1.2. CONCEITOS RELEVANTES.....	8
1.2.1. KEGG.....	8
1.2.2. PathVisio	8
1.2.3. SeedServer.....	8
1.2.4. GeneAges	9
1.2.5. Algoritmo LCA	10
1.2.6. Arcobaleno	10
1.2.7. Gene Ontology	11
1.2.8. ELDOgraph	11
1.3. A LINHAGEM EVOLUTIVA HUMANA.....	12
2. OBJETIVOS	15
2.1. Objetivo Geral	15
2.2. Objetivos Específicos.....	15
3. METODOLOGIA	16
3.1. Montagem do diagrama das vias	16
3.2. Determinação do clado de origem dos componentes das vias	16
3.3. Colorização das vias.....	17
3.3.1. Colorização das vias de acordo com o LCA dos componentes.....	17
3.4. ELDOgraph.....	18
4. RESULTADOS.....	19
4.1. Descrição das vias de secreção do sistema digestivo humano	19
4.1.1. Descrição da via de secreção salivar	19
4.1.2. Secreção gástrica.....	26
4.1.3. Secreção Pancreática	30
4.1.4. Secreção Biliar	35
4.2. Comparando as vias de secreção do sistema digestivo	40
4.3. Comparando as vias de secreção do sistema digestivo em mamíferos.....	43

5. DISCUSSÃO.....	55
6. CONCLUSÃO	59
7. REFERÊNCIAS.....	60

1. INTRODUÇÃO

A manutenção dos processos celulares que sustentam a vida só é possível graças à utilização de energia. Esta pode ser obtida de fontes diversas, metabolizadas de maneiras distintas. Quanto mais energia um organismo dispõe, mais processos metabólicos ele é capaz de realizar. Desde o surgimento das primeiras células a obtenção de energia para a reprodução é um mecanismo básico para garantir a manutenção da vida. Assim, a origem de organismos multicelulares, a diferenciação de células com funções distintas e o surgimento de tecidos diferenciados promoveram maior sucesso na obtenção de alimentos e seu processamento compartimentalizado. O desenvolvimento de um trato digestivo, formado por um conjunto de órgãos especializados na digestão dos alimentos, representa um grande passo na evolução dos animais, uma vez que esse sistema permite um maior aproveitamento dos nutrientes ingeridos. O surgimento do trato digestivo permite que os alimentos ingeridos sejam processados extracelularmente, preparando os nutrientes para serem absorvidos mais facilmente pelas células de modo que estas possam aproveitar melhor a energia dos alimentos. O maior aproveitamento dos nutrientes permitiu aos primeiros animais acumularem essa energia que conseqüentemente pôde ser realocada para auxiliar no desenvolvimento de outras funções como a sua mobilidade ou na formação de um corpo e cérebro maiores. Uma descrição detalhada sobre o processo digestivo em Metazoa pode ser encontrada no trabalho clássico de YONGE (1937). Esse trabalho de anatomia comparada representa uma valiosa fonte de informações a respeito das estruturas encontradas entre os clados de metazoários. No entanto, muito pouco é explorado a respeito das proteínas envolvidas nos processos digestivos, o que se deve à escassez de informações genômicas na época de sua publicação.

Atualmente, apesar de muitos trabalhos discorrerem sobre genômica comparativa como resultado da grande quantidade de informações oriundas do sequenciamento de genomas completos, a maioria deles abrangem diversos aspectos relacionados à biologia do organismo sequenciado, não se atendo a um assunto específico com profundidade. Embora o conhecimento sistêmico de diversos processos seja reportado em muitas bases de dados, como KEGG (OGATA et al., 1999) e Reactome (CROFT et al., 2011), estudos de genômica comparativa voltados para temas específicos são escassos.

1.1. SISTEMA DIGESTIVO HUMANO

O sistema digestivo em todos os metazoários é homólogo e é o primeiro sistema a se formar durante o desenvolvimento embrionário (SMITH et al., 2000). Apesar disso, sua estrutura apresenta variações quando comparado entre os diversos taxa, que muitas vezes são devidas aos diferentes hábitos alimentares mantidos pelos animais dos diferentes grupos (YONGE, 1937). Algumas estruturas podem não estar presentes ou terem se perdido em algumas espécies (SMITH et al., 2000; YONGE, 1937). O mesmo pode ter ocorrido com o padrão de expressão de alguns genes (SMITH et al., 2000).

O processo de digestão consiste na quebra do alimento, por mecanismos químicos e mecânicos, em partes cada vez menores, até que restem apenas moléculas orgânicas simples. Estas moléculas, por sua vez, são absorvidas e distribuídas para as demais células.

1.1.1. SALIVA

A digestão tem início na boca onde a mastigação e a secreção de saliva representam a primeira etapa da digestão (Figura 1). Durante essa fase, o alimento é triturado pelos dentes com auxílio da língua. Nesse processo, a saliva desempenha um papel importante limpando e lubrificando a boca, umedecendo o alimento e secretando enzimas que fazem a digestão inicial do alimento. Essa digestão inicial é possível graças às amilases salivares capazes de digerir carboidratos (AMADO et al., 2010; HELMERHORST; OPPENHEIM, 2007). Após esse pré-processamento na boca, o alimento está úmido e com uma área de superfície maior, o que facilita os processos digestivos que se seguem no trato digestivo. A saliva liberada na boca é uma mistura do conteúdo de três tipos de glândulas salivares, a glândula parótida, a glândula submandibular e a glândula sublingual. Cada tipo de glândula salivar pode ser classificado pela sua composição proteica que, por sua vez, varia de acordo com a dieta de cada espécie e, por isso, a classificação das glândulas não é a mesma para todas elas (ARMSTRONG; TURTURRO, 2013; PROCTOR; CARPENTER, 2007).

A glândula salivar é composta por ácino (ou alvéolo), duto secretor e duto coletor. O ácino corresponde a um grupo de células secretoras que podem ser classificadas como serosas, mucosas (ou mucilaginosas) ou mistas (ARMSTRONG; TURTURRO, 2013). As secreções classificadas como serosas contêm uma baixa fluidez devido à presença de

proteínas em concentrações maiores; as glândulas parótidas são as principais produtoras desse tipo de secreção. Já as secreções mucosas contêm mucina e são mais fluidas. As glândulas do tipo submandibular e sublingual produzem secreção serosa e mucosa (ARMSTRONG; TURTURRO, 2013; HALL; GUYTON, 2011).

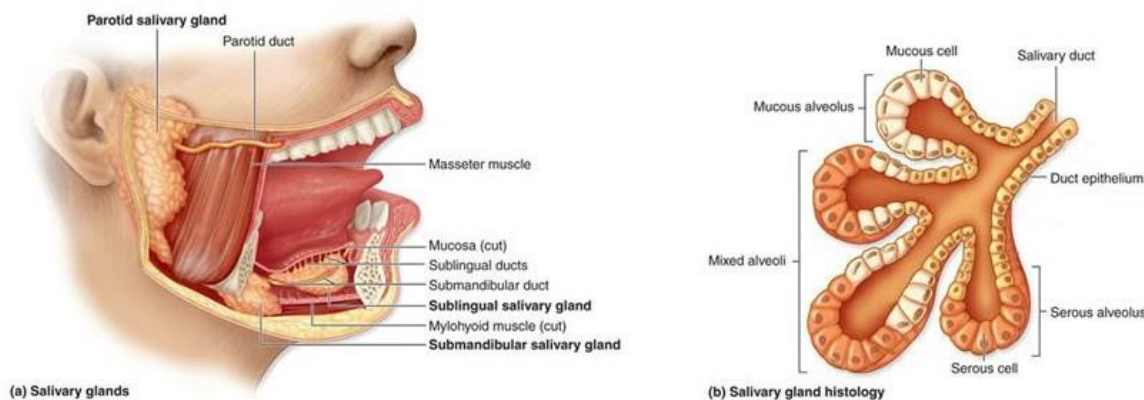


Figura 1 – Anatomia das glândulas salivares. (A) Localização das glândulas; (B) representação geral da histologia da glândula salivar. Adaptado de Mescher AL. Junqueira's Basic Histology, 13e; 2013.

As glândulas salivares produzem entre um e 1,5 litros de saliva por dia, sendo a glândula submandibular a principal produtora de saliva (71%), seguida pela parótida (25%) e sublingual (3-4%) (ARMSTRONG; TURTURRO, 2013).

1.1.2. ESTÔMAGO

Após o alimento ser deglutido, ele passa pela faringe e esôfago até chegar ao estômago onde é misturado ao suco gástrico que dá continuidade ao processo de digestão (GELBERG, 2014). As glândulas gástricas são compostas por uma variedade de células epiteliais sendo a célula parietal a responsável por produzir o ácido clorídrico (HCl), principal componente do suco gástrico (Figura 2 e Figura 3). O ácido gástrico auxilia no processo de digestão de proteínas, pois converte o pepsinogênio em pepsina, que é sua forma ativa. Além disso, a absorção de ferro, cálcio, vitamina B12 e alguns medicamentos é facilitada pelo HCl presente no estômago (SCHUBERT, 2014). Outra função do HCl é a de esterilizar os alimentos ingeridos, uma vez que a acidez do meio inibe o desenvolvimento bacteriano (KOPIC; MUREK; GEIBEL, 2010; SCHUBERT, 2014).

O estômago é dividido em três regiões anatômicas (fundo, corpo e ântró) e duas regiões funcionais, chamadas de mucosa oxíntica e pilórica, que se distinguem pela

presença de glândulas oxínticas (ou glândulas gástricas) e pilóricas, respectivamente (Figura 2). Essas glândulas são compostas por diferentes combinações de tipos celulares e estão localizadas em diferentes regiões. As glândulas oxínticas ocupam cerca de 80% do tecido estomacal e se localizam na região do fundo e do corpo do estômago, enquanto as glândulas pilóricas ocorrem na região antral, ocupando os 20% distais do estômago (SCHUBERT, 2014).

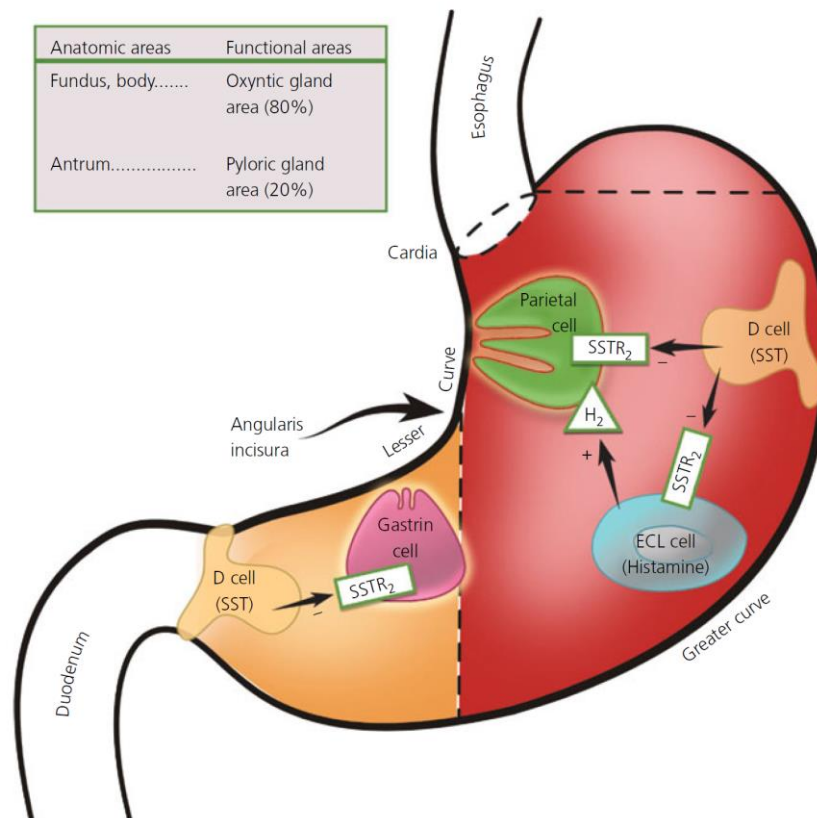


Figura 2 – Divisão fisiológica do estômago. Representação anatômica das regiões do estômago indicando os principais grupos de células presentes. As áreas denominadas Fundo e Corpo estão coloridas de vermelho, e o antrum de laranja. Reproduzido de (SCHUBERT, 2014).

O epitélio estomacal invagina-se formando fossetas e glândulas, ambos compostos por diferentes tipos celulares que secretam substâncias responsáveis por regular a secreção do ácido gástrico (Figura 3). Os principais tipos celulares encontrados nas glândulas oxínticas são as células parietais, responsáveis pela secreção do ácido clorídrico; as células semelhantes às enterocromafins (ECL - *enterochromaffin-like cells*), produtoras e secretoras de histamina; as células delta (ou células D), que produzem o hormônio peptídeo somatostatina; as células principais, que produzem pepsinogênio; e as células enterocromafins, que produzem serotonina, peptídeo natriurético atrial (*atrial natriuretic peptide* - ANP) e adrenomedulina (SCHUBERT, 2010, 2014; SCHUBERT;

PEURA, 2008). Nas glândulas pilóricas são encontradas apenas as células D, enterocromafins e as células da gastrina (ou células G), produtoras do hormônio gastrina (SCHUBERT, 2014). Em ambas as glândulas são encontradas as células mucosas de superfície e do colo, cuja função é secretar muco que auxilia na proteção do tecido estomacal.

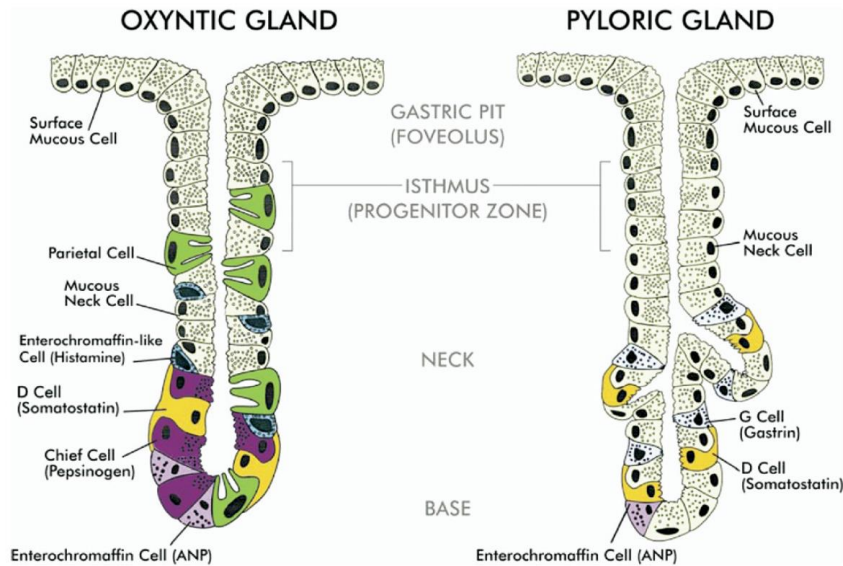


Figura 3 – Anatomia das glândulas gástricas. As glândulas oxíntica e pilórica possuem diferentes grupos de células, como indicado na figura. Reproduzido de (SCHUBERT; PEURA, 2008).

A célula parietal, secretora de HCl, possui adaptações que maximizam sua função secretora. Essa célula conta com invaginações da membrana apical, o canalículo apical, que funciona como microvilosidades, aumentando a área de superfície da membrana. Além disso, seu citoplasma possui estruturas membranosas similares a vesículas chamadas de Tubulovesículas. É na membrana dessas estruturas que ficam armazenadas as H/K-ATPases, responsáveis pelo bombeamento de prótons para o lúmen da glândula oxíntica onde se forma o ácido clorídrico (FORTE; ZHU, 2010; SCHUBERT, 2014; YAO; FORTE, 2003).

1.1.3. PÂNCREAS

O pâncreas é uma glândula acessória do sistema digestivo que tem um importante papel fisiológico (Figura 4A). No pâncreas são produzidas e secretadas diversas enzimas que realizam o processo de digestão dos nutrientes; dentre estas enzimas, encontram-se amilases, lipases e proteases (HALL; GUYTON, 2011). As enzimas são produzidas nas células acinares do pâncreas e ficam armazenadas em vesículas chamadas grânulos de zimogênio (Figura 4B e C). Além disso, grandes quantidades de solução de bicarbonato de sódio são secretadas pelas células do ducto (

Figura 4C). Como glândula endócrina, o pâncreas é responsável por produzir hormônios como a insulina e glucagon, importantes para o funcionamento do sistema endócrino (Figura 4D).

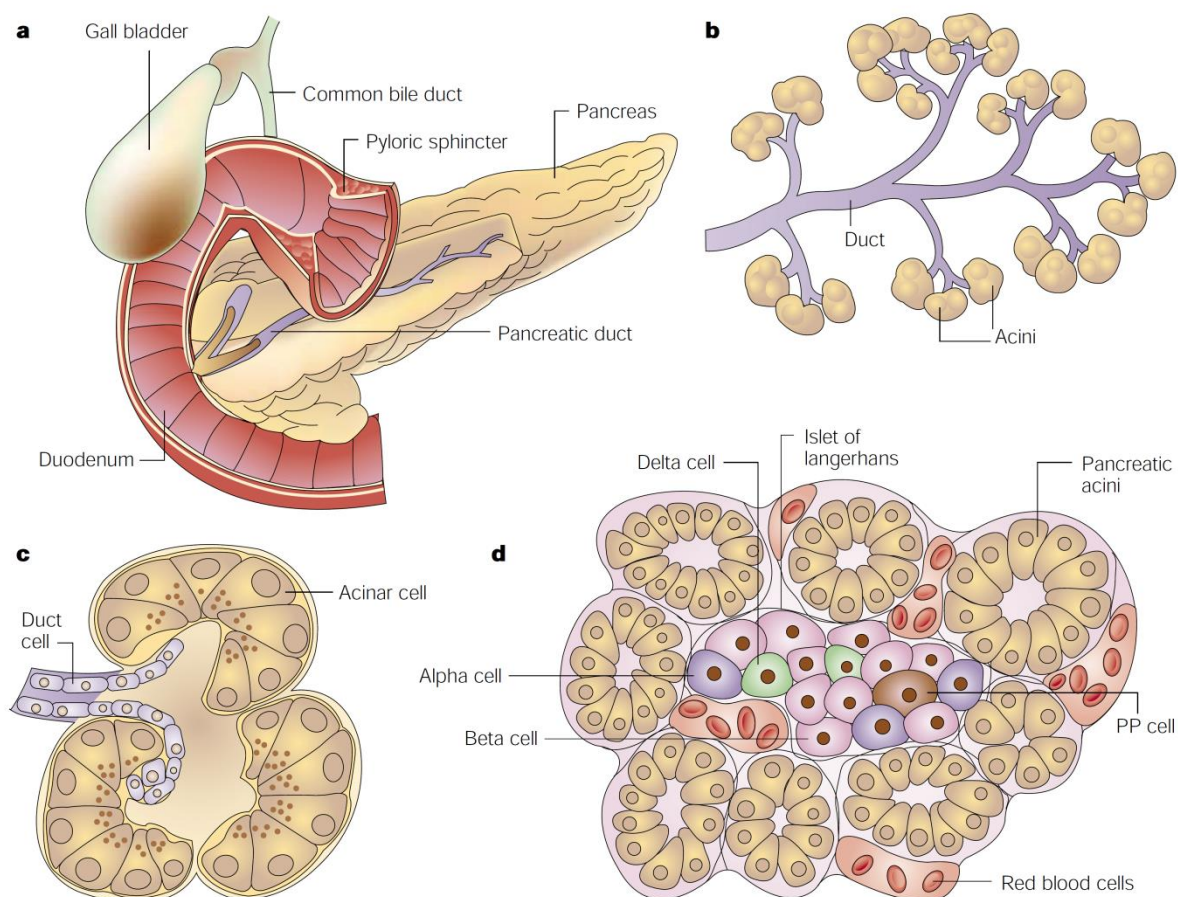


Figura 4 – Anatomia do pâncreas. (A) Representação anatômica do pâncreas; (B) Estrutura das glândulas exócrinas pancreáticas; (C) Representação do ácino pancreático, onde é formado o suco pancreático; (D) Anatomia de uma ilhota de Langherans que secretam insulina e glucagon. Essa ilhota encontra-se rodeada por ácinos pancreáticos. Reproduzido de (BARDEESY; DEPINHO, 2002).

1.1.4. BILE

A bile (suco biliar) é um fluido produzido pelos hepatócitos, sendo composto por água, sais biliares, pigmentos, aminoácidos, colesterol, esteróis, ácidos graxos, enzimas e sais inorgânicos. Apesar de ser produzida no fígado, a bile sofre modificações à medida que é absorvida e secretada pelo sistema de transporte do duto biliar a caminho da vesícula biliar (BOYER, 2013). A bile tem diversas funções, entre elas a eliminação de substâncias nocivas exógenas de natureza lipofílica e a eliminação de colesterol; além disso, os sais biliares ajudam a emulsificar os ácidos graxos, facilitando sua absorção pelo intestino (BOYER, 2013; O'BYRNE et al., 2003).

A formação da bile ocorre em um processo de filtração gerado pelo gradiente de concentração decorrente do transporte ativo de solutos para o lúmen no canalículo biliar. Inicialmente, a bile é secretada pelos hepatócitos para os canalículos biliares (Figura 5). Ao fluir pelo duto biliar, ela sofre uma série de transformações decorrentes de processos de secreção e reabsorção das células que compõem o duto, os colangiócitos (ESTELLER, 2008). Após passar pelo canalículo, a bile passa para o duto biliar onde sofre um processo de reabsorção e secreção que torna o suco biliar mais fluido que o original (HOWE et al., 2017; KOSTERS; KARPEN, 2008; MEIER; STIEGER, 2002; TAKIKAWA, 2002; ZELCER et al., 2009).

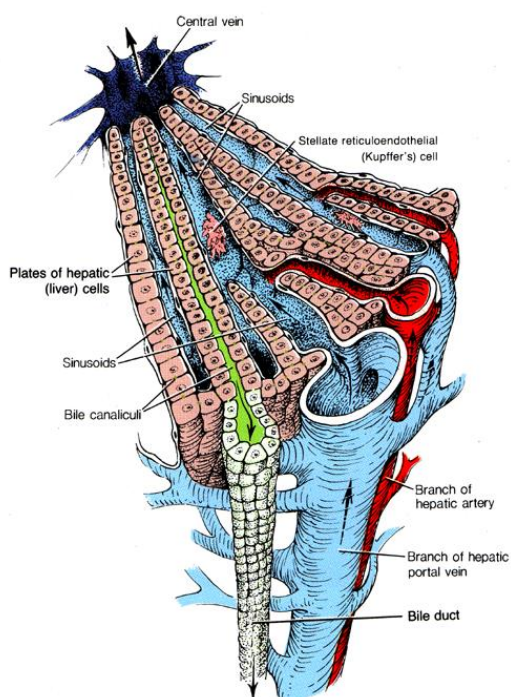


Figura 5 – Sistema de canais hepáticos. Os hepatócitos produzem o suco biliar que é transportado para os canalículos biliares e fluem para o duto biliar.

1.2. CONCEITOS RELEVANTES

1.2.1. KEGG

A base de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*) reúne informações genômicas que representam dados genéticos (genes, genomas, homologia), vias metabólicas e informações químicas (reações químicas, compostos e ligantes) (OGATA et al., 1999). Nessa base de dados as informações são integradas na forma de diagramas de modo que os genes e produtos gênicos podem ser associados com as reações, interações e processos biológicos. É partindo desse princípio que a informação biológica é armazenada em três bases de dados: GENES, que reúne informações sobre anotações de funções de genes e informações a respeito de grupos de ortólogos; a base PATHWAY que corresponde a representações gráficas de processos celulares; e a base LIGAND, que fornece informações sobre compostos químicos, moléculas e reações enzimáticas. O conjunto de informações catalogado no KEGG representa um valioso recurso computacional para interpretação de dados biológicos. As vias utilizadas neste trabalho foram obtidas de KEGG PATHWAYS.

1.2.2. PathVisio

O PathVisio é um software para análises de vias metabólicas, onde é possível não só visualizá-las mas também editá-las (KUTMON et al., 2015; VAN IERSEL et al., 2008). Esse software também oferece recursos para a visualização de dados e para análises estatísticas, bem como permite a instalação de plugins para adição de outras funcionalidades ao programa (KUTMON et al., 2015). PathVisio é o software de preparação de vias para depósito na base WikiPathways (<http://wikipathways.org>).

1.2.3. SeedServer

O SeedServer é um programa que integra diferentes abordagens para agrupamento de proteínas homólogas (GUEDES, 2013). Para isso, o programa recebe o identificador UniProtKB de uma proteína, denominada *seed*, recupera sua sequência e os identificadores taxonômicos relacionados a ela. O agrupamento se inicia com a execução do programa Seed Linkage que agrupa possíveis ortólogos (BBHs, *bidirectional best*

hits) a partir da sequência *seed* e também dos seus parálogos, recrutados pelo programa. Em seguida é realizada uma consulta aos bancos KO/UEKO para verificar se a *seed* está presente, e caso esteja os membros complementares desse grupo são recrutados, em caso negativo o agrupamento contaria apenas com os homólogos recuperados pelo programa Seed Linkage. No final desse processo o programa faz a validação do cluster com PSIBLAST (RIBEIRO, 2013) e informações sobre as proteínas recrutadas, como descrição, tamanho, presença de curadoria manual (Swiss-Prot), são adicionadas e apresentadas no arquivo de saída do programa. Seed Linkage (BARBOSA-SILVA et al., 2008) e UEKO (FERNANDES et al., 2008) foram desenvolvidos no laboratório de Biodados.

1.2.4. GeneAges

GeneAges consiste em um trabalho de síntese utilizando vários métodos de agrupamento de genes ortólogos com intuito de inferir a idade dos genes (LIEBESKIND; MCWHITE; MARCOTTE, 2016). Foram utilizados 13 algoritmos de inferência de ortólogos, que inclui o InParanoid (O'BRIEN, 2004), Panther (MI; THOMAS, 2009), Grupos OMA (ALTENHOFF et al., 2015), EggNOG (HUERTA-CEPAS et al., 2016; JENSEN et al., 2007), Ensembl-Compara (VILELLA et al., 2008) e PhylomeDB (HUERTA-CEPAS et al., 2007). A base resultante classifica a ancestralidade dos genes em oito clados: Cellular organisms, Euk+Bacteria, Euk+Archaea, Eukaryota, Opisthokonta, Eumetazoa, Vertebrata e Mammalia, especificando para os níveis mais ancestrais se o gene ocorre em duplas ou nas três vertentes da vida (classificação Cellular organisms). Neste trabalho comparamos esses três grupos a Cellular organisms nas linhagens do NCBI Taxonomy. Assim, se inferimos origem em Euteleostomi consideramos de acordo com Geneages se estiver mapeado em Vertebrata, grupo que inclui o clado Euteleostomi. Geneages já havia sido utilizado no estudo de conteúdo de estrutura 2D em proteínas realizado por Ricardo Vialle (VIALLE, 2017) em nosso laboratório.

1.2.5. Algoritmo LCA

O ancestral comum mais recente (LCA – *Lowest Common Ancestor*) de dois nós (x e y) em uma árvore é o nó anterior mais profundo que tem tanto x quanto y como descendentes (Figura 6).

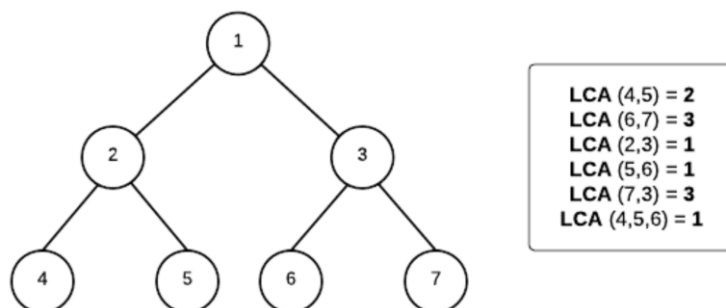


Figura 6 – Exemplo de LCA. Reproduzido da tese de doutorado de Henrique Velloso Ferreira Melo (2014).

Uma implementação do algoritmo LCA como *webservice* foi desenvolvido por MELO (2014), que carrega a árvore taxonômica completa na memória e mapeia os identificadores taxonômicos fornecidos como entrada nessa árvore, retornando o LCA. Esta aplicação está disponível em <http://biodados.icb.ufmg.br/services>.

1.2.6. Arcobaleno

O ARCoBALeno (*Application for Rapidly COloring Biological pathways by Ancestrality or gene ontoLogy-Enriched iNformatiOn*) é uma ferramenta que colore automaticamente arquivos do PathVisio (com extensão .gpml) com foco em estudos de evolução. Essa ferramenta desenvolvida por Carlos Alberto Xavier Gonçalves durante seu mestrado em Bioinformática (GONÇALVES, 2016) está disponível para uso em interface web em <http://biodados.icb.ufmg.br/arcobaleno/>. Essa ferramenta colore cada gene em uma via metabólica de acordo com sua origem (LCA) e com informações extraídas da base Gene Ontology (ASHBURNER et al., 2000) como função molecular mais recente (LMF – *Last Molecular Function*) ou do processo biológico mais recente (LBP – *Last Biological Process*).

1.2.7. Gene Ontology

O *Gene Ontology* (GO) é o resultado de uma iniciativa de unificação da informação biológica relacionada a genes e produtos gênicos em todas as espécies (ASHBURNER et al., 2000). Esse consórcio cataloga termos próprios, também referidos como ontologias, que abrangem os mais diversos aspectos biológicos. As ontologias são divididas em três categorias: processos biológicos, que reúne informações sobre os fenômenos biológicos dos quais os genes ou produtos gênicos participam; funções moleculares, que definem a atividade química que os produtos gênicos desempenham; e componentes celulares, que representam as partes da célula onde os produtos gênicos podem ser encontrados.

Todos os termos são organizados de forma hierárquica na base de dados, sendo alguns termos considerados ancestrais de outros. Assim, os termos do GO representam uma ferramenta útil para a caracterização de produtos gênicos.

1.2.8. ELDOgraph

O programa ELDOgraph (SAKAMOTO, 2016) realiza análises comparativas com base nas distâncias evolutivas entre espécies em uma árvore, o que permite verificar quais espécies possuem mais genes mais similares com os genes da espécie em análise. O termo ELDO deriva do termo LDO (*Least Divergent Ortholog*) que representa como o nome sugere, o ortólogo que menos divergiu entre duas espécies. Assim, um par de proteínas de espécies distintas é um LDO se a distância filogenética entre elas for menor que a distância entre as demais proteínas dessas duas espécies. Já o ELDO (*External Least Diverged Ortholog*) de uma proteína representa o ortólogo mais próximo que pertence a um táxon distinto (externo) ao táxon da proteína em análise. Assim, diferente do LDO, que é uma propriedade associada a cada par de organismos distintos, o ELDO é uma propriedade individual de cada proteína. O programa ELDOgraph permite que o usuário escolha um nível taxonômico (gêneros, famílias, ordens, etc.) para que o programa procure para cada proteína o ortólogo menos divergente que pertença a um táxon distinto ao da proteína em análise no nível taxonômico escolhido. Por último, a ferramenta ELDOgraph oferece aos usuários uma interface que permite uma fácil visualização das relações entre as espécies na forma de grafos (Figura 7). O produto da análise é a OTU com mais proteínas similares, independentemente dela ser ou não um grupo irmão na

árvore filogenética e foi desenvolvido para utilização em genômica comparativa independente de filogenia.

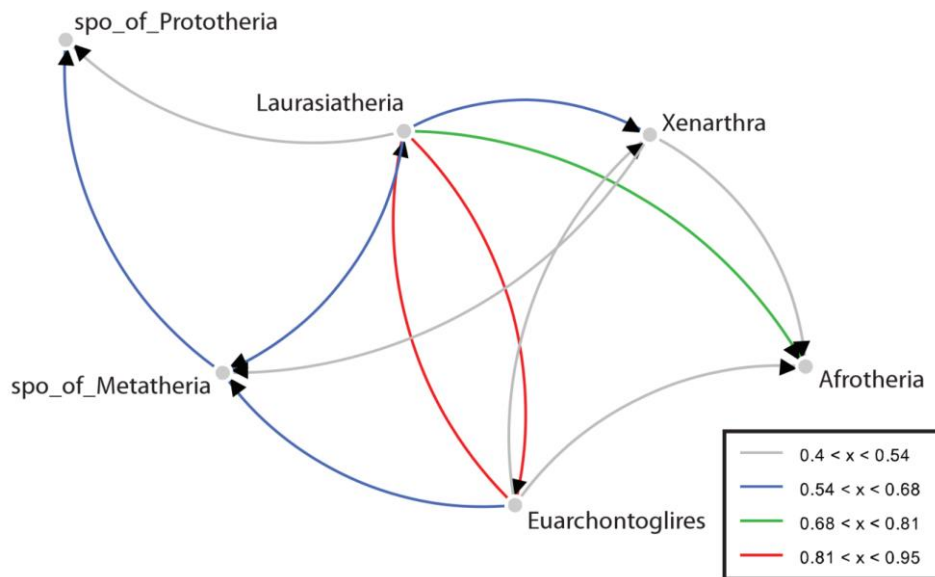


Figura 7 – Saída do programa ELDOgraph. Reproduzida da tese de doutorado de Tetsu Sakamoto (2016). A porcentagem de genes da OTU (neste caso ordens de mamíferos) que está em análise (ponta da aresta) é representada pela cor da aresta como indicado na legenda.

1.3. A LINHAGEM EVOLUTIVA HUMANA

Nesse trabalho utilizamos como modelo o sistema digestivo humano, assim, para facilitar o entendimento do trabalho apresentamos a seguir os 31 níveis da linhagem humana (Figura 8) encontrada no NCBI Taxonomy, seguida de uma breve descrição.

Nível 1 – Organismos celulares. É o táxon mais ancestral da árvore da vida. Compreende todos os organismos constituídos por células. Isso significa que esse táxon agrupa bactérias, arqueias e eucariotos.

Nível 2 – Eukaryota. Compreende todos os organismos que possuem membrana nuclear (carioteca). Isso significa que esse táxon agrupa protistas, plantas, fungos e animais.

Nível 3 – Opisthokonta. Engloba fungos e animais.

Nível 4 – Metazoa. Engloba todos os animais. Isso significa que esse táxon agrupa tanto os animais basais (poríferos, placozoários) quanto os animais verdadeiros.

Nível 5 – Eumetazoa. Contém os animais verdadeiros, isto é, os animais que possuem tecidos verdadeiros. Isso significa que esse táxon agrupa tanto os animais com simetria radial (cnidários e ctenóforos) quanto os animais com simetria bilateral.

Nível 6 – Bilateria. Contém os animais com simetria bilateral. Estão incluídos nesse táxon os artrópodes, moluscos, anelídeos, platelmintos, nematelmintos e deuterostomados.

Nível 7 – Deuterostomia. Engloba equinodermos (ouriços e estrelas-do-mar) e cordados.

Nível 8 – Chordata. Compreende todos os animais que apresentam notocorda durante o desenvolvimento embrionário. Isso significa que esse táxon agrupa cefalocordados (anfioxo), tunicados e craniados.

Nível 9 – Craniata. Compreende todos os animais que desenvolvem crânio. No entanto, devido à incerteza na classificação taxonômica, todos os craniados atualmente são classificados como vertebrados.

Nível 10 – Vertebrata. Compreende todos os animais que possuem esqueleto. Isso significa que esse táxon agrupa peixes sem mandíbula (lampreias e peixes-bruxa) e vertebrados com mandíbula.

Nível 11 – Gnathostomata. Contém os vertebrados com mandíbula. Estão incluídos nesse táxon tanto os peixes de esqueleto cartilaginoso (tubarão e raia) quanto os vertebrados de esqueleto ósseo.

Nível 12 – Teleostomi. Todos os teleostomi atualmente são classificados como euteleostomi.

Nível 13 – Euteleostomi. Engloba os peixes extintos Acanthodii e os vertebrados com esqueleto ósseo verdadeiro.

Nível 14 – Sarcopterygii. Contém os vertebrados com membros lobados com suporte muscular.

Nível 15 – Dipnotetrapodomorpha. Engloba peixes pulmonados e tetrápodes.

Nível 16 – Tetrapoda. Compreende todos os vertebrados com quatro membros. Isso significa que esse táxon agrupa anfíbios e amniotas.

Nível 17 – Amniota. Compreende todos os vertebrados que apresentam âmnio durante o desenvolvimento embrionário. Estão incluídos nesse táxon répteis, aves e mamíferos.

Nível 18 – Mammalia. Engloba todos os mamíferos. Isso significa que esse táxon agrupa prototérios (ou monotremados: ornitorrinco e équidna), metatérios e eutérios.

Nível 19 – Theria. Engloba metatérios (ou marsupiais, canguru, coala, gambá e outros) e eutérios.

Nível 20 – Eutheria. Contém os placentários verdadeiros. Alguns exemplos são elefante e peixe-boi (Afrotheria) e boreoeutérios.

Nível 21 – *Boreoeutheria*. Contém placentários agrupados a partir de dados moleculares excluídos os Afrotheria. Alguns exemplos são carnívoros, baleia, ruminantes, porco, camelo, morcegos, cavalo, rinoceronte, anta (todos Laurasiatheria) e Euarchontoglires.

Nível 22 – *Euarchontoglires*. Contém lêmures, coelhos, roedores, musaranhos e primatas.

Nível 23 – *Primates*. Engloba todos os primatas.

Nível 24 – *Haplorrhini*. Engloba tarsos e símios.

Nível 25 – *Simiiformes*. Contém os símios, isto é, macacos do novo mundo, macacos do velho mundo e hominoídeos.

Nível 26 – *Catarrhini*. Engloba macacos do velho mundo e hominoídeos.

Nível 27 – *Hominoidea*. Contém os hominoídeos, isto é, gibões, orangotangos e hominíneos.

Nível 28 – *Hominidae*. Engloba orangotangos e hominíneos.

Nível 29 – *Homininae*. Contém os hominíneos, isto é, gorilas, bonobos, chimpanzés e humanos.

Nível 30 – *Homo*. Engloba os humanos extintos e humano atual.

Nível 31 – *Homo sapiens*. Humano atual

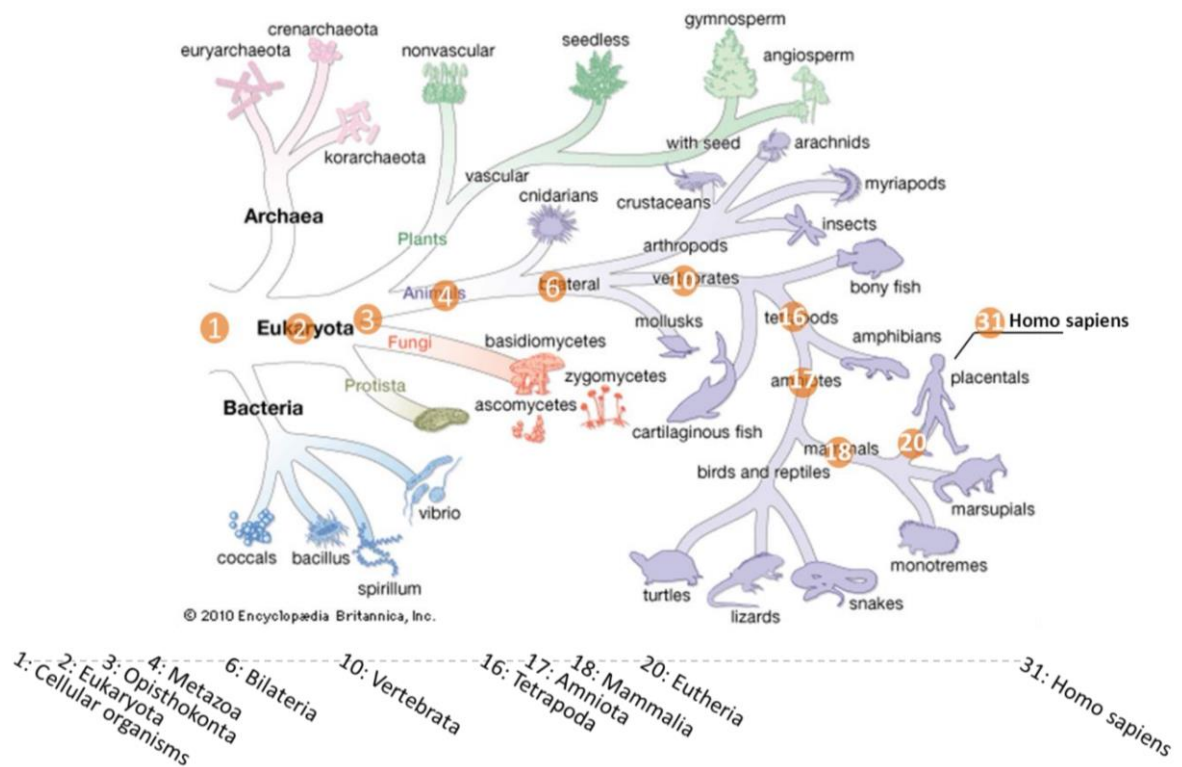


Figura 8. Representação simplificada da linhagem evolutiva humana. Reproduzido da dissertação de mestrado de Carlos Alberto Xavier Gonçalves (2016).

2. OBJETIVOS

2.1. Objetivo Geral

Estudar a origem sequencial dos genes envolvidos na heterotrofia usando como modelo o sistema digestivo humano.

2.2. Objetivos Específicos

- Elaborar diagramas com extensão *.gpml* das vias de secreção salivar, gástrica, pancreática e biliar com base nos modelos disponíveis no KeggPathways;
- Estimar o clado LCA das proteínas presentes nas vias de secreção do sistema digestivo humano;
- Analisar a origem dos processos de secreção do sistema digestivo humano com base na origem dos seus genes;
- Realizar uma genômica comparativa e filogenia molecular das proteínas presentes nas vias em espécies de mamíferos;
- Disponibilizar os diagramas das vias gerados, juntamente com a descrição do processo e do LCA obtido, na base de dados WikiPathways (<http://www.wikipathways.org/index.php/WikiPathways>).

3. METODOLOGIA

3.1. Montagem do diagrama das vias

As vias estudadas nesse trabalho representam modelos dos processos encontrados em humanos. Elas estão depositadas no KEGG (OGATA et al., 1999) e foram reproduzidas usando o software PathVisio (KUTMON et al., 2015; VAN IERSEL et al., 2008). Símbolos específicos foram utilizados para representar os componentes (produtos gênicos e compostos químicos), compartimentos celulares e interações e/ou reações. Os produtos gênicos e compostos químicos estão representados como retângulos. As interações entre os componentes são representadas com setas (interação positiva, ativação) e barras em T (interação negativa, repressão). As setas contínuas indicam interação molecular ou relação direta, enquanto as setas pontilhadas indicam relação indireta. Subvias são coloridas em vermelho. Os parálogos de proteínas foram organizados em um quadro abaixo do esquema da via. Os nomes dos componentes foram baseados nos identificadores usados na base UniProtKB.

3.2. Determinação do clado de origem dos componentes das vias

O identificador UniProtKB de cada produto gênico representado nas vias foi recuperado manualmente e associado ao nome de seu componente correspondente. Nesta etapa, foram selecionados os identificadores curados manualmente (Swiss-Prot).

A inferência da origem dos componentes de cada via foi realizada em duas etapas. Primeiramente os identificadores UniProtKB foram utilizados como *input* para a ferramenta SeedServer que realiza a busca por homólogos desses componentes. O SeedServer pode ser encontrado para acesso local ou via *Bioinformatics Open Web Services* (BOWS) (<http://maxixe.icb.ufmg.br:8080/BOWSWeb/index.jsp>). Nesse trabalho foi utilizada apenas a versão local que recebe como parâmetro de entrada um arquivo de texto contendo o identificador UniProtKB e os parâmetros para busca dos ortólogos. No presente estudo foram utilizados os parâmetros padrão (Figura 9A). Como resultado o SeedServer fornece uma lista dos identificadores das proteínas consideradas homólogas pelo programa. A Figura 9B ilustra o arquivo de saída do SeedServer listando os homólogos da proteína Lactoperoxidase humana (acesso P22079).

A

```
<processID>output</processID>
<SLtable>allorg</SLtable>
<input>P22079</input>
<parameter_r>50</parameter_r>
<parameter_c>50</parameter_c>
<parameter_s>50</parameter_s>
<parameter_l>0.3</parameter_l>
<e_value>10</e_value>
<self_score>0.7</self_score>
```

B

uniprot_ID	cluster_number	SL	KO	size	txi	clade	validated	FSI	description
P22079	1	2	1	712	9606	1	1	1	PERL_HUMAN Lactoperoxidase OS=Homo sapiens GN=LPO PE=1 SV=2
H2QDI8	1	1	2	711	9598	1	1	1	H2QDI8_PANTR Uncharacterized protein OS=Pan troglodytes GN=LPO PE=4 SV=1
H2NTI4	1	1	2	1228	9601	1	1	1	H2NTI4_PONAB Uncharacterized protein OS=Pongo abelii GN=LPO PE=4 SV=1
G7NHR9	1	1	2	715	9544	1	1	1	G7NHR9_MACMU Putative uncharacterized protein OS=Macaca mulatta GN=EGK_08576 PE=4 SV=1
F7BAC5	1	1	2	715	9544	1	1	1	F7BAC5_MACMU Uncharacterized protein OS=Macaca mulatta GN=LPO PE=4 SV=1
F7BAB9	1	1	1	1252	9544	1	1	1	F7BAB9_MACMU Uncharacterized protein OS=Macaca mulatta GN=LPO PE=4 SV=1
G7PUF6	1	1	2	715	9541	1	1	1	G7PUF6_MACEA Putative uncharacterized protein OS=Macaca fascicularis GN=EGM_07763 PE=4 SV=1
G1R9M9	1	1	2	712	61853	1	1	1	G1R9M9_NOMLE Uncharacterized protein OS=Nomascus leucogenys GN=LPO PE=4 SV=1
G1LZG5	1	1	2	712	9646	1	1	1	G1LZG5_AILME Uncharacterized protein OS=Ailuropoda melanoleuca GN=LPO PE=4 SV=1
F7C3Z3	1	1	2	710	9796	1	1	1	F7C3Z3_HORSE Uncharacterized protein OS=Equus caballus GN=LPO PE=4 SV=1
M3VY73	1	1	2	712	9669	1	1	1	M3VY73_MUSEP Uncharacterized protein OS=Mustela putorius furo GN=LPO PE=4 SV=1
E2RPI9	1	1	2	718	9615	1	1	1	E2RPI9_CANFA Uncharacterized protein OS=Canis familiaris GN=LPO PE=4 SV=1
G5BJT5	1	1	2	741	10181	1	1	1	G5BJT5_HETGA Lactoperoxidase OS=Heterocephalus glaber GN=GW7_05214 PE=4 SV=1
HDWL87	1	1	2	1635	30611	1	1	1	HDWL87_OTOGA Uncharacterized protein OS=Otolemur garnettii GN=LPO PE=4 SV=1
Q9MZY2	1	1	2	712	9940	1	1	1	Q9MZY2_SHEEP Airway lactoperoxidase OS=Ovis aries PE=1 SV=1
W5PGP1	1	1	2	717	9940	1	1	1	W5PGP1_SHEEP Uncharacterized protein OS=Ovis aries GN=LPO PE=4 SV=1
G1SXQ6	1	1	2	711	9986	1	1	1	G1SXQ6_RABIT Uncharacterized protein OS=Oryctolagus cuniculus GN=LPO PE=4 SV=1
A3F9D6	1	1	1	782	9925	1	1	1	A3F9D6_CAPHI Lactoperoxidase OS=Capra hircus GN=LPO PE=1 SV=1
G1P676	1	1	2	1060	59463	1	1	1	G1P676_MYOLU Uncharacterized protein OS=Myotis lucifugus GN=LPO PE=4 SV=1

Figura 9 – Modelos de arquivos de entrada (A) e saída (B) do SeedServer.

O resultado fornecido pelo SeedServer foi utilizado posteriormente para determinar o clado de origem do grupo de homólogos usando o algoritmo do LCA (*Lowest Common Ancestor*). A base desse algoritmo é obter as linhagens taxonômicas de todos os organismos representados em um grupo de homólogos e determinar o clado mais recente que seja comum em todas as linhagens taxonômicas. Para determinar o LCA de cada grupo de homólogos obtidos em escala, foi utilizado um script que tem como entrada o arquivo de saída do SeedServer e que requisita o serviço LCA presente em <http://biodados.icb.ufmg.br/services>, protocolo SOAP).

3.3. Colorização das vias

3.3.1. Colorização das vias de acordo com o LCA dos componentes

Os componentes das vias foram coloridos de acordo com sua origem utilizando a ferramenta ARCoBALeno (GONÇALVES, 2016). Como entrada para o programa foram fornecidos o diagrama da via no formato GPML (Graphical Pathway Markup Language) gerado com o PathVisio e a lista com os nomes de cada componente (*Gene Symbol*) e seus respectivos LCAs, indicado como um número que corresponde ao nível hierárquico do táxon na linhagem evolutiva humana.

3.4. ELDOgraph

O programa ELDOgraph foi utilizado para gerar grafos que ilustram a similaridade entre os organismos envolvidos na análise com base nos dados de distância filogenética. Para isso, a primeira etapa desse processo consistiu na recuperação do proteoma completo de 92 espécies de mamíferos, cujo genoma está completamente sequenciado. Esses dados foram recuperados da base Refseq (<https://www.ncbi.nlm.nih.gov/refseq/>). Em seguida, os proteomas foram analisados pelo OrthoDB (WATERHOUSE et al., 2013) que agrupou as proteínas homólogas. Para cada agrupamento foi feito o alinhamento das sequências usando o programa Clustal Omega (SIEVERS et al., 2011; SIEVERS; HIGGINS, 2014). Os alinhamentos foram refinados com o programa trimAl (CAPELLA-GUTIÉRREZ; SILLA-MARTÍNEZ; GABALDÓN, 2009). Com esses alinhamentos foram geradas árvores gênicas com o programa FastTree (PRICE; DEHAL; ARKIN, 2009) que realiza inferências filogenéticas com uma grande quantidade de sequências, de modo similar ao método de máxima verossimilhança rapidamente. As árvores geradas servem de entrada para o programa ELDOgraph para extrair a distância par-a-par de todas as proteínas contra as proteínas de humano. O programa gera então um arquivo HTML que pode ser aberto em qualquer programa de navegação da internet. O arquivo HTML ilustra um grafo que sumariza as proporções de genes de um organismo que possui um gene de um segundo organismo como o ortólogo menos divergente. A execução dos passos é controlada por *scripts* executados sequencialmente, disponíveis em <http://biodados.icb.ufmg.br/taxonphylotools>.

4. RESULTADOS

4.1. Descrição das vias de secreção do sistema digestivo humano

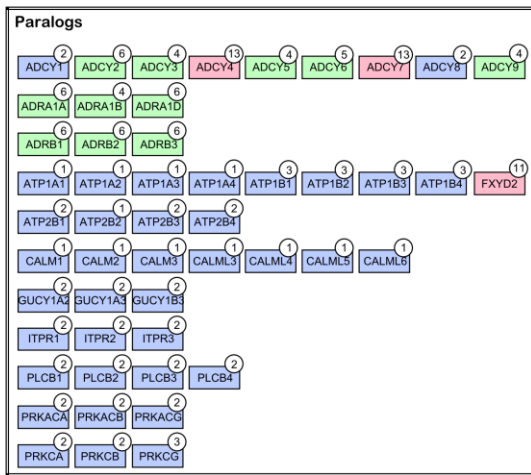
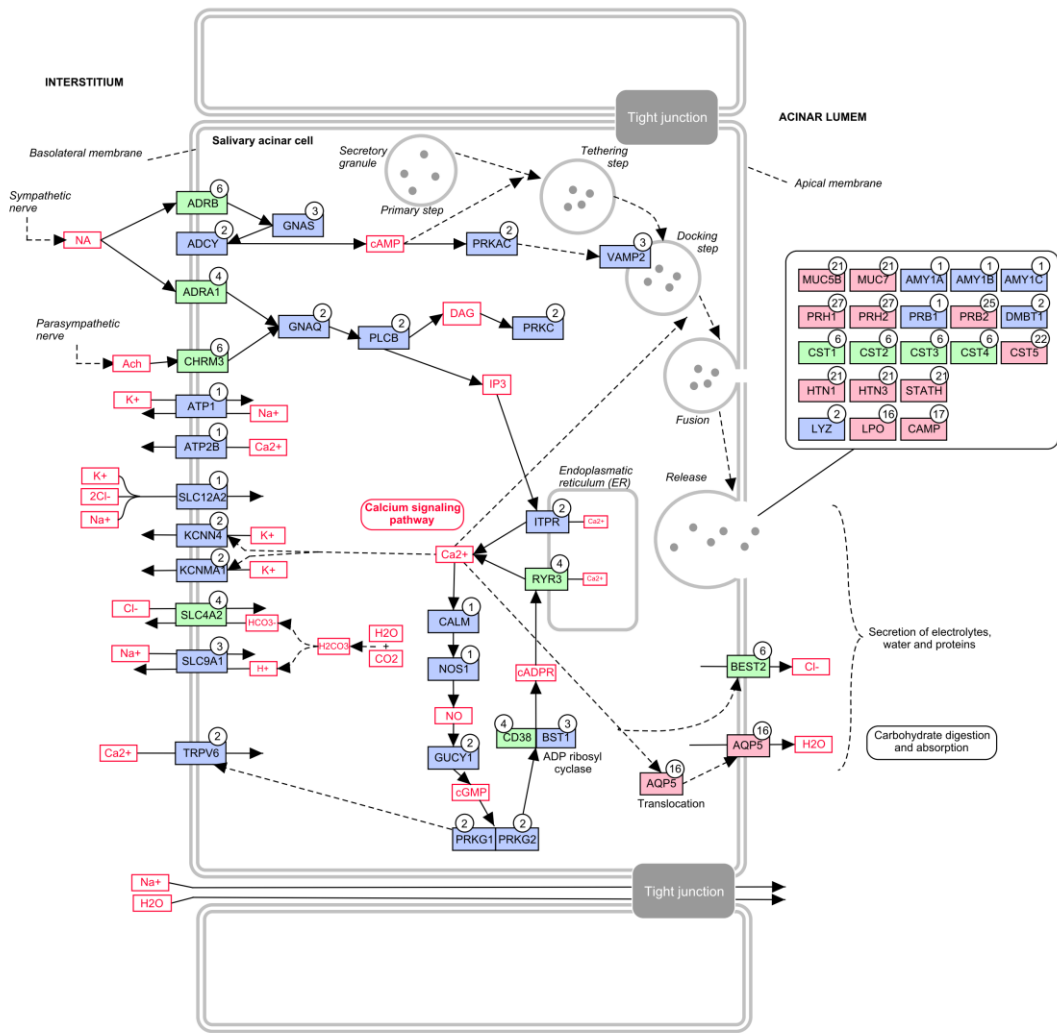
Neste trabalho analisamos as principais vias de secreção do sistema digestivo: secreção salivar, gástrica, biliar e pancreática. Conforme apresentado a seguir, foi possível distinguir três pontos na evolução da linhagem humana onde ocorre o surgimento de genes responsáveis por funções-chave no processo de secreção nas quatro vias analisadas. As vias foram obtidas da base Kegg Pathways e correspondem a via de secreção salivar (hsa04970), secreção de ácido gástrico (hsa04971), secreção pancreática (hsa04972) e secreção biliar (hsa04976), todavia uma inspeção da literatura foi necessária para compor a descrição do papel de cada produto gênico.

Os arquivos suplementares citados no texto, assim como outras figuras presentes neste documento, podem ser encontrados no endereço <http://biodados.icb.ufmg.br/heterotrophyFiles/>.

4.1.1. Descrição da via de secreção salivar

A Figura 10 mostra a via de secreção salivar extraída da base Kegg Pathways (hsa04970), sendo a sua descrição apresentada em seguida. A secreção da saliva é estimulada pelo sistema nervoso autônomo simpático e parassimpático (PROCTOR, 2016; PROCTOR; CARPENTER, 2007, 2014), responsáveis por desencadear a secreção do fluido salivar e eletrólitos, assim como do seu componente proteico. O nervo simpático estimula os receptores noradrenérgicos do tipo alfa (ADRA1) e beta (ADRB) (Figura 10, canto superior esquerdo). O receptor ADRA1 já é evidenciado em Metazoa (nível 4 da linhagem, indicado pelo número incluso no círculo) mas somente em Bilateria (n6) é que se verifica o ADRB (n6) e ambos já são de ocorrência restrita a animais (coloração verde). O estímulo dos receptores ADRB são a principal via para a secreção dos grânulos contendo as proteínas salivares, no entanto, pouco fluido é secretado (MELVIN et al., 2005). Esses receptores ativam proteínas G (GNAS) que ao se acoplarem com adenilato ciclases (ADCY) elevam os níveis de cAMP, que por sua vez ativa a proteína cinase A (PRKAC), esta última ativa a proteína VAMP2, que está ancorada à membrana do grânulo proteico. De maneira geral, a sinalização citoplasmática é mais antiga que a origem dos Metazoa (n4), aparecendo em clados coloridos em azul na Figura 10. A proteína VAMP2 é importante no processo de fusão do grânulo com a membrana plasmática, o que leva à secreção do seu conteúdo na saliva (BAUM, 1993; MESSENGER; FALKOWSKI;

GROBLEWSKI, 2014; PROCTOR, 2016; TURNER; SUGIYA, 2002). Por sua vez, os receptores noradrenérgicos do tipo alfa (ADRA1) desencadeiam uma via de sinalização de cálcio (Ca^{2+}) e respondem da mesma forma que o receptor de origem em Bilateria (n6) CHRM3: receptor muscarínico de acetilcolina (ACh). Este último é estimulado por nervos parassimpáticos e, junto com o receptor alfa (ADRA1), são responsáveis pela secreção do fluido salivar (BAUM, 1993; MELVIN et al., 2005; PAK; PHAM; ROTIN, 2002; PROCTOR; CARPENTER, 2007; TURNER; SUGIYA, 2002). Esta sinalização, embora estudada em humanos, poderia, portanto, ser efetiva já em Bilateria. Alguns genes, identificados na literatura e na figura pelo *Gene Symbol*, frequentemente têm parálogos na base UniProt, e a ancestralidade delas, quando isso ocorre, é mostrada na parte inferior da figura, sendo a ancestralidade mais remota mostrada na via. Observa-se que um parálogo do receptor do tipo alfa só ocorre em Euteleostomi (n13).



Color Legend:

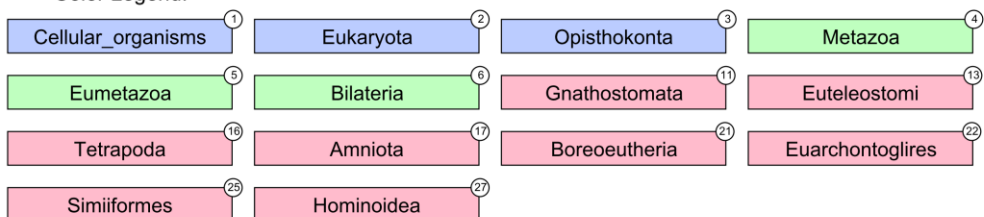


Figura 10 - Esquema da via de secreção salivar. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); verde: genes originados entre Metazoa (n4) e Chordata (n8); rosa: genes originados entre Vertebrata (n10) e *Homo sapiens* (n31). Os números dentro dos círculos correspondem ao clado no qual os genes se originaram. Fonte: adaptado do KEGG, desenhado com PathVisio.

Os receptores ADRA1 e CHRM3 estão acoplados à proteína GNAQ (MELVIN et al., 2005; MESSENGER; FALKOWSKI; GROBLEWSKI, 2014) e, quando estimulados por seus agonistas, levam à ativação da PLCB (fosfolipase $C\beta$) que, por sua vez, cliva o fosfatidilinositol 1,4, bisfosfato (PIP2) gerando diacilglicerol (DAG) e inositol 1,4,5-trifosfato (IP₃) (BAUM, 1993; CATALÁN; NAKAMOTO; MELVIN, 2009; MELVIN et al., 2005; PAK; PHAM; ROTIN, 2002; PROCTOR; CARPENTER, 2014; TURNER; SUGIYA, 2002). O DAG está envolvido na ativação de proteína cinase C (PRKC) e o IP₃ ativa seu receptor na membrana do retículo endoplasmático (RE) gerando um influxo de Ca²⁺ para o citoplasma (MELVIN et al., 2005). Os receptores ITPR e RYR3 (receptor de rianodina, este mais recente, originado em Bilateria (n6)) são os principais responsáveis pelo transporte de Ca²⁺ do RE para o citoplasma. O Ca²⁺ tem um importante papel na regulação de outras proteínas envolvidas no processo de secreção da saliva, como será discutido adiante, e na regulação do balanço de Ca²⁺ citoplasmático em um processo conhecido como liberação de cálcio induzida por cálcio (MELVIN et al., 2005; MESSENGER; FALKOWSKI; GROBLEWSKI, 2014). O complexo Ca²⁺/calmodulina estimula a produção de óxido nítrico pela óxido nítrico sintase (NOS1) que ativa a síntese de cGMP pela guanilato ciclase (GUCY1) (MELVIN et al., 2005; PROCTOR; CARPENTER, 2007). O cGMP ativa a proteína cinase G (PRKG) que fosforila uma ADP ribosil ciclase (CD38/BST1) que produz cADP-ribose (cADPR), responsável por ativar o receptor RYR3. Apenas CD38 nessa sinalização é mais recente (Metazoa – n4). Além disso, a proteína PRKG estimula a captação de Ca⁺ por proteínas transportadoras na membrana basal, incrementando o nível desse íon no citoplasma (LEHEN'KYI; PREVARSKAYA, 2011). O RYR3 é um receptor da membrana do Retículo Endoplasmático (RE) que, assim como o ITPR, é responsável por liberar o Ca²⁺ do interior do RE, de modo que a presença de Ca²⁺ no citoplasma é autorregulada pelo sistema NO-cGMP-cADP-ribose (MELVIN et al., 2005; RAH et al., 2007).

A liberação de Ca²⁺ no citoplasma é um processo chave para a liberação de proteínas e fluidos na saliva. O Ca²⁺ ativa canais de K⁺, como KCNN4 e KCNMA1, desencadeando um potencial de membrana que ativa outros transportadores de membrana

responsáveis pela translocação de íons importantes para manutenção do ambiente citosólico (LEE et al., 2012; MELVIN et al., 2005), assim como a translocação da AQP5 para a membrana apical, promovendo a secreção de H₂O (ISHIKAWA et al., 2006; ISHIKAWA; ISHIDA, 2000). A origem de AQP5 é discutível, a base Geneages aponta como antiga (encontrando-a em bactérias), mas o ortólogo mais similar ao humano é mapeado pelo SeedServer em Tetrapoda (n16). Análise detalhada com TaxOnTree sugere resultado próximo a SeedServer: Aminiota (n17) (Figura 11) já que os ramos vizinhos (não mostrados) contêm parálogos nomeados diferentemente. Quando estimuladas, as células acinares estão sob a influência de vias de sinalização de Ca²⁺ que ativam as proteínas que fazem o transporte de K⁺ para o fluido extracelular, hiperpolarizando a membrana. Esse fenômeno potencializa o transporte de Cl⁻ na membrana apical. Para restabelecer o equilíbrio eletroquímico citoplasmático, o K⁺ é reabsorvido pela célula através de transportadores como SLC12A2 e ATP1 (CANCELA et al., 2000; HEITZMANN; WARTH, 2008) originados antes dos Metazoa (n4). Esse mecanismo é comum em células acinares de diversas glândulas como saliva, célula parietal gástrica e ácinos pancreáticos, que serão descritos posteriormente. A principal proteína envolvida na entrada de Cl⁻ nas células acinares das glândulas salivares é o co-transportador de Na⁺/K⁺/2Cl⁻ (SLC12A2) (LEE et al., 2012). O Cl⁻ absorvido é liberado no lúmen da glândula salivar por canais de Cl⁻ como a proteína BEST2 encontrada em Bilateria (n6). O efluxo de Cl⁻ para o lúmen e K⁺ para o espaço intersticial, é seguido pela liberação de água para o lúmen. O desequilíbrio eletroquímico gerado pela perda de carga negativa (Cl⁻) é compensado pela entrada de Na⁺. As proteínas transportadoras de Na⁺/H⁺ e Cl⁻/HCO₃⁻, SLC9A1 e SLC4A2, respectivamente, mantêm a captação de Cl⁻ necessária para a secreção do fluido salivar. Essas proteínas são antigas surgindo em Opisthokonta (n3) e Metazoa (n4), respectivamente. Geneages as mapeia proximamente, em Cellular Organisms (n1) e Eumetazoa (n5), respectivamente, resultados relativamente próximos. Essas enzimas também desempenham uma importante função na regulação do pH citoplasmático (LEE et al., 2012; MELVIN et al., 2005). As proteínas ATP2A e ATP2B (mais antigas que os Metazoa) são transportadores de Ca⁺ que atuam removendo esse íon do citoplasma para o meio extracelular ou para o RE, respectivamente. No entanto, apesar de estudos relatarem a proteína ATP2A nas células acinares da glândula salivar (AMBUDKAR, 2000; LEE et al., 2012; MELVIN et al., 2005; PROCTOR; CARPENTER, 2014), esta não está presente na representação depositada no KEGG. Essas enzimas são importantes para recuperação do estado de repouso da célula

(BERRIDGE; BOOTMAN; RODERICK, 2003; LEE et al., 2012; MELVIN et al., 2005). O Ca²⁺ também tem um papel no processo de fusão do grânulo de zimogênio na membrana apical, pois as proteínas presentes na membrana do grânulo são dependentes de Ca²⁺ como, por exemplo, VAMP2 (MESSENGER; FALKOWSKI; GROBLEWSKI, 2014).

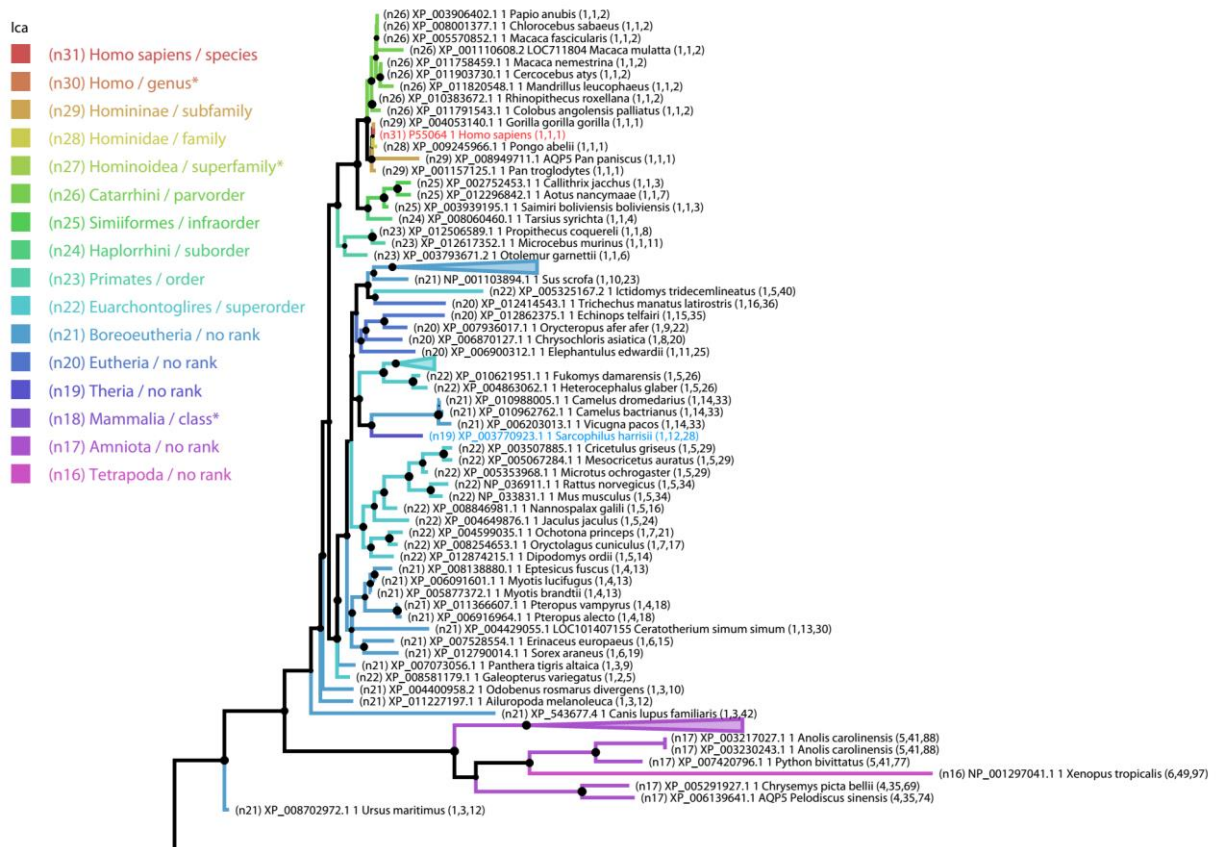


Figura 11. Análise filogenética de AQP5. Resultado de BLAST (base RefSeq sem fragmentos) e de árvore filogenética construída com TaxOnTree, mostrando a proteína AQP5 humana em vermelho e a mais distante taxonomicamente no mesmo ramo em azul. LCAs indicados na legenda e nos ramos. Ao final da notação do ramo, entre parênteses, estão elencadas as classes, ordens e famílias distintas encontradas na análise: (1,12,28) significa a 28ª família, a 12ª ordem e a mesma classe da proteína humana (1). Ramos contendo (n21) em azul, (n22) em verde e (n17) em lilás foram colapsados para clareza. AQP5 parece restrita a Amniota (n17).

A porção proteica da saliva é formada por enzimas com diferentes funções, são essas enzimas as responsáveis pelas propriedades que caracterizam a saliva. A viscosidade típica da saliva é determinada pela secreção de dois tipos de glicoproteínas, a MUC5B e MUC7, restritas a Boreoeutheria (n21) segundo SeedServer enquanto Geneages mapeia a primeira em Eumetazoa (n5) e apenas a segunda como recente, em Mammalia (n18), que inclui Boreoeutheria. Essas proteínas além de ajudar no processo de mastigação e deglutição dos alimentos, lubrificando os tecidos da mucosa oral, também

atuam como proteção contra proliferação de bactérias e fungos (HELMERHORST et al., 2008; TAKEHARA et al., 2013; XU et al., 2016). As proteínas do tipo PRH e PRB (*acid proline-rich protein* e *basic proline-rich protein*, respectivamente) evitam que sejam formados cristais de cálcio e hidroxiapatita, sendo esta última o principal componente do esmalte que recobre os dentes, no fluido salivar (BENNICK, 1982; HELMERHORST et al., 2008). A PRH mais antiga é detectada em Hominoidea (n27) pelo SeedServer, resultado que é corroborado pelo Geneages pois sua classificação mais recente é Mammalia (n18). SeedServer encontra uma PRB1 com origem antiga, mas PRB2 somente em Simiiformes (n25) (com Geneages mapeando de acordo em Mammalia). Função semelhante já foi descrita para outra proteína secretada na saliva, a Estaterina (STATH), encontrada restrita a Boroetheria (n21) pelo SeedServer. O papel principal dessa proteína é prevenir a precipitação de fosfato de cálcio e o crescimento de cristais de hidroxiapatita na superfície dos dentes (AMADO et al., 2010; LI et al., 2004). Outras proteínas encontradas na saliva atuam como uma barreira protetora da cavidade oral, com ação antimicrobiana e/ou antifúngica como as HTN1 e HTN3 (de Boroetheria – n21), lisozima (LYZ) de Bilateria (n6), lactoperoxidase (LPO) e CAMP (catelicidina) (AMADO et al., 2010; HELMERHORST et al., 2008; IDE et al., 2016; LIGTENBERG et al., 2007; PROCTOR, 2016). Há conflito para origem de LPO, Geneages encontrando ortólogos em Bacteria (n1) e SeedServer mapeando em Tetrapoda (n16), enquanto concordam para CAMP na ocorrência em Aminiota (n17) (Vertebrata para Geneages). A conferência com TaxOnTree mapeia a origem de LPO em Gnathostomata (n11), mais favorável à análise com SeedServer (Figura 12). Algumas proteínas da secreção, no entanto, são antigas. A saliva contém uma classe de proteínas conhecidas como cistatinas (CST1, CST2, CST3, CST4 e CST5) que apresentam propriedades antibacterianas e antivirais (AMADO et al., 2010). O início da família teria sido em Bilateria (n6), com outros parálogos mais recentes. Também se encontram presentes na saliva as aglutininas codificadas pelo gene DMBT1 (existentes desde os primeiros animais) que interage com patógenos, promovendo a aglutinação de bactérias e vírus (POLLEY et al., 2015). Segundo AMADO *et al.* (AMADO et al., 2010), a proteína mais abundante na saliva é a amilase (de origem bastante ancestral), compreendendo 25% do total de proteínas da saliva. Diferentes isoformas de amilase são secretadas (AMY1A, AMY1B e AMY1C), principalmente nas glândulas parótidas, sendo glicosiladas como não-glicosiladas (HELMERHORST; OPPENHEIM, 2007). As células na glândula salivar, assim como nas glândulas gástricas e pancreáticas, estão ligadas às suas vizinhas por junções

celulares. Estas estruturas não permitem a passagem de compostos grandes, no entanto, água e cátions têm livre passagem no espaço entre as células em resposta à diferença no potencial elétrico durante o estímulo (MELVIN et al., 2005). Como se vê, apesar de conter proteínas de origem remota, a secreção salivar contém produtos bastante recentes na evolução, os mais recentes da via.

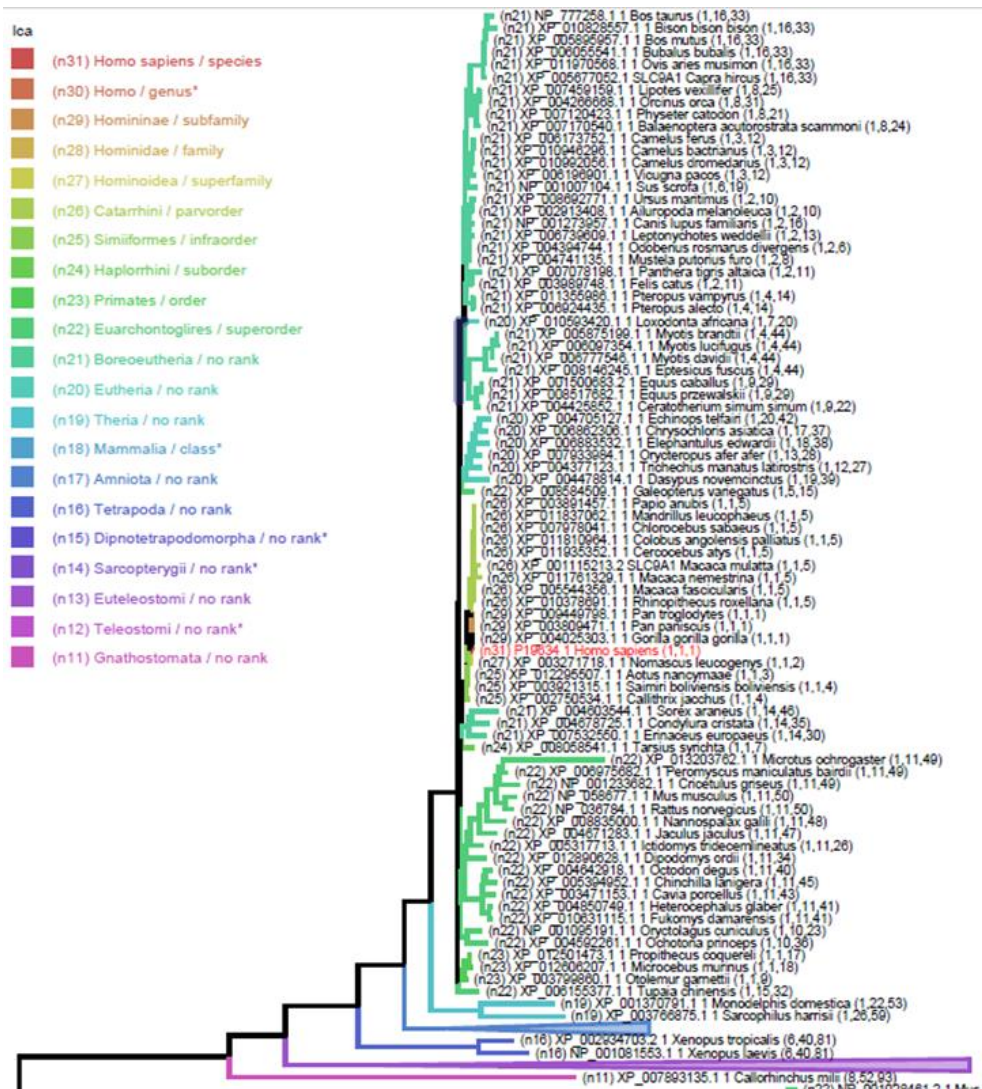


Figura 12. Análise filogenética de LPO. Resultado de BLAST (base RefSeq sem fragmentos) e de árvore filogenética construída com TaxOnTree, mostrando a proteína LPO humana em vermelho. Notações similares às da figura anterior. Ramos contendo (n11) em lilás e (n21) em azul foram colapsados para clareza. LPO parece restrita a Gnathostomata (n11).

4.1.2. Secreção gástrica

O estímulo para secreção do ácido gástrico é uma ação coordenada de sinais nervosos, hormonais e parácrinos (SCHUBERT, 2010, 2014, 2016; SCHUBERT; PEURA, 2008; TAKEUCHI et al., 2016). Esses sinais podem ser disparados em

diferentes etapas do processo de digestão: quando o alimento está sendo ingerido e durante a sua passagem pelo sistema digestivo, e mesmo antes dele, pelo estímulo dos sentidos (visão, olfato e paladar). A acetilcolina (ACh), gastrina (GAST) e histamina são os principais ativadores da secreção gástrica (Figura 13). Essas moléculas são reconhecidas por receptores específicos na membrana basolateral das células parietais, CHRM3, CCKBR e HRH2, respectivamente (FORTE; ZHU, 2010; KOPIC; MUREK; GEIBEL, 2010; SCHUBERT, 2014; YAO; FORTE, 2003). Além da gastrina, o receptor CCKBR também reconhece a colecistocinina (SCHUBERT, 2010; TRIPATHI et al., 2015). Tanto o CHRM3, quanto o CCKBR surgem em Bilatéria (n6), enquanto o receptor HRH2 já é encontrado em Metazoa (n4) (Geneages mapeando próximo, em Eumetazoa – n5). O estímulo do nervo vago, na forma de acetilcolina, e da gastrina (que surge em Theria (n19), com concordância de Geneages que mapeia em Mammalia – n18), ativam a PLCB (fosfolipase C β), por intermédio de proteínas GNAQ, que gera DAG e IP3 a partir do fosfatidilinositol 1,4, bisfosfato (PIP2). O DAG estimula a PRKC que por sua vez fosforila a proteína Ezrin (EZR). As proteínas que atuam nesse processo têm uma origem antiga, no entanto, a origem da proteína EZR é controversa. Enquanto as análises do SeedServer apontam sua origem em Opisthokonta (n3), o GeneAges indica que esta surge em Vertebrata (n10). Porém, a conferência com TaxOnTree (não mostrado) mostra que o resultado do SeedServer está mais apropriado. A EZR é uma proteína encontrada na membrana canalicular apical das células parietais; ela associa-se com actina do citoesqueleto promovendo a formação de microvilosidades, tanto no estômago como no intestino, e de tubovesículas (SCHUBERT, 2016). Nas células parietais essa proteína está associada à fusão das vesículas contendo HCl à membrana canalicular culminando na secreção do seu conteúdo no lúmen da glândula oxíntica. O IP3 ativa o efluxo de Ca²⁺ do retículo endoplasmático, acionando uma cascata de sinais, intermediada pela calmodulina (CALM) e proteínas dependentes de Ca²⁺ (CAMK2 e MYLK), que resulta na fusão das tubulovesículas na membrana apical e na ativação da bomba de prótons (ATP4A/ATP4B) encontrada na membrana dessas vesículas (HERRING et al., 2006; SCHUBERT, 2016). A gastrina também estimula a liberação de histamina pelas células ECL. Ao ligar-se ao seu receptor na membrana das células parietais (HRH2), a histamina ativa a proteína ADCY, por intermédio de proteínas GNAS, gerando cAMP. Este último aciona diversas proteínas, entre elas a PRKAC que fosforila a proteína EZR, cuja função já foi mencionada anteriormente, culminando na secreção do ácido gástrico (PANULA et al., 2015). Os componentes que participam dessa via de sinalização surgem nos clados mais

ancestrais, Cellular organisms (n1), Eukaryota (n2) e Opisthokonta (n3). A exceção são os receptores HRH2, cuja ancestralidade remonta aos primeiros animais, e a proteína ADCY que tem seu mais antigo parálogo originado em Eukaryota (n2), mas outros surgiram em linhagens mais recentes, como Euteleostomi (n13). O heterodímero formado pelas proteínas ATP4A e ATP4B tem origens diferentes, sendo a primeira mais antiga (Metazoa – n4) que a segunda tendo surgido em Bilateria (n6). A somatostatina (SST), através de seu receptor SSTR2, age como inibidor da secreção do HCl por meio da proteína GNAI, que inibe a proteína ADCY e conseqüentemente a produção de cAMP (TAKEUCHI et al., 2016). A SST é recente, surge em Gnathostomata (n11) (Vertebrata (n10) para Geneages, em acordo), depois do seu receptor (Bilateria – n6).

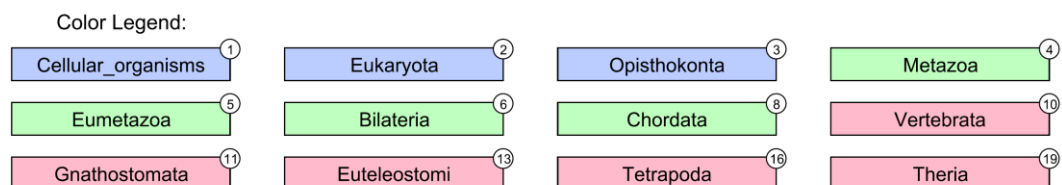
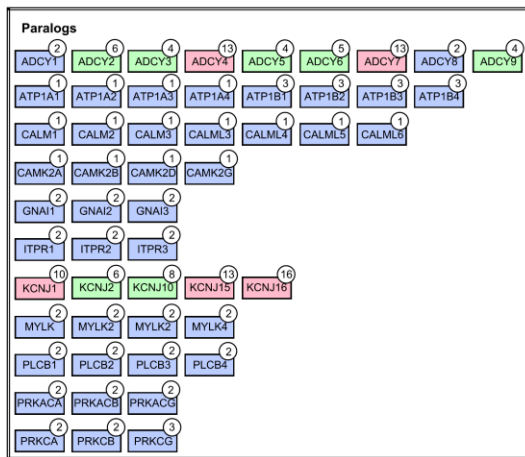
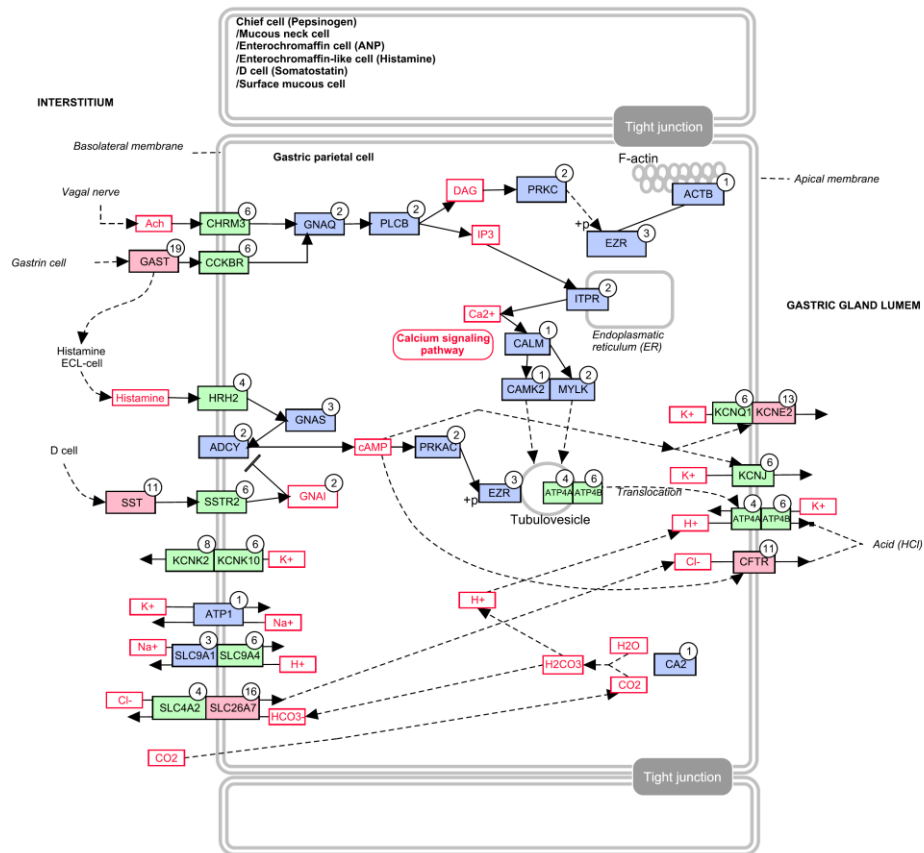


Figura 13 - Esquema da via de secreção gástrica representando uma célula parietal. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); verde: genes originados entre Metazoa (n4) e Chordata (n8); rosa: genes originados entre Vertebrata (n10) e *Homo sapiens* (n31). Os números dentro dos círculos correspondem à linhagem na qual os genes se originaram. Fonte: Adaptado do KEGG.

A formação de HCl ocorre principalmente pela ação da bomba de H^+K^+ -ATPase composta por duas subunidades ATP4A e ATP4B. O processo tem início no citoplasma onde a anidrase carbônica (CA2 - *carbonic anhydrase 2*) produz bicarbonato (HCO_3^-) e H^+ . Na membrana basolateral a proteínas do tipo AE (*anion exchanger*), SLC4A2 e SLC26A7, fazem o transporte de HCO_3^- para o fluido extracelular e a captação um íon cloreto (Cl^-) para o citoplasma. Como mencionado anteriormente a proteína SLC4A2 tem origem antiga, mas seu parálogo, SLC26A7, tem origem indicada por SeedServer em Tetrapoda (n16), mesmo clado apontado por TaxOnTree (não mostrado), embora Geneages conflite indicando a origem em Eumetazoa (n5). Posteriormente, os íons Cl^- são secretados para o lúmen da glândula por meio da proteína CFTR, sendo a ação desta regulada pela ação do cAMP (KOPIC; MUREK; GEIBEL, 2010; SCHUBERT, 2014). Essa proteína originada em Gnathostomata (n11) (Vertebrata (n10) para Geneages, de acordo), desempenha um importante papel na secreção do HCl juntamente com a H^+K^+ -ATPase. O bombeamento de H^+ para o lúmen da glândula é realizado pela H^+K^+ -ATPase em troca de um íon K^+ . Na membrana basolateral, as proteínas KCNK10 e KCNK2, ATP1 e SLC9A1 e SLC9A4 promovem um influxo de Na^+ e de K^+ do citoplasma, a redução do nível de Na^+ e K^+ resulta na reabsorção de K^+ na membrana apical, a qual é a essencial para o funcionamento da bomba H^+K^+ -ATPase (FORTE; ZHU, 2010; KOPIC; MUREK; GEIBEL, 2010). Os parálogos KCNK10 e KCNK2 surgem em Bilateria (n6) e Chordata (n8), respectivamente. Já a proteína ATP1 tem um grupo de parálogos que surge em Cellular organisms (n1) e outro em Opisthokonta (n3). Na célula parietal são expressos dois parálogos do transportador do tipo NHE (*Na⁺/H⁺ exchanger*), SLC9A1 e SLC9A4, originados aproximadamente na mesma época, sendo o SLC9A1 em Opisthokonta (n3) e SLC9A4 em Eumetazoa (n5) (Geneages detecta homólogos em bactérias). As proteínas KCNQ1, KCNE2 e KCNJ, são responsáveis por transportar o K^+ para o lúmen da glândula, restaurando o equilíbrio iônico da célula (FORTE; ZHU, 2010; KOPIC; MUREK; GEIBEL, 2010; SCHUBERT, 2014). Destas, KCNE2 seria uma adição recente em Euteleostomi (n13) (Vertebrata (n10) para Geneages, em acordo).

4.1.3. Secreção Pancreática

Quando recebe estímulo do sistema nervoso ou hormonal, a célula acinar desencadeia uma cascata de sinalizações que levam à secreção de proteínas, sais ($NaCl$) e água no duodeno (KONTUREK et al., 2003; WILLIAMS, 2010). A Figura 14 mostra a via de secreção pancreática, sendo a sua descrição apresentada em seguida.

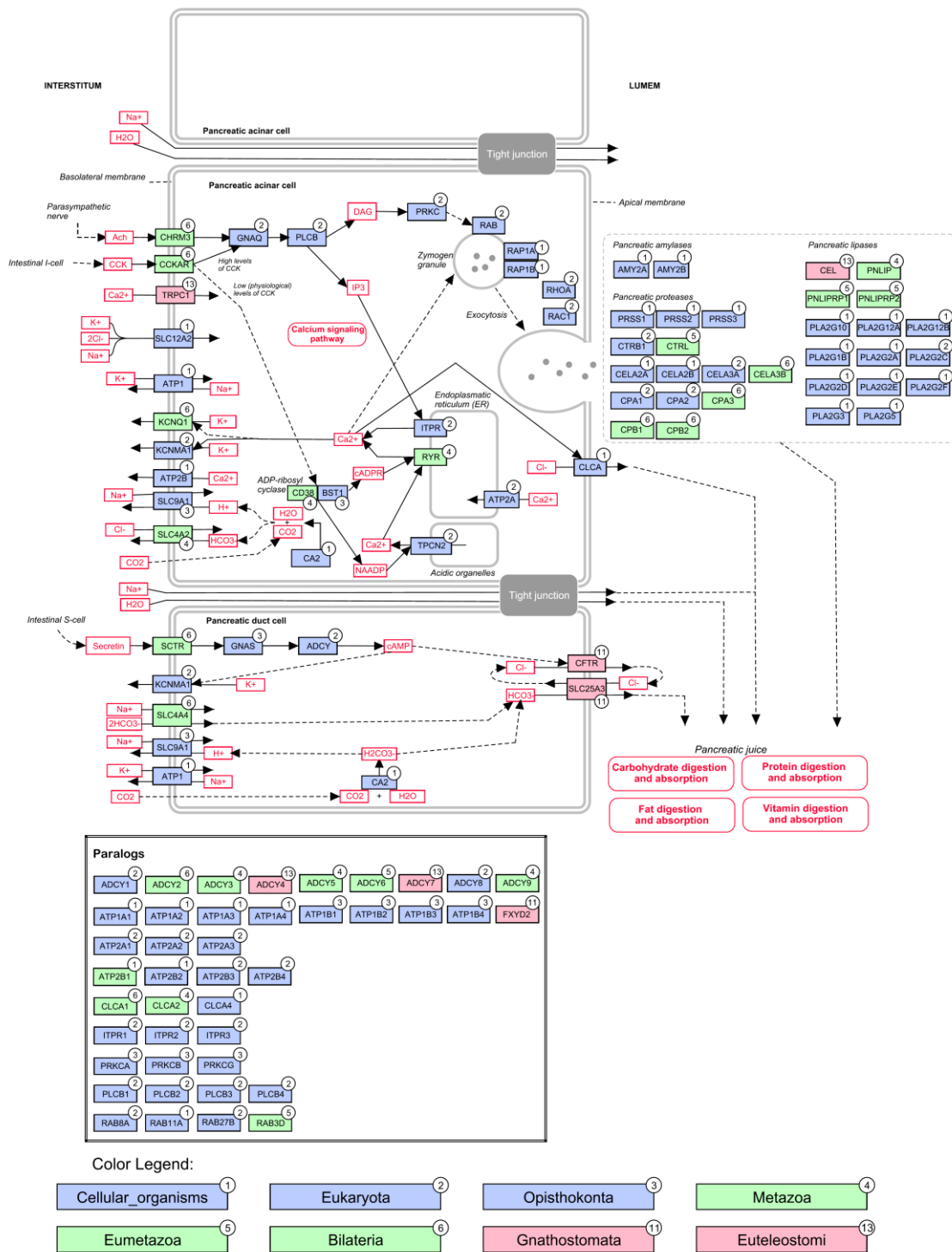


Figura 14 - Esquema da via de secreção pancreática. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); Verde: genes originados entre Metazoa (n4) e Chordata (n8); Rosa: genes originados entre Vertebrata (n10) e *Homo sapiens* (n31). Os números dentro dos círculos correspondem à linhagem na qual os genes se originaram. Fonte: Adaptado do KEGG.

A secreção das proteínas pancreáticas é estimulada pelo sistema nervoso parassimpático por meio da acetilcolina (Ach) e também pela Colecistocinina (CCK) secretada pela mucosa duodenal (Figura 14). Ambos os compostos mencionados anteriormente estimulam os ácinos do pâncreas ao se ligarem a seus respectivos receptores, CHRM3 e CCKAR, ambos originados em Bilateria (n6) (com concordância de Geneages). Esses receptores desencadeiam a liberação de Ca^{2+} intracelular em um mecanismo similar ao descrito para a célula acinar salivar. Por intermédio da proteína GNAQ, a proteína PLCB é ativada, produzindo DAG e IP3. O DAG, no ácino pancreático, ativa um conjunto de proteínas responsáveis pela ancoragem e fusão do grânulo de zimogênio à membrana apical do ácino (RAB, RAP1A, RAP1B, RHOA e RAC1) (MESSENGER; FALKOWSKI; GROBLEWSKI, 2014; NAJI; PACHOLSKY; ASPENSTRÖM, 2011; WILLIAMS, 2006; YAO; FORTE, 2003). A origem dessas últimas proteínas remonta aos clados Cellular organisms (n1) e Eukaryota (n4). O IP3 tem um papel direto na ativação da via de sinalização de Ca^{2+} pois estimula o transporte deste do retículo endoplasmático (RE) para o citoplasma (BI; WILLIAMS, 2004; KANNO, 1998; KONTUREK et al., 2003; WILLIAMS, 2010). As proteínas CD38 (Metazoa – n4) e BST1 (Bilateria – n6) são ADP-ribosil ciclases que, além de produzirem cADPR, geram NAADP, que ativam, respectivamente, o receptor RYR (*Ryanodine Receptor 2*) (Bilateria – n6) na membrana do retículo endoplasmático e o receptor TPCN2 (Eukaryota – n2) nos lisossomos, resultando na liberação de Ca^{2+} dessas organelas para o citoplasma (CANCELA et al., 2000; LIU et al., 2005). Nesse sistema de sinalização não há conflito com Geneages. Parte do Ca^{2+} presente no citoplasma é bombeada para o fluido intercelular pela proteína ATP2B (*plasma-membrane- Ca^{2+} -activated-ATPase*) (PETERSEN; TEPIKIN, 2008). Para restaurar as reservas desse íon nos compartimentos de armazenamento, como RE, o canal de cálcio TRPC1 (*Short transient receptor potential channel 1*) transporta o Ca^{2+} do fluido intersticial para o citoplasma (AMBUDKAR, 2000; BERRIDGE; BOOTMAN; RODERICK, 2003; MELVIN et al., 2005; PETERSEN; TEPIKIN, 2008). As proteínas transportadoras de K^+ encontradas na membrana basolateral, como a KCNQ1 e KCNMA1, regulam a voltagem da membrana (CANCELA et al., 2000; HEITZMANN; WARTH, 2008). Nesse sistema, as proteínas mais recentes são KCNQ1 (Bilateria (n6), com acordo de Geneages) e TRPC1 (bem mais recente, de Euteleostomi – n13), embora segundo Geneages esta última seria remota (Opisthokonta – n3). No entanto, a conferência com TaxOnTree sugere que TRPC1 teria surgido em Gnathostomata (n11), próximo de Euteleostomi (n13). A proteína KCNMA1

é mais antiga (Eukaryota (n2), Geneages corroborando). No citoplasma, a anidrase carbônica produz H^+ e HCO_3^- que são trocados por Na^+ e Cl^- , pelas proteínas SLC9A1 (Opisthokonta – n3) e SLC4A2 (Metazoa – n4), respectivamente. A proteína ATP1 transporta K^+ para o citoplasma em troca de Na^+ . Além desses canais de entrada de Na^+ e Cl^- , a proteína SLC12A2 (Cellular organisms – n1) faz o transporte de Na^+ , $2Cl^-$ e K^+ para o citoplasma (HEITZMANN; WARTH, 2008). Na membrana apical, a proteína CLCA (n1) transporta o Cl^- para o lúmen. O Na^+ e H_2O atravessam o espaço intersticial para o lúmen, onde são formados os sais de NaCl. O mecanismo descrito gera um potencial de membrana no ácino pancreático, tornando a face citoplasmática negativa e a externa positiva, fenômeno importante para ativação de canais e bombas dependentes de voltagem (HEITZMANN; WARTH, 2008). As enzimas secretadas pelos ácinos são responsáveis pela digestão dos principais grupos de nutrientes (isto é, carboidratos, proteínas e lipídeos). As principais responsáveis pela digestão de carboidratos são as amilases. Estas pertencem a uma classe diferente das amilases expressas na saliva, mas assim como as salivares, as proteínas AMY2A e AMY2B surgem em Cellular organisms (n1). Dentre as proteases, são encontradas serino proteases (Tripsina, Quimotripsina e Elastase) e carboxipeptidases. O pâncreas secreta três isoformas de Tripsinogênio, PRSS1, PRSS2 e PRSS3, todos datados de Cellular organisms (n1), embora Geneages sugira origem de PRSS1 em Opisthokonta (n3) e PRSS3 em Eukaryota (n2). Uma vez secretado, o tripsinogênio sofre uma proteólise tornando-o ativo, forma conhecida como tripsina. A tripsina é um importante componente do suco pancreático, pois além de hidrolisar polipeptídeos junto a resíduos de lisina e arginina essas endopeptidases regulam outras enzimas digestivas (WHITCOMB; LOWE, 2007). As quimotripsinas, CTRB1 (*Chymotrypsinogen B*) e CTRL (*Chymotrypsin-like*), são responsáveis por hidrolisar polipeptídeos junto a aminoácidos aromáticos. Estas endopeptidases que surgem em Eukaryota (n2) e Eumetazoa (n5) respectivamente (ambas em Eumetazoa para Geneages), são a segunda mais importante classe de proteínas secretadas pelo pâncreas (SZABÓ et al., 2016; WHITCOMB; LOWE, 2007). As elastases são proteínas capazes de digerir elastina rompendo as ligações peptídicas nos resíduos de alanina, glicina e serina. Quatro elastases são encontradas no suco pancreático CELA2A (Chymotrypsin-like elastase 2A), CELA2B, CELA3A e CELA3B (WHITCOMB; LOWE, 2007). As CELA2 surgem em Cellular organisms (n1), enquanto CELA3A em Eukaryota (n2) e CELA3B em Bilateria (n6). Para Geneages, todas de Eumetazoa (logo não há conflito para a origem de CELA3B em Bilateria). As carboxipeptidases são exopeptidases capazes

de clivar cadeias peptídicas c-terminais. As CPAs quebram as ligações carboxiterminais nos resíduos de tirosina, triptofano e fenilalanina, já as CPBs clivam ligações nos resíduos de arginina e lisina. As CPA1 e CPA2 são as primeiras a surgirem em Eukaryota (n2), enquanto as demais (CPA3, CPB1 e CPB2) surgem em Bilateria (n6). Desta forma, o repertório de proteases é em geral antigo, mas algumas adições aparecem somente nos Bilateria, completando o repertório que é encontrado em humanos.

No suco pancreático são encontradas diversas lipases que digerem os ácidos graxos. A principal lipase pancreática é a carboxil esterase PNLIP (*Pancreatic triacylglycerol lipase*). Esta enzima cuja origem remonta a linhagem dos Eumetazoa (n5) (com suporte de Geneages), é responsável por hidrolisar acilglicerídeos (VAN BENNEKUM et al., 2000; WHITCOMB; LOWE, 2007). Os parálogos PNLIPRP1 e PNLIPRP2 conservam cerca de 65% de identidade com a PNLIP, e assim como esta surgem em Eumetazoa (n5) (novamente com suporte de Geneages). Porém, a PNLIPRP1 não possui ação lipolítica, atuando como inibidor da digestão de lipídeos devido a duas substituições no seu sítio catalítico (BERTON et al., 2009; EYDOUX et al., 2008; VAN BENNEKUM et al., 2000). Por outro lado, a PNLIPRP2 age sobre uma ampla variedade de substratos, atuando como lipase, fosfolipase e galactolipase (EYDOUX et al., 2008; VAN BENNEKUM et al., 2000). A CEL (*carboxyl ester lipase*) é uma lipase com ampla variedade de substratos e que atua em diversos processos como triacilglicerídeos, ésteres de colesterol, fosfolipídeos entre outros. Esse produto gênico é expresso no pâncreas, glândulas mamárias, fígado, glândula pituitária entre outros (HOLMES; COX, 2011; HUI; HOWLES, 2002; WHITCOMB; LOWE, 2007). Estudos indicam ação dessa proteína não apenas no metabolismo de lipídeos, mas também na proliferação de células endoteliais, regulação da secreção de hormônios e até mesmo como potencial agente contra aterosclerose (HUI; HOWLES, 2002). Esta enzima tem origem em Euteleostomi (n13) (com suporte de Geneages que a mapeia em Vertebrata (n10)), sendo então a mais recente enzima secretada pelo pâncreas. Outra família proteica muito abundante no suco pancreático é a Fosfolipase A2 (PLA2). Essas proteínas são responsáveis por hidrolisar ligações ester em fosfolipídeos (WHITCOMB; LOWE, 2007). Essa família de proteínas tem origem ancestral que remonta a Cellular organisms (n1) com um parálogo que surgiu em Eukaryota (n2), enquanto Geneages as mapeia proximalmente, todas em Eumetazoa (n5).

As células que compõem o ducto da glândula pancreática têm uma importante função na regulação do produto final secretado. As células do ducto são estimuladas pelo

hormônio Secretina, produzido no intestino. Quando o receptor de secretina (SCTR) é estimulado, ativa a ADCY que aumenta os níveis de cAMP, este aciona o transportador de Cl^- , o CFTR, na membrana apical. O HCO_3^- produzido no citoplasma é secretado no lúmen da glândula pela proteína SLC25A3 que faz o transporte de um Cl^- para cada HCO_3^- secretado. Assim, as proteínas CFTR e SLC25A3, que são os componentes mais recentes desse sistema (Gnathostomata - n11), atuam em conjunto para secretar bicarbonato no lúmen, produzindo o suco pancreático final que flui para o duto pancreático e é então lançado no duodeno (CHU; YUNG; CHOW, 2006; HEITZMANN; WARTH, 2008; KIM; STEWARD, 2009). O receptor de secretina (SCTR) é um caso peculiar de conflito pois SeedServer sugere origem em Bilateria (n6) mas verificação com TaxOnTree mapeia em Gnathostomata (n11), de acordo com a inferência de Geneages, Vertebrata (n10). A entrada de HCO_3^- é intermediada pela proteína SLC4A4, também datado de Bilateria (n6) por SeedServer, mas assim como a SCTR, análise com TaxOnTree arbitra o conflito concordando com Geneages (Vertebrata – n10), mapeando-o em Gnathostomata (n11). A concentração de bicarbonato no suco pancreático pode ser cinco vezes maior que no plasma, e age como uma barreira protetora que ajuda a neutralizar a acidez do ácido clorídrico vindo do estômago, evitando que este agrida o tecido duodenal (ALLEN; FLEMSTRÖM, 2005; HEITZMANN; WARTH, 2008; KIM; STEWARD, 2009). Nota-se que essas últimas atividades dependem de muitas proteínas presentes somente em Gnathostomata (n11).

4.1.4. Secreção Biliar

A via de secreção biliar é mostrada na Figura 15 e descrita a seguir.

Figura 15 - Esquema da via de secreção biliar. Abaixo da via estão listadas as demais isoformas das proteínas presentes na via. Azul: genes originados entre Cellular organisms (n1) e Opisthokonta (n3); Verde: genes originados entre Metazoa (n4) e Chordata (n8); Rosa: genes originados entre Vertebrata (n10) e *Homo sapiens*. Os números dentro dos círculos correspondem à linhagem na qual os genes se originaram. Fonte: Adaptado do KEGG.

Os ácidos biliares (BA) são compostos esteroidais provenientes do metabolismo do colesterol como ácido quenodesoxicólico dihidroxilado e ácido cólico trihidroxilado. Os receptores de colesterol LDLR e SCARB1, que surgiram em Gnathostomata (n11) e Eukaryota (n2), respectivamente, permitem a entrada de colesterol nas células hepáticas por meio de endocitose (KISS; SNIDERMAN, 2017; ZELCER et al., 2009). Geneages conflita indicando origem de LDLR em Eumetazoa (n5), mas TaxOnTree suporta Gnathostomata. Uma vez no citoplasma, o colesterol sofre modificações pelas proteínas HMGCR (*hydroxymethylglutaryl-CoA reductase*) e NCEH1 (*cholesterol ester hydrolase*), ambas surgidas em Cellular organisms (n1). A HMGCR é uma proteína essencial na via de síntese de colesterol, de modo que essa proteína é um fator limitante no processamento de BA (HOWE et al., 2017; SHARPE; BROWN, 2013; ZHONG et al., 2014). A NCEH1 hidrolisa o colesterol no citoplasma permitindo que este seja secretado (GHOSH, 2012; ZHAO; NATARAJAN; GHOSH, 2005; ZHAO; SONG; GHOSH, 2008). Outra fonte de BA nas células é através da reabsorção desses ácidos pelas proteínas presentes na membrana basolateral dos hepatócitos: duas dessas proteínas de origem ancestral (Cellular organisms – n1) são SLC10A1 (*sodium taurocholate-cotransporting polypeptide*) e EPHX1 (*microsomal epoxide hydrolase*) que transportam BA conjugados com Na⁺. Além delas, também estão presentes a SLCO1 (*organic anion-transporting polypeptides*), que faz a troca de BA por HCO₃⁻ e glutatona (GSH), e as SLC22As (*organic anion transporters*), que além de BA também transportam ânions orgânicos (OA) (ALREFAI; GILL, 2007; BOYER, 2013; ESTELLER, 2008; MEIER; STIEGER, 2002; THOMAS et al., 2008; TRAUNER; BOYER, 2003). Além da captação de ácidos biliares e ânions orgânicos, a membrana basolateral também tem proteínas que fazem a captação de cátions, como a SLC22A1 (*organic cation transporters*) e transporte de potássio, como a KCNN2 e ATP1, cuja função na regulação do gradiente eletroquímico nas células já foi comentada anteriormente (BOYER, 2013; ESTELLER, 2008). Dentre essas proteínas as mais recentes são as SLCO1s e SLC22As que têm origem do grupo em Bilateria (n6), apesar de parálogos dessas proteínas surgirem ao longo da linhagem até Gnathostomata (n11), dados confirmados com TaxOnTree. No citoplasma, uma proteína

que surge em Eukaryota (n2), CYP7A1 (*cholesterol-7 α -hydroxylase*) faz a síntese de BA a partir do colesterol (BOYER, 2013; ESTELLER, 2008; THOMAS et al., 2008; TRAUNER; BOYER, 2003). Os BA são então conjugados com taurina e glicina, tornando-os mais hidrofílicos e facilitando sua excreção. Muitas proteínas podem realizar a conjugação dos BA, entre elas UGT2B4 (*UDP-glucuronosyltransferase 2B4*), a mais ancestral delas (Cellular organisms – n1), SULT2A1 (*bile salt sulfotransferase*), BAAT (*bile acid-CoA:amino acid N-acyltransferase*), ambas originadas após Opisthokonta (n3), e SLC27A5 (*bile acyl-coa synthetase - BACS*), proteína que tem origem em Cellular organisms (n1) mas gera parálogos em Euteleostomi (n13), segundo TaxOnTree (BOYER, 2013; ESTELLER, 2008; MEIER; STIEGER, 2002). Além disso, os BA também podem sofrer destoxificação pela proteína CYP3A4 (*cytochrome P450 3A4*), que surgiu em Metazoa (n4), antes de serem secretados (TRAUNER et al., 2005).

Após a síntese e modificação, os BA são secretados nos canalículos. As proteínas ABCB11 (*bile salt export pump*) e ABCC2 (*multidrug resistant associated protein 2*) são as principais vias de saída dos BA para os canalículos. Segundo GeneAges, ambas as proteínas surgiram em clados basais, Cellular organisms (n1) e Eukaryota (n2), no entanto, SeedServer conflita colocando a origem do ABCB11 em Chordata (n8), com suporte de TaxOnTree. Além dos BA, esteróis e colesterol também são transportados para os canalículos respectivamente pelas proteínas ABCG2 (*breast cancer resistance protein*) e o heterodímero ABCG5/G8 (BOYER, 2013; ESTELLER, 2008), todos originados em Eukaryota (n2). As proteínas ABCB1 e ABCB4 (*multidrug resistance proteins*) translocam respectivamente cátions e fosfolipídios (PL) para os canalículos (BOYER, 2013; ESTELLER, 2008). Nossos resultados indicam que o transportador de cátions surgiu primeiro, em Eukaryota (n2), enquanto a origem da proteína MDR3 ocorreu em Metazoa (n4), todavia sendo remotos. Além das proteínas citadas, a SLC4A2 (Metazoa – n4) também está presente na membrana apical do canalículo. Essa proteína ajuda a regular o pH do citoplasma secretando Cl⁻ e captando HCO₃⁻. A ação da SLC4A2 depende dos níveis de bicarbonato que pode ser produzido pela anidrase carbônica (CA2), presente no citoplasma, ou bombeado para o citoplasma pela proteína SLC4A5 (*sodium bicarbonate cotransporter*), esta última tendo surgido mais recentemente em Euteleostomi (n13) (com suporte de GeneAges) e que está presente na membrana basolateral (ESTELLER, 2008). Os BA são transportados para o canalículo, como anteriormente descrito, e também para o sangue através de proteínas presentes na membrana basolateral como as ABCC3 e ABCC4 (*multidrug-resistance-associated protein*), cuja origem ocorreu em Eukariota

(n2), e o heterodímero SLC51A/SLC51B (*organic solute transporter*) (ALREFAI; GILL, 2007; BOYER, 2013; ESTELLER, 2008; THOMAS et al., 2008). O primeiro componente desse heterodímero a surgir foi o SLC51A, em Bilateria (n6), já a origem de SLC51B interessantemente ocorreu em Euteleostomi (n13) (Geneages inclusive mapeia sua origem em Mammalia (n18) e TaxOnTree em Theria (n19), tendo um viés para recente). Essas proteínas são um mecanismo para evitar o fenômeno da colestase, ou retenção de BA nos hepatócitos (MAILLETTE DE BUY WENNIGER; BEUERS, 2010). A expressão de muitas proteínas responsáveis pelo fluxo dos BA é regulada por um conjunto de três proteínas: a RXRA (*retinoid-X nuclear receptor*), que surgiu em Opisthokonta (n3), NR1H4 (*farnesoid X-receptor*), cuja origem ocorreu em Bilateria (n6), e NR0B2 (*short heterodimer protein*), o mais recente desses reguladores que aparece em Euteleostomi (n13), com suporte de TaxOnTree (BOYER, 2013; ESTELLER, 2008; TRAUNER; BOYER, 2003).

No duto biliar os BA são processados pelos colangiócitos e assim como as células que compõe o duto da glândula pancreática, os colangiócitos são estimulados pela secretina e respondem a esse estímulo de forma muito similar às células do duto pancreático, resultando na secreção de HCO_3^- (BOYER, 2013; ESTELLER, 2008). Os componentes que atuam nesse processo nos colangiócitos são os mesmos encontrados nas células do duto pancreático, por isso esse processo não será descrito novamente.

Na membrana luminal dessas células encontram-se proteínas como SLC10A2 (*apical sodium-dependent bile salt transporter - ASBT*) e SLC5A1 (*Sodium/glucose cotransporter 1*), ambas originadas em Cellular organisms (n1)', que reabsorvem os BA e glicose, respectivamente. Os colangiócitos expressam alguns transportadores que ficam ancorados na membrana basolateral e secretam compostos na corrente sanguínea, entre eles encontram-se ASBT, MRP3 e MRP4, e SLC51A/OST- β , que com exceção do heterodímero SLC51A/OST- β , surgem em linhagens basais (Cellular organisms (n1) e Eukaryota (n2)), e são responsáveis por reciclar os BA que voltam para as células hepáticas ou seguem para o sistema circulatório. Esse processo ocorre normalmente na vesícula biliar mesmo em períodos de jejum (BOYER, 2013). A proteína SLC2A1 (Eukaryota - n2) transporta a glicose reabsorvida para o sangue. Além das proteínas citadas, aquaporinas são responsáveis pelo fluxo de H_2O tanto nos hepatócitos (AQP9 e AQP8) quanto nos colangiócitos (AQP4 e AQP1) (BOYER, 2013; ESTELLER, 2008; MAILLETTE DE BUY WENNIGER; BEUERS, 2010). Esses canais surgem todos em

Cellular organisms (n1). O suco biliar produzido nesse processo segue para o intestino ou pode ser armazenado na vesícula biliar.

4.2. Comparando as vias de secreção do sistema digestivo

As quatro vias apresentadas compartilham um conjunto de 20 genes, sendo que 17 deles pertencem a duas famílias de proteínas. Essas proteínas desempenham funções nas vias de sinalização interna que resultam na secreção dos compostos digestivos. A lista desses genes pode ser encontrada na Tabela 1.

Tabela 1 – Proteínas compartilhados pelas vias de secreção salivar, gástrica, pancreática e biliar

Gene symbol	Descrição da proteína	Uniprot ID
ADCY1	Adenylate cyclase 1	Q08828
ADCY2	Adenylate cyclase 2	Q08462
ADCY3	Adenylate cyclase 3	O60266
ADCY4	Adenylate cyclase 4	Q8NFM4
ADCY5	Adenylate cyclase 5	O95622
ADCY6	Adenylate cyclase 6	O43306
ADCY7	Adenylate cyclase 7	P51828
ADCY8	Adenylate cyclase 8	P40145
ADCY9	Adenylate cyclase 9	O60503
SLC4A2	Anion exchange 2	P04920
ATP1A1	Sodium/potassium-transporting ATPase subunit alpha 1	P05023
ATP1A2	Sodium/potassium-transporting ATPase subunit alpha 2	P50993
ATP1A3	Sodium/potassium-transporting ATPase subunit alpha 3	P13637
ATP1A4	Sodium/potassium-transporting ATPase subunit alpha 4	Q13733
ATP1B1	Sodium/potassium-transporting ATPase subunit beta 1	P05026
ATP1B2	Sodium/potassium-transporting ATPase subunit beta 2	P14415
ATP1B3	Sodium/potassium-transporting ATPase subunit beta 3	P54709
ATP1B4	Sodium/potassium-transporting ATPase subunit beta 4	Q9UN42
GNAS	Guanine nucleotide-binding protein G(s)	P63092
SLC9A1	Sodium/hydrogen exchanger 1	P19634

As vias de secreção salivar, gástrica e pancreática, por se tratarem de glândulas do tecido epitelial, compartilham muitas semelhanças entre si, relativas ao mecanismo de ação. Essas glândulas compartilham 12 genes, além dos 20 já mencionados, como por exemplo, o receptor muscarínico (CHRM3) (Figura 16).

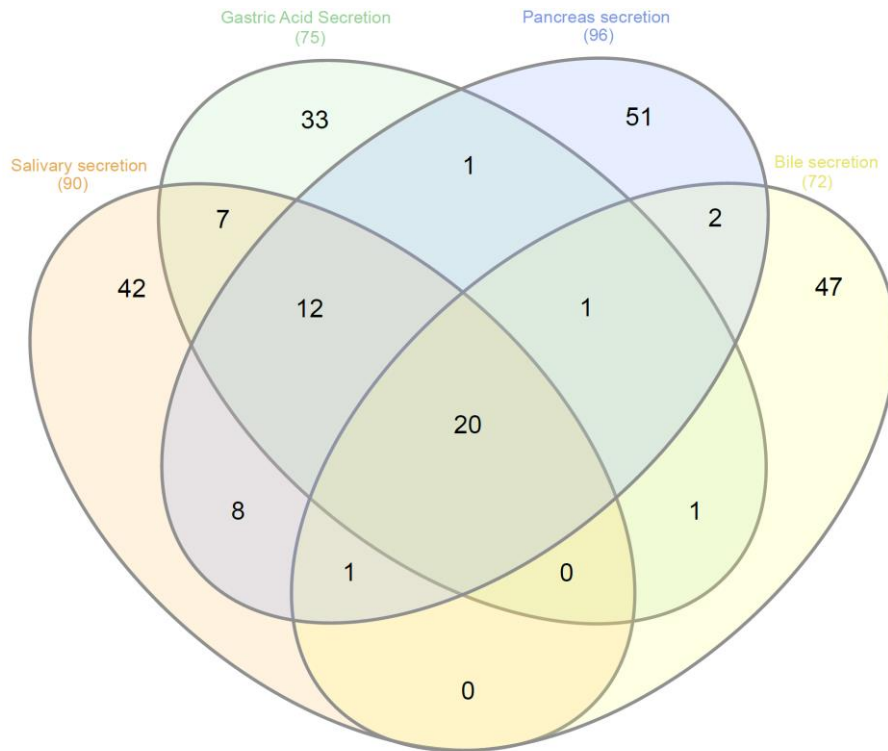


Figura 16 – Diagrama de Venn das proteínas encontradas nas vias da digestão. Os números entre parênteses indicam o número total de genes em cada via.

A análise do LCA das vias indica que em torno de 70-80% dos genes das vias estudadas já estavam presentes em Bilatéria (n6) (Figura 17). Em Tetrapoda (n16) quase todos os componentes já estavam presentes com exceção da via de secreção salivar. Esta última está completa após a evolução de Hominidae (n28). Como pode ser observado na Figura 18, as quatro vias estudadas possuem muitos genes que surgiram em Cellular organisms (n1), Eukaryota (n2), e em seguida o aparecimento de novos genes se reduz. Ao longo da linhagem humana, novos componentes foram sendo acrescentados às vias em menor quantidade. A glândula salivar é a única que apresenta inovações nos seus componentes após o surgimento dos Primatas (n23), com picos além de Boreutheria (n21) que não refletem a distribuição do total de genes. Essas novas proteínas da saliva fazem parte do arsenal proteico secretado na cavidade oral. Para as quatro vias a contribuição de genes de Bilateria (n6) é alta, mais que o esperado pela distribuição dos genes totais. E com exceção da secreção biliar, há bem menos genes participantes sendo originados em Euteleostomi (n13), em comparação com os genes totais.

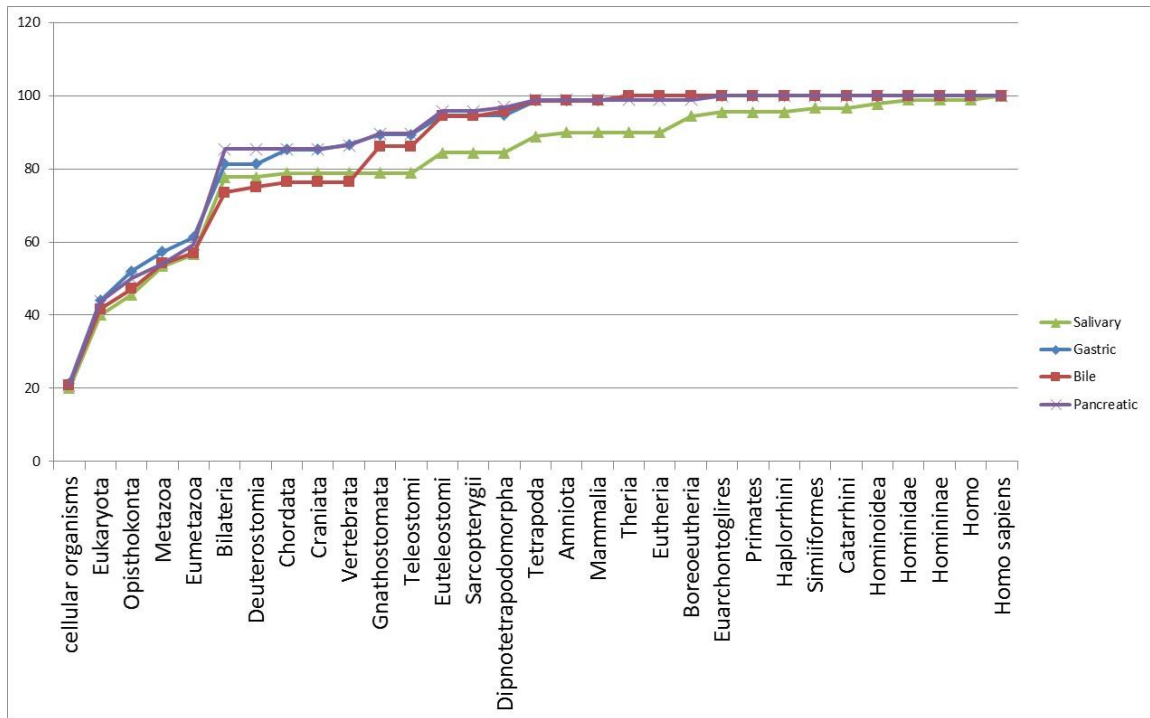


Figura 17 - Distribuição acumulada dos genes para as vias de secreção salivar (verde), gástrica (azul), pancreática (roxo) e biliar (vermelho). Valores em porcentagem do total de genes em cada via.

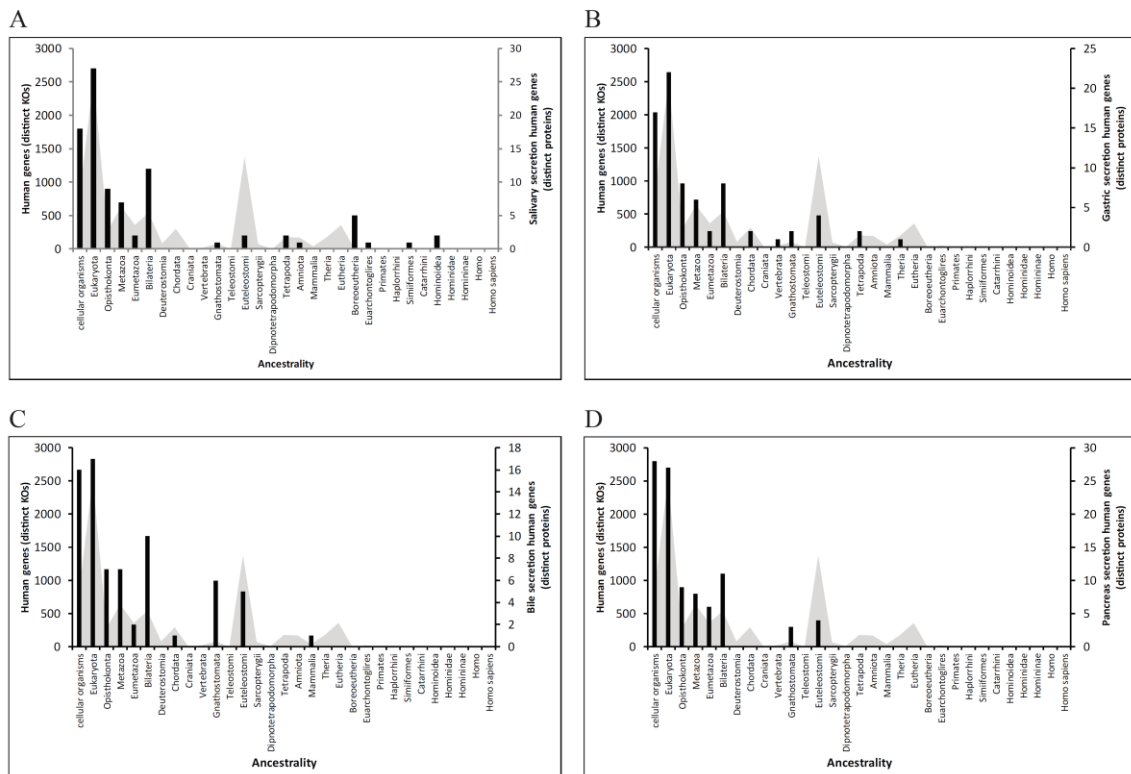


Figura 18 – Distribuição absoluta dos genes na linhagem evolutiva humana. (A) Secreção salivar; (B) Secreção gástrica; (C) Secreção biliar; (D) Secreção pancreática. A área cinza representa o surgimento de todos os agrupamentos Kegg Orthology contendo genes humanos (base KEGG).

4.3. Comparando as vias de secreção do sistema digestivo em mamíferos

As vias estudadas nesse trabalho representam modelos da secreção de glândulas do sistema digestivo de humanos. Nesse trabalho utilizamos dados de filogenia molecular envolvendo o proteoma completo de 93 espécies de mamíferos ([Tabela Suplementar 1](#)) provenientes do NCBI RefSeq para comparar os componentes das vias de secreção do sistema digestivo humano com as dos demais mamíferos. Para isso, utilizamos o programa ELDOgraph, que analisa as distâncias filogenéticas e determina para cada proteína o ortólogo menos divergente externo (ELDO) à unidade taxonômica considerada. As análises do ELDOgraph apresentadas neste trabalho foram realizadas utilizando a “espécie” como unidade taxonômica e o *threshold* de 0,05. O parâmetro *threshold* permite que outras proteínas que não são as menos divergentes sejam também contabilizadas como ELDO caso a diferença entre a sua distância e a menor distância encontrada seja suficientemente pequena. Um *threshold* de 0,05, por exemplo, permite que proteínas sejam consideradas como ELDO caso a diferença entre a sua distância e a menor distância encontrada não ultrapasse o valor correspondente a 5% da menor distância. Em termos práticos, essa operação permite empates, quando algumas proteínas têm praticamente a mesma distância.

Realizando uma análise generalizada do ELDOgraph envolvendo todas as proteínas humanas (n=19.589), é possível verificar que a maioria das proteínas humanas possuem como ELDO proteínas de duas espécies do gênero *Pan*, o *Pan troglodytes* (50,1%) e *Pan paniscus* (50,3%), e de *Gorilla gorilla* (31,4%). Analisando o ELDO das proteínas envolvidas em cada processo de secreção do sistema digestivo (Figura 19B, C, D e E), é possível observar que as espécies que estão mais presentes como ELDO das proteínas humanas continuam sendo as mesmas, mas com uma ligeira diferença nas suas proporções. Nas vias de secreção gástrica, biliar e pancreática, é possível observar que o *Pan paniscus* é a espécie que mais se apresenta como ELDO das proteínas humanas. Em contrapartida, na via da secreção salivar, podemos verificar que a maior parte das proteínas humanas nesta via possuem como ELDO proteínas de *Pan troglodytes*.

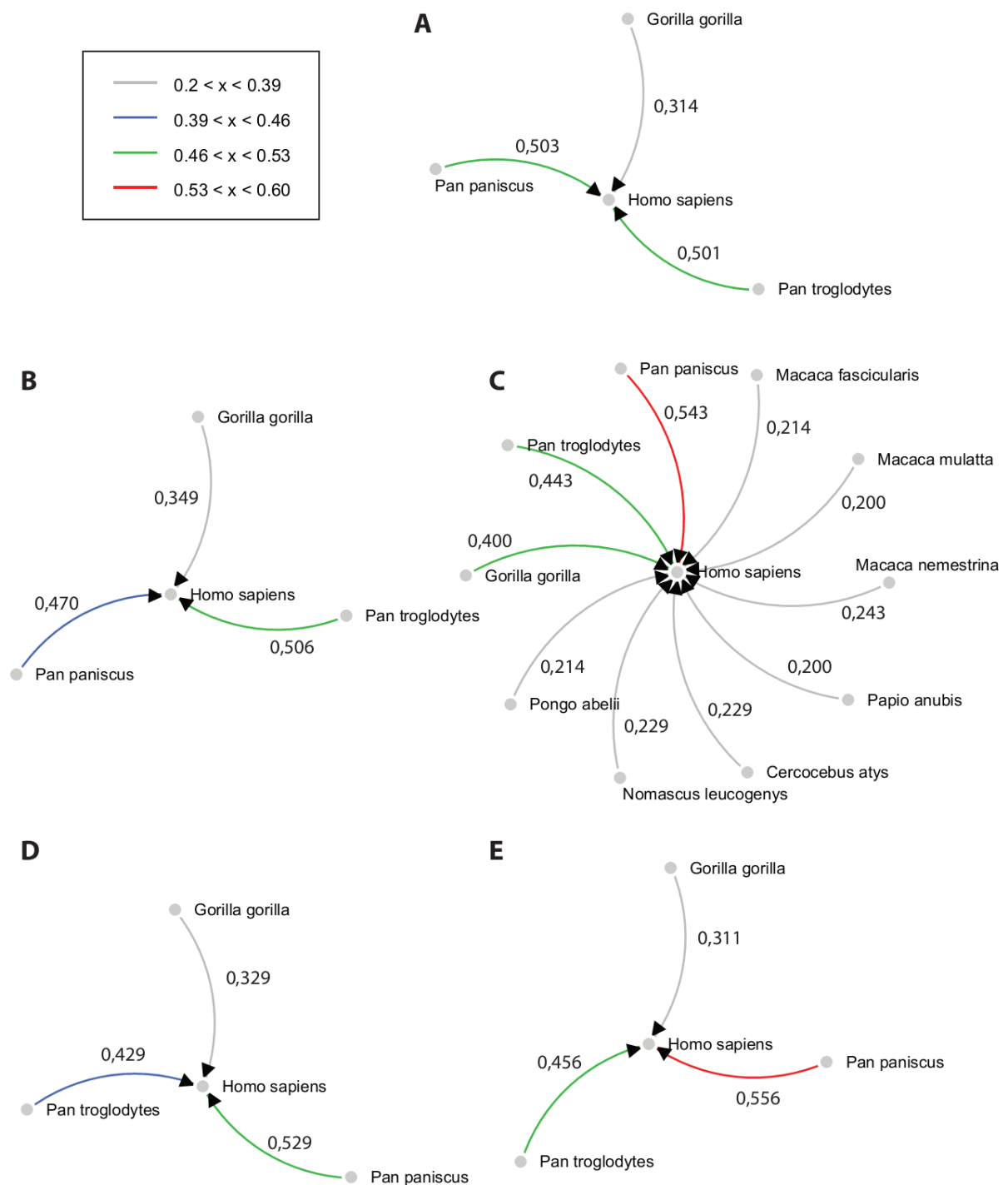


Figura 19 – Grafo gerado pelo programa ELDOgraph. O grafo mostra a proporção dos genes compartilhados entre diferentes espécies de organismos. (A) Genoma completo de humano; (B) Secreção salivar; (C) Secreção gástrica; (D) Secreção biliar; (E) Secreção pancreática. Os números sobre as arestas indicam a proporção dos genes humanos analisados que cada espécie são ELDO.

Procuramos também verificar quais são as ordens diferentes de primatas que compartilham mais ELDO com as proteínas humanas. Para isso, realizamos novamente a análise com o programa ELDOgraph, mas, desta vez, utilizando a ordem como a unidade

taxonômica. A análise do ELDO, considerando todas as proteínas humanas, indicou que a maior parte das proteínas humanas possuem como ELDO proteínas de espécies da ordem Dermoptera (29,4%) seguida por Rodentia (19%), sendo ambos pertencentes a superordem Euarchontoglires, que é a mesma superordem de Primatas. As outras ordens onde pelo menos 10% das proteínas humanas têm elas como ELDO são: Scadentia (superordem Euarchontoglires), Perissodactyla, Carnivora, e ord_of_Cetartiodactyla (superordem Laurasiatheria) (Figura 20A). Analisando o ELDO das proteínas envolvidas em cada processo de secreção do sistema digestivo (Figura 20B, C, D e E), é possível observar que as ordens que estão mais presentes como ELDO dos genes humanos continuam as mesmas. No entanto, as proporções variam, sendo Dermoptera a ordem que mais compartilha os componentes das vias da saliva e da bile e Rodentia a que compartilha mais genes das vias de secreção gástrica. Os componentes da via de secreção pancreática, por sua vez, apresentam ELDO com espécies das ordens Rodentia e Carnivora na mesma proporção. Note-se que Carnivora não pertence à superordem do homem, mas à superordem Laurasiatheria.

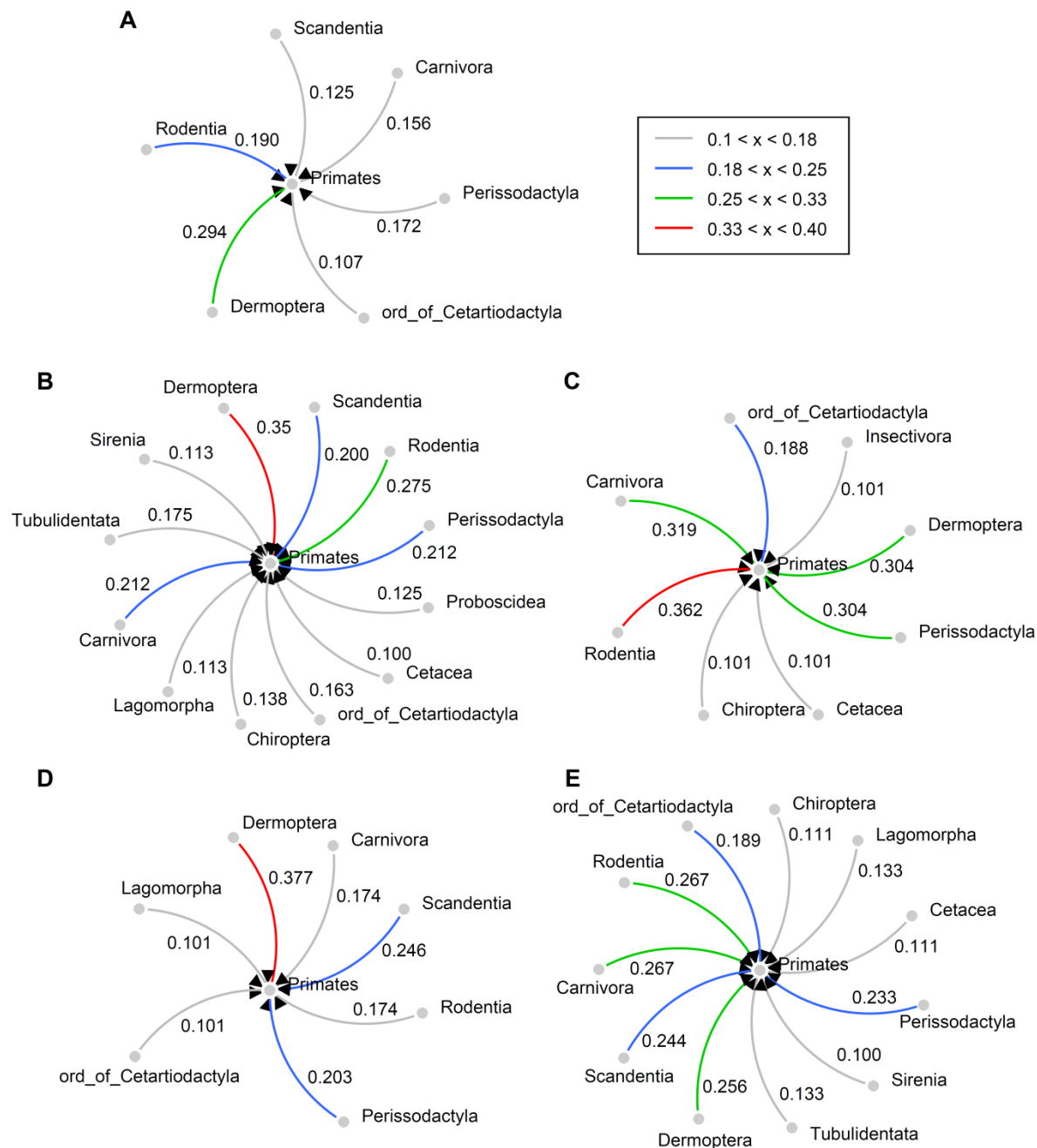


Figura 20 – Grafo gerado pelo programa ELDOgraph. O grafo mostra a proporção dos genes compartilhados entre diferentes organismos. (A) Genoma completo de humano; (B) Secreção salivar; (C) Secreção gástrica; (D) Secreção biliar; (E) Secreção pancreática. Os números sobre as arestas indicam a proporção dos genes humanos analisados que cada ordem é ELDO.

Outros arquivos de saída do programa ELDOgraph foram aproveitados para complementar na análise comparativa dos componentes das vias analisadas neste trabalho. A partir destes arquivos, foi possível analisar a relação das proteínas que se encontram presentes ou ausentes no proteoma de cada mamífero presentes nos dados como também verificar o comportamento dessas proteínas ao longo da evolução dos

mamíferos ([Tabela Suplementar 2](#)). A partir desta análise, foi possível verificar que alguns componentes exclusivos da via de secreção da saliva são exclusivos na ordem primatas, como por exemplo o HTN1, HTN3 e STATH, o que provavelmente resultou em um ganho de função nesta via entre os primatas. Existem também proteínas que foram deletadas em alguns clados específicos dentro de mamífero, com destaque a ordem Cetacea, onde foi verificado a ausência de seis proteínas envolvidas nas vias em análise. Na Tabela 2 encontra-se sumarizados os componentes que foram deletados ou que são exclusivos para um determinado clado taxonômico.

Tabela 2. Proteínas deletadas ou mantidas nas espécies de mamíferos

Via	Proteína	Descrição da proteína	Ordens
Presença			
Saliva	HTN1	Histatin-1	Primatas
Saliva	HTN3	Histatin-3	Primatas
Saliva	STATH	Statherin	Hominídeos
Bile/gástrica	PRKACG	Protein kinase alpha	Primatas
Ausência			
Saliva	MUC7	Mucin-7	Cetáceos, ruminantes e na maioria dos roedores
Bile/gástrica/ saliva	CALML3	Calmodulin-like protein 3	Cetáceos, insetívoras, cetartiodactílos, e na maioria dos perissodactílos
Bile/gástrica/ saliva	CALML5	Calmodulin-like protein 5	Cetáceos, ruminantes e em histicomorfos (Rodentia)
Bile/gástrica/ saliva	CALML6	Calmodulin-like protein 6	Insetívoras e na maioria dos roedores
Bile	BAAT	Bile acid-CoA:amino acid N-acyltransferase	Microquirópteros, insetívoros e na maioria dos cetartiodactílos
Bile	UGT2B4	UDP-glucuronosyltransferase 2B4	Felinos, cetáceos
Bile	SULT2A1	Bile salt sulfotransferase	Carnívoros e na maioria dos cetáceos
Pancreas	CELA2A	Chymotrypsin-like 2A	Carnívoros
Pancreas	CELA2B	Chymotrypsin-like 2B	Carnívoros
Pancreas	PLA2G2A	Phospholipase A2	Carnívoros e cetáceos
Pancreas	CTRL	Chymotrypsin-like protease	Lagomorfos, ruminantes e na maioria dos perissodactílos

Aproveitamos também os dados de distância fornecidos pelo programa ELDOgraph para realizar uma ampla verificação do seu comportamento ao longo da evolução. Para isso, dados de distância entre a proteína humana com a proteína de cada mamífero

analisado foram colocados em um gráfico de dispersão (Figura 21 a Figura 24). Estes gráficos nos permitiram analisar comparativamente o grau de modificações ocorridas entre pares de proteínas de organismos (humano *versus* demais mamíferos) que possuem diferentes níveis de ancestralidade comum. Em geral, a dispersão dos pontos na maior parte das proteínas envolvidas nas vias em análise não ultrapassa 0,5 unidades de distância. No entanto, existem outras proteínas que se destacam por apresentarem uma dispersão mais ampla por terem sofrido um maior número de modificações durante a evolução dos mamíferos. Uma parte destes componentes são aqueles que se encontram envolvidos em duas ou mais vias analisadas neste trabalho, que é o caso do ATP1B3, CALML5, KCNQ1 e CD38. No entanto, é interessante notar que a maior parte dos componentes que parecem ter sofrido mais modificações são exclusivos de cada via. Nas vias de secreção salivar e pancreática, que são vias que apresentam proteínas em suas secreções, pudemos notar que as proteínas que mais sofreram modificações foram aquelas presentes nas próprias secreções. Entre elas podemos citar as cistatinas (CST), LL37, LYZ, MUC5B e MUC7, na via da secreção salivar (Figura 10), e o CELA, PLA2G, PLRP e PRSS, na via da secreção pancreática (Figura 14). Já na via da secreção gástrica (Figura 13), verifica-se que a GAST e MYLK3, componentes exclusivos desta via, foram as que mais sofreram modificações. Na via da secreção biliar, por sua vez, as proteínas que mais sofreram modificações foram aquelas envolvidas nos processos de conjugação e detoxicação do ácido biliar, como o SLC27A5, BAAT, CYP3A4 e o SULT2A1, e também a SLCO1 e a SLC51B, que são proteínas de membrana responsável pelo transporte do ácido biliar entre o meio intra e extracelular.

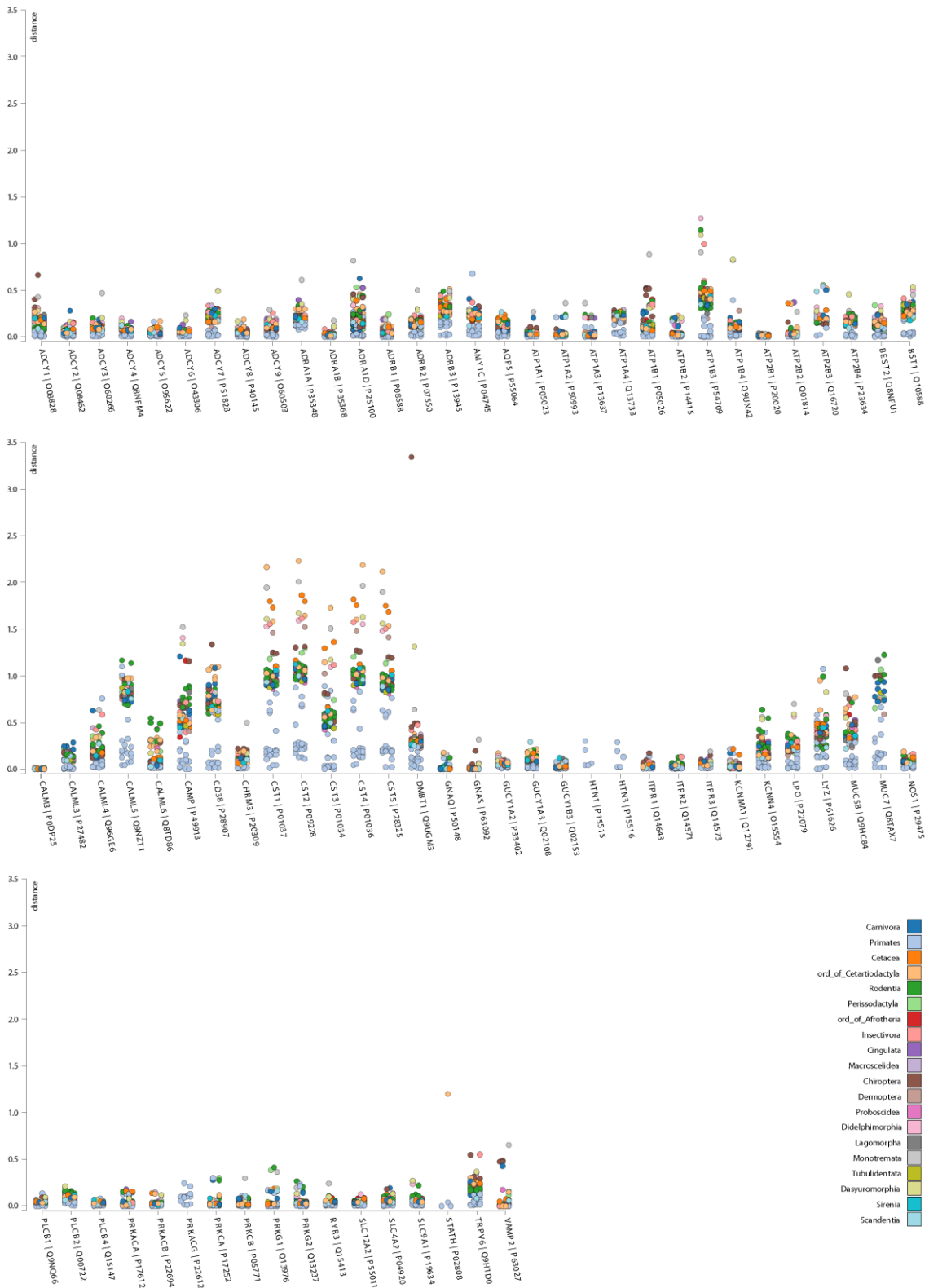


Figura 21 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção salivar em espécies de mamíferos.

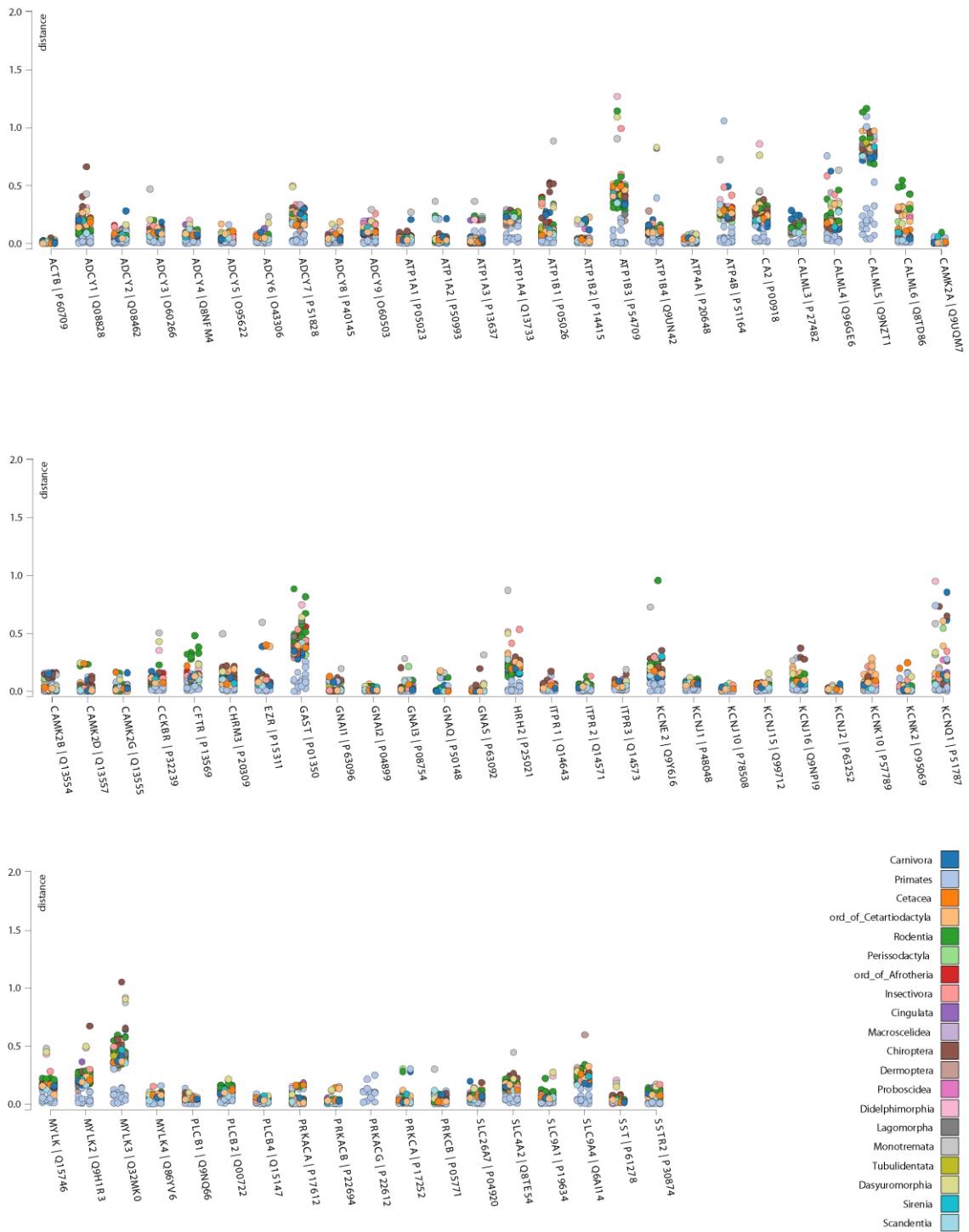


Figura 22 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção gástrica em espécies de mamíferos.

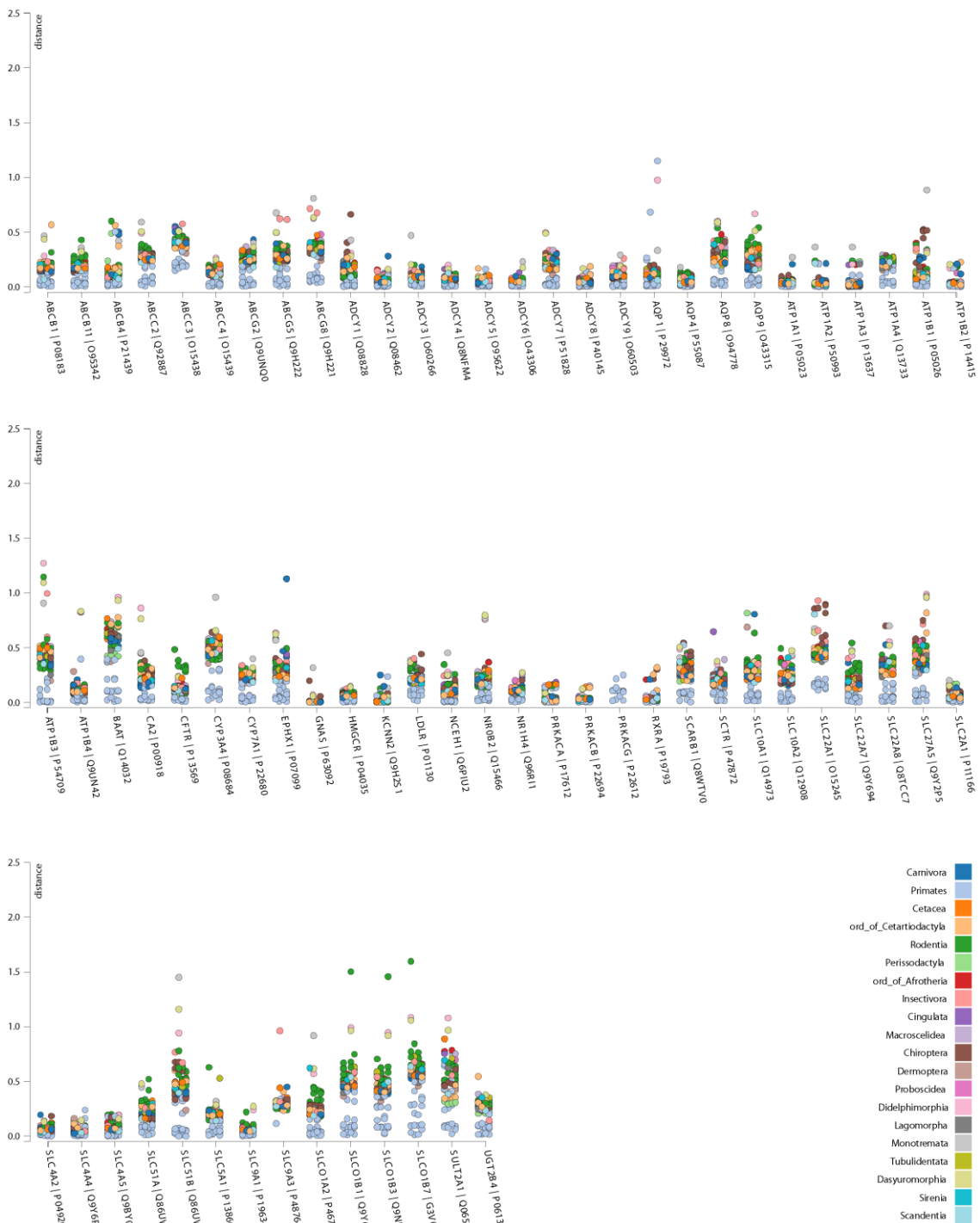


Figura 24 – Gráfico de dispersão representando as distâncias filogenéticas entre os genes da via de secreção biliar em espécies de mamíferos.

As distâncias filogenéticas utilizadas na análise anterior para comparar o comportamento dos genes foram utilizadas novamente para realizar um estudo comparativo entre os 92 mamíferos utilizados nesta análise. Para isso, para cada organismo foi calculada a média das distâncias de todos os componentes que ele apresenta na via em relação às proteínas humanas e o valor desenhado em um gráfico de barras

(Figura 25). Neste gráfico, podemos perceber que em todas as vias analisadas, os organismos que pertencem a ordem Primata possuem a menor média de distância e ao mesmo tempo maior número de componentes na via de humano (Figura 25, barras pretas). No entanto, existem casos onde espécies de outras ordens apresentam uma média de distância menor que a distância do primata mais divergente (Figura 25, barras cinzas). Isso pode ser verificado em 14 espécies na via de secreção da saliva (Figura 25A), 21 espécies na via de secreção gástrica (Figura 25B) e em 3 espécies na via de secreção pancreática (Figura 25C). Em algumas espécies, essa média baixa pode ser explicada pela ausência de alguns componentes que seriam divergentes e passam a não estar no mesmo grupo de ortólogos. Na via da secreção salivar, seria o caso das espécies *Condylura cristata*, *Miniopterus natalinas*, *Panthera tigris*, *Acynonyx jubatus* e *Ochtona princeps* que apresentam uma média de distância menor à média encontrada no primata mais divergente *Carlito syrichta*. Nestas espécies, apesar da média baixa, observa-se que o número de componentes presentes nestes organismos também é baixa, permitindo-nos supor que existe uma grande chance de que as proteínas ausentes sejam aquelas que tendem a acumular mais modificações. Os genes da secreção biliar tendem a ser mais distantes que os genes das demais vias (Figura 25D). Já outros organismos apresentam ser promissores para os estudos de temas sob o contexto deste trabalho em comparação com as vias humanas. É o caso das espécies *Tupaia chinensis* (ordem Scandentia) e *Odobenus rosmarus* (ordem Carnivora), que apresentaram uma média menor que o primata mais divergente e ao mesmo tempo mantiveram um número alto de componentes em pelo menos três vias (secreção gástrica, pancreática e salivar).

Os organismos que possuem os genes mais divergentes e menos componentes para todas as vias pertencem às superordens Prototheria e Metatheria, que são clados que divergiram mais cedo na escala evolutiva de mamíferos. Com base nos resultados observados podemos supor que muitos genes associados a secreções do trato digestivo tenham sido perdidos em metatérios e em ornitorrinco, já que a origem de muitos genes que se encontram ausentes nesses organismos antecede à época da origem de mamíferos. De fato, existem relatos da perda de genes da secreção gástrica em *O. anatinus* (ORDOÑEZ et al., 2008), indicando que a hipótese da perda de genes nas outras vias também é possível.

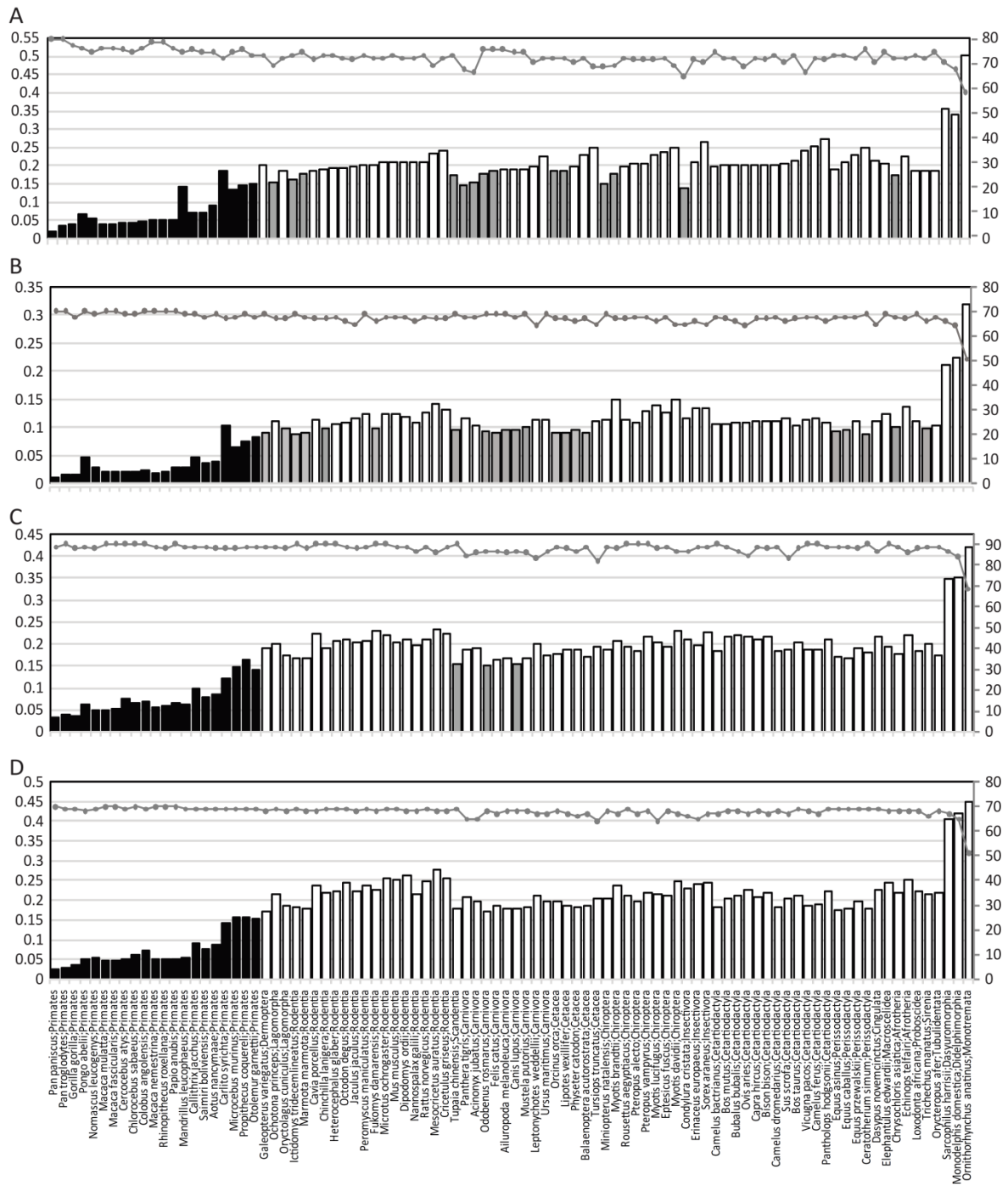


Figura 25. Gráfico das médias das distâncias filogenéticas das proteínas de mamífero. As barras pretas representam os primatas. As barras cinzas representam espécies cuja média das distâncias é menor que o primata com a maior média. O eixo da esquerda indica as distâncias filogenéticas dos genes em relação aos genes de humano. O eixo da direita indica a quantidade de proteínas encontradas em cada espécie. (A) Secreção salivar; (B) Secreção gástrica; (C) Secreção pancreática; (D) Secreção biliar.

5. DISCUSSÃO

Neste trabalho foram realizadas diversas análises no estudo comparativo e evolutivo da heterotrofia usando como modelo as principais vias de secreção do sistema digestivo humano: secreção salivar, gástrica, pancreática e biliar, responsáveis pela digestão química dos alimentos. Apesar de muitos trabalhos sobre sequenciamento de genomas apresentarem estudos de genômica comparativa, estes trabalhos se atêm a abordagens generalistas a respeito do genoma sequenciado. No presente trabalho utilizamos informações depositadas em bases de dados e software de comparação de sequências para realizar um estudo temático a respeito da heterotrofia focando nos processos digestivos e buscando entender como o sistema se forma a partir do estudo de seus componentes. Para isso, contamos com as ferramentas à nossa disposição, tanto aquelas já publicadas quanto as desenvolvidas ou em desenvolvimento em nosso grupo de pesquisa.

Com os resultados obtidos observamos que muitos componentes dos sistemas já existiam antes da existência do órgão no qual atuam. Após analisar os resultados do LCA das vias estudadas nesse trabalho foram determinados três níveis de origem dos genes: o nível mais antigo que corresponde aos genes que surgiram entre as linhagens de Cellular organisms e Opisthokonta (azul); o nível intermediário que corresponde aos genes originados entre Metazoa e Chordata (verde); e o nível mais recente que corresponde aos genes que surgiram entre as os clados de Vertebrata ao *Homo sapiens* (rosa). Para alguns genes não foi possível indicar com precisão um clado de origem, pois análises com diferentes abordagens geraram resultados discordantes. Neste caso privilegiamos análise manual com um software que executa busca com BLAST, produz árvore filogenética e as anota com informações de LCA e taxonômicas, facilitando a observação de ramos com a função da proteína em análise. Algumas vezes a proteína relatada na via está em um dos ramos da árvore, e o ramo irmão contém um gene humano com nome distinto (*Gene Symbol* distinto), embora sejam observadas duplicações remotas dando origem à família, a identidade do *Gene Symbol* facilita a determinação da origem. Seria simples para um software determinar esta separação se a ocorrência de *Gene Symbols* não fosse tão rara. Em alguns casos, todavia, a separação de ramos é tamanha que é possível ser identificada por software. Neste trabalho só foram submetidos a análise manual inferências onde havia conflito entre o método escolhido, SeedServer, e o Geneages.

Como pode ser observado nas figuras 10, 13, 14 e 15 (Figura 1) os primeiros produtos gênicos a surgirem nas vias estão envolvidos em processos de sinalização celular e alguns desses genes estão presentes nas quatro vias estudadas. Evidentemente esses produtos existem antes do sistema ser originado, pois atuam de forma genérica nestas e, supostamente, em mais vias. A maioria desses genes surgiu em Cellular organisms (n1) e Eukariota (n2), como mostra a Figura 18, antes da origem dos animais. Embora a maioria dos componentes encontrados nas vias seja antiga, os componentes mais recentes, como os receptores e alguns transportadores, essenciais para a função de secreção, surgem a partir do clado Metazoa (n4), ou seja, são restritos aos animais. Para as vias estudadas, há um sensível marco em Bilateria (n6), sendo que a ocorrência de genes originados nesse momento ultrapassa o *background* do conjunto de todos os genes do KEGG (Figura 18). Mesmo que em metazoários primitivos, como Protozoa e Porifera, a digestão ainda seja um processo intracelular, o desenvolvimento de organismos com órgãos especializados na digestão extracelular requer um controle do processo de secreção de enzimas e ácidos digestivos. Nossos resultados mostram que os receptores de membrana que desencadeiam o processo de secreção, assim como genes que regulam a expressão de outros componentes desse sistema, surgem a partir do aparecimento dos metazoários. Subsequentemente, são observados genes com ocorrência mais recente, em vertebrados e mesmo em mamíferos. Era esperado, com base em outros estudos desenvolvidos no nosso grupo de pesquisa, que alguns genes com funções mais especializadas adicionassem funcionalidades ao sistema, mas ocorrendo de forma esparsa. Todavia, foi verificado que na resposta a interferons, uma origem próxima para ligantes, receptores, transdutores de sinal em um período concentrado (COSTA, 2015). Assim, seria possível que alguns módulos das vias tivessem ocorrência alta de genes modernos.

Nas vias de secreção pancreática e biliar as células do duto e os colangiócitos desempenham uma função acessória nesses sistemas, respectivamente. Essas células são responsáveis por secretar bicarbonato, o qual auxilia a balancear o pH das secreções. Os componentes que exercem as funções chave nesse processo tem origem recente, indicando que a secreção de bicarbonato é um refinamento do fenômeno de secreção nessas vias. No pâncreas a secreção das células do duto confere vantagem ao evitar que as enzimas presentes na secreção pancreática agridam o tecido onde é secretado.

A saliva contém uma grande quantidade de proteínas recentes, com muitos genes exclusivos de mamíferos. Essas proteínas auxiliam na lubrificação e proteção da cavidade oral, servindo como uma barreira contra a entrada de microrganismos. Apesar de muitos

organismos apresentarem glândulas que secretam fluido com função de lubrificação do alimento, Yonge (1937) sugere que o termo glândula salivar seria mais adequado para as glândulas dos vertebrados, dada sua composição. De fato, estudos recentes mostram que muitas proteínas presentes na saliva são um produto da evolução de espécies de mamíferos em resposta à pressão patogênica e aos hábitos alimentares baseados em carboidratos (DE SOUSA-PEREIRA et al., 2014; HARDY et al., 2015; TAKEHARA et al., 2013).

No caso da secreção gástrica, genes mais recentes são indutores da secreção, agonistas de receptores, no entanto, originados na categoria anterior. Isto significa que os ligantes mais similares aos humanos são verificados somente em clados mais recentes, enquanto uma proteína diferente talvez tenha ocorrido mais remotamente, ou então o receptor atual seja derivado de receptores mais antigos, mas suficientemente similares. Nossa abordagem foi direcionada ao sistema como existe em humanos, seria necessário estudar as vias de secreções digestivas de *Drosophila melanogaster* por exemplo para entender como as proteínas interagem no organismo que existe há mais tempo.

Apesar da datação obtida, para muitos genes não há uma relação entre sua origem e a origem de sua função. Assim, alguns genes recentes desempenham funções antigas, como o canal de cálcio TRPC1 que atua na via pancreática, e surge em Gnathostomata (n11), mas sua função é datada em Cellular organisms (n1). Essa proteína contribui para elevar os níveis de cálcio no citoplasma, uma vez que essa molécula é essencial no processo de secreção. A ocorrência de proteínas recentes com funções remotas sugere que sejam mais especializadas em sua regulação, etc. Deste modo, novos genes que desempenham funções já existentes podem contribuir para tornar o processo mais eficiente, maximizando a resposta da célula. Por outro lado, casos como o do gene NR1H4 também foram observados. Esse gene surge em Metazoa (n4), mas a sua função na regulação da expressão de genes da secreção biliar é uma característica de Amniota (n17). Genes antigos podem ter sido recrutados para atuar em novos processos como parece ser o caso do gene NR1H4, ou sub ou neo funcionalizado. Uma análise detalhada da origem das funções bioquímicas, utilizando a base Gene Ontology, está em andamento e será apresentada em trabalhos futuros. Atualmente a ferramenta ARCoBALeno realiza a marcação das proteínas pela função molecular mais recente, mas estamos colaborando com o acoplamento de um formulário de curadoria, aparentemente necessária.

Como as vias do sistema digestivo humano foram usadas como modelo neste estudo, nós realizamos uma análise para identificar quais organismos ou OTU

(*Operational Taxonomic Unit*) tinham maior número de proteínas próximas às de humano. Nesse trabalho, as OTUs utilizadas foram espécies e ordens de mamíferos. Quando analisado em nível de espécie as que possuem mais ELDO com humano são as do gênero *Pan* (*P. paniscus* e *P. troglodytes*); esse resultado se mantém considerando-se todos os genes de humano ou apenas aqueles presentes nas vias estudadas. Isso já era esperado, uma vez que essas espécies formam um grupo irmão com humanos. Quando usamos ordem como OTU, verificamos que aquelas que possuem mais ELDO com humanos são Dermoptera e Rodentia, nesta ordem, que pertencem à mesma superordem dos Primatas. Todavia, os dados sugerem grande proximidade com Perysodactyla e Carnivora dentre as ordens pertencentes à superordem irmã, Laurasiatheria. Analisando-se as proteínas da bile, a ordem Perysodactyla apresenta mais ELDO que Rodentia. Em conjunto as análises mostram que as vias do sistema digestivo em mamíferos compartilham muitas similaridades, algumas proteínas apresentando distâncias bem pequenas e outras, maiores. Embora em determinadas espécies seja observada a ausência de alguns componentes, é interessante notar que existem proteínas que se mantiveram muito conservadas mesmo entre espécies distantes. Nota-se que membros de ordens mais distantes algumas vezes apresentam distâncias médias menores às proteínas humanas que as das vias de mamíferos mais distantes.

6. CONCLUSÃO

Nesse trabalho apresentamos um estudo comparativo dos genes envolvidos nos principais processos de secreção do sistema digestivo. Nossos resultados mostram que, ao longo da evolução, os genes com função de sinalização citoplasmática surgiram primeiro, antecedendo o aparecimento de receptores e de outras proteínas que desempenham atividades mais específicas associadas à digestão. Assim, após o aparecimento de genes que realizam a manutenção de processos celulares básicos, como regulação de pH e sinalização de cálcio, surgem componentes com funções fisiológicas mais sofisticadas. Além disso, foi possível identificar que alguns componentes seguem conservados em diferentes espécies de mamíferos, enquanto alguns apresentam distâncias maiores.

A partir do estudo dos componentes das vias de secreção salivar, gástrica, pancreática e biliar foi possível desenhar um panorama a respeito da evolução do processo digestivo, contribuindo assim para a história da evolução desse sistema. Muitas funções aparecem com a origem da heterotrofia animal e algumas só ocorrem em vertebrados, quando os órgãos estudados se desenvolvem. Este trabalho, apesar de utilizar vias publicadas na base KEGG, adiciona uma descrição biológica detalhada, permitindo um rápido entendimento do cenário no qual os produtos gênicos atuam.

7. REFERÊNCIAS

- ALLEN, A.; FLEMSTRÖM, G. Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin. **American Journal of Physiology - Cell Physiology**, v. 288, n. 1, p. C1–C19, 2005.
- ALREFAI, W. A.; GILL, R. K. Bile acid transporters: Structure, function, regulation and pathophysiological implications. **Pharmaceutical Research**, v. 24, n. 10, p. 1803–1823, 2007.
- ALTENHOFF, A. M. et al. The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. **Nucleic Acids Research**, v. 43, n. D1, p. D240–D249, 28 jan. 2015.
- AMADO, F. et al. Salivary peptidomics. **Expert Review of Proteomics**, v. 7, n. 5, p. 709–721, 2010.
- AMBUDKAR, I. S. Regulation of Calcium in Salivary Gland Secretion. **Critical Reviews in Oral Biology & Medicine**, v. 11, n. 1, p. 4–25, 2000.
- ARMSTRONG, M. A.; TURTURRO, M. A. Salivary Gland Emergencies. **Emergency Medicine Clinics of North America**, v. 31, n. 2, p. 481–499, 2013.
- ASHBURNER, M. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25–29, maio 2000.
- BARBOSA-SILVA, A. et al. Clustering of cognate proteins among distinct proteomes derived from multiple links to a single seed sequence. **BMC bioinformatics**, v. 9, p. 141, 2008.
- BARDEESY, N.; DEPINHO, R. A. Pancreatic cancer biology and genetics. **Nature reviews. Cancer**, v. 2, n. 12, p. 897–909, 2002.
- BAUM, B. J. Principles of saliva secretion. **Annals of the New York Academy of Sciences**, v. 694, p. 17–23, 1993.
- BENNICK, A. Salivary proline-rich proteins. **Molecular and Cellular Biochemistry**, v. 45, n. 2, p. 83–99, nov. 1982.
- BERRIDGE, M. J.; BOOTMAN, M. D.; RODERICK, H. L. Calcium signalling: dynamics, homeostasis and remodelling. **Nature reviews. Molecular cell biology**, v. 4, n. 7, p. 517–29, 2003.
- BERTON, A. et al. Individual and combined action of pancreatic lipase and pancreatic lipase-related proteins 1 and 2 on native versus homogenized milk fat globules. **Molecular Nutrition and Food Research**, v. 53, n. 12, p. 1592–1602, 2009.
- BI, Y.; WILLIAMS, J. {A}. Receptor biology and signal transduction in pancreatic acinar cells. **Current Opinion in Gastroenterology**, v. 20, n. 5, p. 427–434, 2004.
- BOYER, J. L. Bile Formation and Secretion. In: **Comprehensive Physiology**.

Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013. v. 3p. 1035–1078.

CANCELA, J. M. et al. Two different but converging messenger pathways to intracellular Ca(2+) release: the roles of nicotinic acid adenine dinucleotide phosphate, cyclic ADP-ribose and inositol trisphosphate. **The EMBO journal**, v. 19, n. 11, p. 2549–2557, 2000.

CAPELLA-GUTIÉRREZ, S.; SILLA-MARTÍNEZ, J. M.; GABALDÓN, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. **Bioinformatics**, v. 25, n. 15, p. 1972–1973, 2009.

CATALÁN, M. A; NAKAMOTO, T.; MELVIN, J. E. The salivary gland fluid secretion mechanism. **The journal of medical investigation : JMI**, v. 56 Suppl, p. 192–196, 2009.

CHU, J. Y. S.; YUNG, W. H.; CHOW, B. K. C. Secretin: A pleiotrophic hormone. **Annals of the New York Academy of Sciences**, v. 1070, p. 27–50, 2006.

COSTA, V. R. DE M. **Estudo da ancestralidade de genes componentes do sistema imune**. [s.l.] UFMG, 2015.

CROFT, D. et al. Reactome: a database of reactions, pathways and biological processes. **Nucleic Acids Research**, v. 39, n. Database, p. D691–D697, 1 jan. 2011.

DE SOUSA-PEREIRA, P. et al. Evolution of C, D and S-type cystatins in mammals: An extensive gene duplication in primates. **PLoS ONE**, v. 9, n. 10, p. 1–7, 2014.

ESTELLER, A. Physiology of bile secretion. **World Journal of Gastroenterology**, v. 14, n. 37, p. 5641–5649, 2008.

EYDOUX, C. et al. Structure of human pancreatic lipase-related protein 2 with the lid in an open conformation. **Biochemistry**, v. 47, n. 36, p. 9553–9564, 2008.

FERNANDES, G. R. et al. A procedure to recruit members to enlarge protein family databases--the building of UECOG (UniRef-Enriched COG Database) as a model. **Genetics and molecular research : GMR**, v. 7, n. 3, p. 910–924, 2008.

FORTE, J. G.; ZHU, L. Apical recycling of the gastric parietal cell H,K-ATPase. **Annu Rev Physiol**, v. 72, p. 273–296, 2010.

GELBERG, H. B. Comparative Anatomy, Physiology, and Mechanisms of Disease Production of the Esophagus, Stomach, and Small Intestine. **Toxicologic Pathology**, v. 42, n. 1, p. 54–66, 2014.

GHOSH, S. Early steps in reverse cholesterol transport: cholesteryl ester hydrolase and other hydrolases. **Current opinion in endocrinology, diabetes, and obesity**, v. 19, n. 2, p. 136–41, 2012.

GONÇALVES, C. A. X. **Ferramentas e métodos para estudo da evolução de processos biológicos e funções moleculares do Homo sapiens**. [s.l.] UFMG, 2016.

GUEDES, R. L. M. **Desenvolvimento das ferramentas SeedServer, para**

agrupamento de sequências protéicas homólogas e U-MAGE, para propagação de ontologia funcional. [s.l.] UFMG, 2013.

HALL, J. E.; GUYTON, A. **Guyton & Hall. Tratado de Fisiologia Médica.** 12. ed. [s.l.] Elsevier, 2011.

HARDY, K. et al. The Importance of Dietary Carbohydrate in Human Evolution. **The Quarterly Review of Biology**, v. 90, n. 3, p. 251–268, 2015.

HEITZMANN, D.; WARTH, R. Physiology and pathophysiology of potassium channels in gastrointestinal epithelia. **Physiological reviews**, v. 88, n. 3, p. 1119–1182, 2008.

HELMERHORST, E. J. et al. Identification of Lys-Pro-Gln as a novel cleavage site specificity of saliva-associated proteases. **Journal of Biological Chemistry**, v. 283, n. 29, p. 19957–19966, 2008.

HELMERHORST, E. J.; OPPENHEIM, F. G. Saliva: a Dynamic Proteome. **Journal of Dental Research**, v. 86, n. 8, p. 680–693, ago. 2007.

HERRING, B. P. et al. Regulation of myosin light chain kinase and telokin expression in smooth muscle tissues. **AJP: Cell Physiology**, v. 291, n. 5, p. C817–C827, 1 nov. 2006.

HOLMES, R. S.; COX, L. A. Comparative Structures and Evolution of Vertebrate Carboxyl Ester Lipase (CEL) Genes and Proteins with a Major Role in Reverse Cholesterol Transport. **Cholesterol**, v. 2011, p. 1–15, 2011.

HOWE, V. et al. New insights into cellular cholesterol acquisition: promoter analysis of human HMGCR and SQLE, two key control enzymes in cholesterol synthesis. **Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids**, v. 1862, n. 7, p. 647–657, 2017.

HUERTA-CEPAS, J. et al. PhylomeDB: a database for genome-wide collections of gene phylogenies. **Nucleic Acids Research**, v. 36, n. Database, p. D491–D496, 23 dez. 2007.

HUERTA-CEPAS, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic Acids Research**, v. 44, n. D1, p. D286–D293, 4 jan. 2016.

HUI, D. Y.; HOWLES, P. N. Carboxyl ester lipase. **Journal of Lipid Research**, v. 43, n. 12, p. 2017–2030, dez. 2002.

IDE, M. et al. Relationship between salivary immunoglobulin a , lactoferrin and lysozyme flow rates and lifestyle factors in Japanese children : a cross-sectional study. **Acta Odontologica Scandinavica**, v. 0, n. 0, p. 000, 2016.

ISHIKAWA, Y. et al. Aquaporin-5 water channel in lipid rafts of rat parotid glands. **Biochimica et biophysica acta**, v. 1758, n. 8, p. 1053–60, 2006.

ISHIKAWA, Y.; ISHIDA, H. **Aquaporin water channel in salivary glands.** Japanese

journal of pharmacology, 2000. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/10928320>>

JENSEN, L. J. et al. eggNOG: automated construction and annotation of orthologous groups of genes. **Nucleic Acids Research**, v. 36, n. Database, p. D250–D254, 23 dez. 2007.

KANNO, T. Ca⁺⁺ signaling in the pancreatic acinus. **Pancreas**, v. 16, n. 3, p. 273, 1998.

KIM, D.; STEWARD, M. C. The role of CFTR in bicarbonate secretion by pancreatic duct and airway epithelia. **The journal of medical investigation : JMI**, v. 56 Suppl, p. 336–342, 2009.

KISS, R. S.; SNIDERMAN, A. Shunts, channels and lipoprotein endosomal traffic: a new model of cholesterol homeostasis in the hepatocyte. **Journal of Biomedical Research**, v. 31, n. 2, p. 95–107, 2017.

KONTUREK, S. J. et al. Neuroendocrinology of the pancreas; role of brain-gut axis in pancreatic secretion. **European Journal of Pharmacology**, v. 481, n. 1, p. 1–14, 2003.

KOPIC, S.; MUREK, M.; GEIBEL, J. P. Revisiting the parietal cell. **AJP: Cell Physiology**, v. 298, n. 1, p. C1–C10, 2010.

KOSTERS, A.; KARPEN, S. J. Bile acid transporters in health and disease. **Xenobiotica**, v. 38, n. 7–8, p. 1043–1071, 2008.

KUTMON, M. et al. PathVisio 3: An Extendable Pathway Analysis Toolbox. **PLoS Computational Biology**, v. 11, n. 2, p. 1–13, 2015.

LEE, M. G. et al. Molecular mechanism of pancreatic and salivary gland fluid and HCO₃ secretion. **Physiological reviews**, v. 92, n. 1, p. 39–74, 2012.

LEHEN'KYI, V.; PREVARSKAYA, N. Study of TRP Channels in Cancer Cells. In: **TRP Channels**. [s.l.] CRC Press, 2011.

LI, J. et al. Statherin is an in vivo pellicle constituent: Identification and immuno-quantification. **Archives of Oral Biology**, v. 49, n. 5, p. 379–385, 2004.

LIEBESKIND, B. J.; MCWHITE, C. D.; MARCOTTE, E. M. Towards Consensus Gene Ages. **Genome biology and evolution**, v. 8, n. 6, p. 1812–1823, 2016.

LIGTENBERG, A. J. M. et al. Salivary agglutinin/glycoprotein-340/DMBT1: A single molecule with variable composition and with different functions in infection, inflammation and cancer. **Biological Chemistry**, v. 388, n. 12, p. 1275–1289, 2007.

LIU, Q. et al. Crystal structure of human CD38 extracellular domain. **Structure**, v. 13, n. 9, p. 1331–1339, 2005.

MAILLETTE DE BUY WENNIGER, L.; BEUERS, U. Bile salts and cholestasis. **Digestive and Liver Disease**, v. 42, n. 6, p. 409–418, 2010.

- MEIER, P. J.; STIEGER, B. Bile Salt Transporters. **Annual Review of Physiology**, v. 64, n. 1, p. 635–661, 29 mar. 2002.
- MELO, H. V. F. **FERRAMENTAS E SERVIÇOS ONLINE PARA A ANÁLISE DA ORIGEM CLADÍSTICA DE GENES E VIAS METABÓLICAS**. [s.l.] UFMG, 2014.
- MELVIN, J. E. et al. Regulation of fluid and electrolyte secretion in salivary gland acinar cells. **Annual review of physiology**, v. 67, n. 1, p. 445–69, 2005.
- MESSENGER, S. W.; FALKOWSKI, M. A.; GROBLEWSKI, G. E. Ca²⁺-regulated secretory granule exocytosis in pancreatic and parotid acinar cells. **Cell Calcium**, v. 55, n. 6, p. 369–375, 2014.
- MI, H.; THOMAS, P. PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools. In: [s.l: s.n.]. p. 123–140.
- NAJI, L.; PACHOLSKY, D.; ASPENSTRÖM, P. ARHGAP30 is a Wrch-1-interacting protein involved in actin dynamics and cell adhesion. **Biochemical and Biophysical Research Communications**, v. 409, n. 1, p. 96–102, 2011.
- O'BRIEN, K. P. Inparanoid: a comprehensive database of eukaryotic orthologs. **Nucleic Acids Research**, v. 33, n. Database issue, p. D476–D480, 17 dez. 2004.
- O'BYRNE, J. et al. The Human Bile Acid-CoA:Amino Acid N-Acyltransferase Functions in the Conjugation of Fatty Acids to Glycine. **Journal of Biological Chemistry**, v. 278, n. 36, p. 34237–34244, 5 set. 2003.
- OGATA, H. et al. **KEGG: Kyoto encyclopedia of genes and genomes** **Nucleic Acids Research**, 1999.
- ORDOÑEZ, G. R. et al. Loss of genes implicated in gastric function during platypus evolution. **Genome Biology**, v. 9, n. 5, p. R81, 2008.
- PAK, Y.; PHAM, N.; ROTIN, D. Direct binding of the beta1 adrenergic receptor to the cyclic AMP-dependent guanine nucleotide exchange factor CNrasGEF leads to Ras activation. **Molecular and cellular biology**, v. 22, n. 22, p. 7942–7952, 2002.
- PANULA, P. et al. International Union of Basic and Clinical Pharmacology. XCVIII. Histamine Receptors. **Pharmacological Reviews**, v. 67, n. 3, p. 601–655, 17 jun. 2015.
- PETERSEN, O. H.; TEPIKIN, A. V. Polarized Calcium Signaling in Exocrine Gland Cells. **Annual Review of Physiology**, v. 70, n. 1, p. 273–299, 2008.
- POLLEY, S. et al. Evolution of the rapidly mutating human salivary agglutinin gene (*DMBT1*) and population subsistence strategy. **Proceedings of the National Academy of Sciences**, v. 112, n. 16, p. 5105–5110, 2015.
- PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. **Molecular Biology and Evolution**, v. 26, n. 7, p. 1641–1650, 2009.

- PROCTOR, G. B. The physiology of salivary secretion. **Periodontology** 2000, v. 70, n. 1, p. 11–25, 2016.
- PROCTOR, G. B.; CARPENTER, G. H. Regulation of salivary gland function by autonomic nerves. **Autonomic Neuroscience: Basic and Clinical**, v. 133, n. 1, p. 3–18, 2007.
- PROCTOR, G. B.; CARPENTER, G. H. Salivary Secretion: Mechanism and Neural Regulation. In: **Monographs in oral science**. [s.l: s.n.]. v. 24p. 14–29.
- RAH, S. Y. et al. Association of CD38 with nonmuscle myosin heavy chain IIA and Lck is essential for the internalization and activation of CD38. **Journal of Biological Chemistry**, v. 282, n. 8, p. 5653–5660, 2007.
- RIBEIRO, H. DE A. L. **Desenvolvimento de um serviço de análise de sequências utilizando um modelo baseado em atributos de resultados de PSI-BLAST**. [s.l.] UFMG, 2013.
- SAKAMOTO, T. **Ferramentas para análise filogenética e de distribuição taxonômica de genes ortólogos**. [s.l.] UFMG, 2016.
- SCHUBERT, M. L. Gastric secretion. **Current Opinion in Gastroenterology**, v. 26, n. 6, p. 598–603, 2010.
- SCHUBERT, M. L. Gastric physiology. In: **Gastrointestinal Anatomy and Physiology: The Essentials**. [s.l: s.n.]. p. 58–77.
- SCHUBERT, M. L. Gastric acid secretion. **Current Opinion in Gastroenterology**, v. 32, n. 6, p. 452–460, 2016.
- SCHUBERT, M. L.; PEURA, D. A. Control of Gastric Acid Secretion in Health and Disease. **Gastroenterology**, v. 134, n. 7, p. 1842–1860, 2008.
- SHARPE, L. J.; BROWN, A. J. Controlling cholesterol synthesis beyond 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR). **Journal of Biological Chemistry**, v. 288, n. 26, p. 18707–18715, 2013.
- SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular systems biology**, v. 7, n. 1, p. 539, 2011.
- SIEVERS, F.; HIGGINS, D. G. Clustal Omega. **Current Protocols in Bioinformatics**, v. 2014, n. December, p. 3.13.1-3.13.16, 2014.
- SMITH, D. M. et al. Evolutionary relationships between the amphibian, avian, and mammalian stomachs. **Evolution and Development**, v. 2, n. 6, p. 348–359, nov. 2000.
- SZABÓ, A. et al. Complex formation of human proelastases with procarboxypeptidases A1 and A2. **Journal of Biological Chemistry**, v. 291, n. 34, p. 17706–17716, 2016.
- TAKEHARA, S. et al. Degradation of MUC7 and MUC5B in Human Saliva. **PLoS ONE**, v. 8, n. 7, p. 1–9, 2013.

- TAKEUCHI, K. et al. Activation of Muscarinic Acetylcholine Receptor Subtype 4 Is Essential for Cholinergic Stimulation of Gastric Acid Secretion: Relation to D Cell/Somatostatin. **Frontiers in Pharmacology**, v. 7, n. AUG, p. 1–11, 30 ago. 2016.
- TAKIKAWA, H. Hepatobiliary transport of bile acids and organic anions. **Journal of Hepato-Biliary-Pancreatic Surgery**, v. 9, n. 4, p. 443–447, 2002.
- THOMAS, C. et al. Targeting bile-acid signalling for metabolic diseases. **Nature Reviews Drug Discovery**, v. 7, n. 8, p. 678–693, 2008.
- TRAUNER, M. et al. Molecular regulation of hepatobiliary transport systems: Clinical implications for understanding and treating cholestasis. **Journal of Clinical Gastroenterology**, v. 39, n. 4 SUPPL., p. S111–S124, 2005.
- TRAUNER, M.; BOYER, J. L. Bile Salt Transporters: Molecular Characterization, Function, and Regulation. **Physiological Reviews**, v. 83, n. 2, p. 633–671, 2003.
- TRIPATHI, S. et al. The gastrin and cholecystokinin receptors mediated signaling network: a scaffold for data analysis and new hypotheses on regulatory mechanisms. **BMC systems biology**, v. 9, p. 40, 2015.
- TURNER, R. J.; SUGIYA, H. Understanding salivary fluid and protein secretion. **Oral diseases**, v. 8, n. 1, p. 3–11, 2002.
- VAN BENNEKUM, A. M. et al. Hydrolysis of retinyl esters by pancreatic triglyceride lipase. **Biochemistry**, v. 39, n. 16, p. 4900–4906, 2000.
- VAN IERSEL, M. P. et al. Presenting and exploring biological pathways with PathVisio. **BMC Bioinformatics**, v. 9, p. 399, 2008.
- VIALLE, R. A. **Evidências de mudanças estruturais proteicas em transições macroevolutivas**. [s.l.] UFMG, 2017.
- VILELLA, A. J. et al. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. **Genome Research**, v. 19, n. 2, p. 327–335, 3 dez. 2008.
- WATERHOUSE, R. M. et al. OrthoDB: A hierarchical catalog of animal, fungal and bacterial orthologs. **Nucleic Acids Research**, v. 41, n. D1, p. 358–365, 2013.
- WHITCOMB, D. C.; LOWE, M. E. Human pancreatic digestive enzymes. **Digestive diseases and sciences**, v. 52, n. 1, p. 1–17, 2007.
- WILLIAMS, J. A. Regulation of pancreatic acinar cell function. **Current opinion in gastroenterology**, v. 22, n. 5, p. 498–504, 2006.
- WILLIAMS, J. A. Regulation of acinar cell function in the pancreas. **Current Opinion in Gastroenterology**, v. 26, n. 5, p. 478–483, 2010.
- XU, D. et al. Recent evolution of the salivary mucin MUC7. **Scientific Reports**, v. 6, n. 1, p. 31791, 2016.

YAO, X.; FORTE, J. G. Cell Biology of Acid Secretion by the Parietal Cell. **Annual Review of Physiology**, v. 65, n. 1, p. 103–131, 2003.

YONGE, C. M. EVOLUTION AND ADAPTATION IN THE DIGESTIVE SYSTEM OF THE METAZOA. **Biological Reviews**, v. 12, n. 1, p. 87–114, jan. 1937.

ZELCER, N. et al. Ubiquitination of the LDL Receptor. **Life Sciences**, v. 325, n. July, p. 100–104, 2009.

ZHAO, B.; NATARAJAN, R.; GHOSH, S. Human liver cholesteryl ester hydrolase: cloning, molecular characterization, and role in cellular cholesterol homeostasis. **Physiological genomics**, v. 23, n. 3, p. 304–10, 2005.

ZHAO, B.; SONG, J.; GHOSH, S. Hepatic overexpression of cholesteryl ester hydrolase enhances cholesterol elimination and in vivo reverse cholesterol transport. **Journal of lipid research**, v. 49, n. 10, p. 2212–7, 2008.

ZHONG, C. et al. HMGCR is necessary for the tumorigenicity of esophageal squamous cell carcinoma and is regulated by Myc. **Tumor Biology**, v. 35, n. 5, p. 4123–4129, 2014.