Original Article

Indian Journal of Genetics and Molecular Research
Volume 13 Number 2, July - December 2024
DOI: http://dx.doi.org/10.21088/ijgmr.2319.4782.13124.2

Low Molecular weight Proteins in Eukaryotic Genomes that are Encoded by Smorf, with Particular Emphasis on Insect Reproductive Proteins

Kiran Kumar Halagur Bhogegowda[1], Kiran Kumar D.J[2], Sajida Niketh[3], Sowmyashree B.S.[4]

How to cite this article:
Kiran Kumar Halagur Bhogegowda, Kiran Kumar D.J, Sajida Niketh et al. Low Molecular weight Proteins in Eukaryotic Genomes that are Encoded by smorf, with Particular Emphasis on Insect Reproductive Proteins. Ind J Genet Mol Res. 2024; 13(1):69-84.

Abstract

The majority of non-repetitive DNA in higher eukaryote genomes is transcribed, but only a small portion of them code for proteins, which appears to be a conundrum. The term "smORF-encoded proteins" (SEPs) refers to small open reading frames (smORFs) that encode functional proteins and are located across the genome in various locations, such as coding sequences (CDSs) and the 5' and 3' untranslated regions (UTRs) of genes. Numerous vital processes, such as metabolic, cellular, gene expression, embryogenesis and development are carried out by SEPs. Additionally, they are connected to a number of clinical conditions promote proteome evolution. As omics technology and identification techniques progress, a large number of peptides are currently being annotated. Further, several of these peptides are essential for the regulation of reproductive physiology and development in insects. Male seminal fluid proteins (SFPs) elicit a physiological response in the females during mating that increases the reproductive success of the mating pair. This response is known as the post-mating response (PMR). SFPs exhibit a broad distribution of relative molecular weights and influence a variety of functions, including catalysis, inhibitor activity, and protein folding and binding. It is now known that several of these proteins, such as MSAmiP in Drosophila, are crucial to insect reproduction. The area of smorf is reviewed in this mini-review, covering basic biology, mechanisms and functions, methodology and advancements finally evolutionary trends. Further expansion of the discussion to their involvement in insect reproductive physiology is also included.

Keywords: Small open reading frames (smORFs); post-mating response (PMR); evolutionary-constrained domain residues/ regions (ECRs).

Author Affiliation: [1]Former Post-doc NCBS, [2,3]Associate Professor, [4]Assistant Professor, [2,4]Department of Biotechnology, [3]Department of Chemistry and Biochemistry Government Science College, (Nrupathunga University), Bengaluru 560001, India.

Corresponding Author: Kiran Kumar Halagur Bhogegowda, Former Post-Doc NCBS, Government Science College, (Nrupathunga University), Bengaluru 560001, India.

E-mail: kirankumarhb@gmail.com

Received on: 22.10.2024

Accepted on: 25.11.2024

INTRODUCTION

The majority of non-repetitive DNA in the genomes of higher eukaryotes has been found to be transcribed, but only a small percentage of these transcripts are known to code for proteins. While there are a few of functional long non-coding RNAs (lncRNAs), it is still unclear if the majorities of lncRNAs serves any purpose or are just transcriptional noise (Chris P Ponting et al. 2009). Because they lack large open reading frames (ORFs), these are referred to as non-coding RNAs (lnRNAs). However, new research has revealed that unique sequences known as short open reading frames (smORFs) generate small proteins or peptides with significant biological roles (Clement Immarigeon et al., 2021; Emmanuel Ladoukakis et al., 2011; Alvaro de Andres-Pablo et al., 2017). Based on transcriptional characteristics Diego Guerra-Almeida et al. 2021 classify smORFs into two groups: expressed (genic) smORFs and non-expressed (intergenic) smORFs. These ORFs encode proteins that are 100 amino acids or less in length. Alternate splicing of numerous non-coding genes results in noncoding RNA variations known as circRNAs and lncRNAs, which serve regulatory purposes in addition to their typical protein-coding functions. These proteins are encoded in the genomes of many different organisms, such as yeast, bacteria, plants, fruit flies, and other metazoans (Ruth Steinberg and Hans-Georg Koch 2021). Many mitochondrial-encoded smORFs play important roles in respiration of cells, and some are implicated in stress signaling (Catherine A Makarewich et al., 2020) and in cancer (Nausica Arnoult et al., 2017) such as the mitochondrial outer membrane (MOM) microprotein (PIGBOS protein). The first functional microprotein discovered was a helix-loop-helix protein in 1990 (Y Wang, et al., 1992) known as the "Id" protein with functions in muscle differentiation, whereas in plants the Little Zipper (ZPR) proteins which control stem cell during development were the first (Kaushal Kumar Bhati, et al., 2018), in yeast the *NORF5/HUG1 protein (*Munira A Basrai et al., 1999) *and in bacteria the* Bacitracin produced by Bacillus licheniformis (Jiang Zhu, et al., 2023).Table-1 summarizes diverse class of smORFs identified. Subsequent to these several peptides have been discovered and annotated to date.

Table 1: List of diverse class of smORFs identified in various taxa

smORF peptides/ example	Species	Functions	Reference
Hemotin	Homo sapiens	Phagocytosis	José I. Pueyo, 2016
MOTS-c	Homo sapiens	Regulator of metabolic homeostasis	Kim et al., 2018
MOTS-c	Mice	Insulin resistance and diet-induced obesity and reverses age-related insulin resistance in muscles	Lee et al., 2015
pri-miR171b and pri-miR165a	Medicago truncatula	Root development	Lauressergues et al., 2015
Polaris	A. thaliana	Hormonal crosstalk during	J. Liu et al., 2013
Myoregulin	Caenorhabditis elegans	Muscle development	Douglas M. Anderson et al., 2015
sarcolamban	Drosophila	SER Ca2+ trafficking	E.G. Magny et al., 2013
Pri	Drosophila	Leg developme nt	Markus et al., 2023
SgrS	Escherichia coli	Glucose flux	Wadler, C.S. and Vanderpool, C.K. 2007
SR1P	Bacillus subtilis	Moonlighting activity modulation	Gimpel et al., 2016
hld	Staphylococcus aureus	Antimicrobial activity	Verdon et al., 2009

The functions of these proteins are varied and include metabolic functions like respiration (Annie Rathore et al., 2018), tissue folding (Yoo-Seok Hwang et al., 2013), and homeostasis; cellular functions like phagocytosis (José I Pueyo, et al., 2016), DNA expression and repair (Sarah A Slavoff, et al., 2014), mitochondrial translation (Nadia G D'Lima, et al., 2017), and cell-to-cell signaling during development (José I Pueyo, et al., 2008), and embryogenesis (Andrea Pauli, et al., 2014).

These lnCRNAs and short regulatory RNAs have numerous unidentified uses that are continually being found. Compared to the overall number of transcripts produced in genomes, the role of non-coding transcript variations, such as lncRNAs and circular RNAs (circRNAs), expressed from peptide-coding genes are extremely modest; extensive noncoding transcriptomes and differential splicing contribute significantly to organismal complexity (Sonam Dhamija & Manoj B. Menon et al., 2018).

Furthermore, other regulatory mechanisms, such as alternative transcription initiation, alternative polyadenylation, and alternative CDS translation, could generate isoformic smORFs (Diego Guerra-Almeida et al., 2021).

Intergenic and genic smorf

Intergenic smORFs

Intergenic smORFs are not generally functional or transcribed; they constitute majority of smORFs in the genome and originate randomly via simple arrangement and rearrangement of nucleotides often classified as junk DNA (Couso, J. and Patraquim, P. 2017). Previously several coding smORFs were overlooked in genome annotation screenings and classified as intergenic DNA stretches (Ladoukakis, E., 2011). Recent evidence from bacteria suggests that intergenic smORFs are expressed under differential physiological conditions indicating that intergenic smORFs are a potential source of peptides with condition-dependent expression (Hu¨cker, S.M et al., 2017). Strategies to uncover new intergenic smORFs includes smORF detection in intergenic regions using bioinformatic tools, followed by filters of evolutionary conservation and comparison to RNA sequence databases to verify if the sequence is expressed.

Genic smorf

smORFs in small RNAs

The regulatory roles of small RNAs are widely studied in all domains of life owing to their cellular responses to biotic and abiotic stresses. The classification of genic smorf is depicted in fig. 1. Small RNAs are transcripts smaller than 200- 300 nucleotides. However, the coding capacity of small RNAs has been scarce with few reports of functional smORFs (Friedman, R.C. et al., 2017). Evidence suggests that most coding small RNAs are dual functional. The coding capacity of eukaryotic small RNAs is still unexplored. Whereas, in prokaryotes; the widespread distribution of potentially coding small RNAs has been confirmed by computational approaches. It is shown in 14 bacterial species that 0.5% of all small RNAs contain smORFs are under purifying selection, and reaches 20% in some taxa. An example of smorf small RNA is the Staphylococcus aureus RNAIII encodes the smORF peptide d-haemolysin which regulates the translation and/or stability of virulence factors, cell wall metabolism enzymes and transcription factor mRNAs (Bronesky, D., 2016).

smORFs in long non-coding RNAs

Typically lncRNAs are expressed at low levels and are non-conserved. These molecules are longer than 200 nucleotides and associated with several biological processes (Yuji Kageyama et al., 2011). Due to certain limitations the coding potential of lncRNAs is still disregarded including (i) the absence of major ORFs (ii) degeneration frequency similar to that of introns and much higher than that of exons and (iii) when translated, the instability and rapid degradation. Examples of lncRNAs are the cis-acting regulatory, pri-miR171b and pri-miR165a transcripts, which are miRNA precursors, encode peptides of 18 and 21 amino acids, respectively. Both peptides promote negative modulation of their targets associated with root development in Medicago truncatula and A. thaliana (Lauressergues D., et al., 2015). For many years, circRNAs have been considered atypical products in cells; however, recent reports indicate these molecules are stable and are generated via a mechanism known as back-splicing. An example is the synaptic circMbl1, which encodes a peptide of 10kDa in response to starvation and FOXO expression involved in the mechanisms of neuronal communication (Pamudurti, N.R.,2017).

smORFs as reference CDSs

Reference CDS is the main or unique coding ORF of an mRNA, which can be flanked by translatable alternative ORFs. These molecules are smaller than 100 codons. Eukaryotic mRNAs exhibit more than one small ref CDS, constituting poly-cistronic transcripts which might even contain conserved and duplicated smORFs (Pauli, A., 2015). Strategy based on ribosome profiling and mass spectrometry and analysis of conservation at the amino acid level among different species coupled to transcript and peptide localization techniques are promising approaches for their identification. Example of small refCDS transcripts is the 55 amino acid peptide toddler/apela/ELABELA/ende conserved among vertebrates that acts as an embryonic signal in mesendodermal cell migration (Chu, Q. et al., 2015).

SmORFs as isoforms of major ORFs (isoformic smORFs)

Isoform smORFs are small variants of major ORFs. The main mechanism of generation of isoform smORF is alternative splicing. Further, other regulatory mechanisms, such as alternative transcription initiation, polyadenylation and ref CDS translation could theoretically generate isoform smORFs (Sasaki, K. et al., 2004). Vertebrates are potentially the best models for the discovery of isoform sorts since the genomes are regulated by alternative splicing. Since the coding genes of invertebrates and vertebrates (e.g. 20,000 in humans versus 20,000 genes in Caenorhabditis elegans) are approximately similar hence, evolution differentiation and complexity may be linked to the diversity of variants produced. Further, complications include the transcriptome sequencing limitations such as de novo assemblies. Finally, overlap between smORFs and their major variants is another difficulty. Example of this class of smorf is the tumor suppressor TEL (also known as ETV6) belongs to the ETS transcription factor family. Exon deletion generates five alternative splicing variants of these two isoforms encode the TEL-c and TEL-d (20 to 30 amino acids), which lack the activity of their full-length variant suppressors exhibit considerable increased expression in myelodysplastic syndrome-derived leukemia (Ohno, S. 1984).

SmORFs as alternative CDSs in mRNAs and classes

In early genomics research, the 'one gene, one protein' dogma was confronted by the capacity of mRNAs to encode more than one protein from alternative ORFs. Alternative smORFs are alternatively or pervasively translated from the same mRNAs as canonical large CDSs (Klemke, M., H. 2001). Emerging evidence suggests that alternative ORF translation is also an important biological pathway contributing to genetic variability, either from overlapping genes or alternative ORFs in unique transcripts (Vanderperre, B et a., 2013). Proteomic and transcriptomic approaches corroborate the existence of putative alternative smORFs in the flanking regions of or overlapping with ref CDSs in different frames. Data suggest that 15% of alternative smORFs in humans are preceded by Kozak consensus sequences, suggesting that smORFs can show efficiency in ribosome recognition. Comparative studies implicate hundreds of in human alternative smORFs with evidence of translation are highly conserved in distant taxa from basal vertebrates to invertebrates and yeast. Upstream ORFs (uORFs) are smORFs located in 5'UTRs and have been reported in 50% of eukaryotic mRNAs. uORFs play roles in several post-transcriptional control mechanisms(Kochetov, A.V. 2008). uORFs perform their functions through both cis and trans-regulatory mechanisms. Example of uORF is the 5'UTR of the c-akt proto-oncogene transcript which displays a uORF encoding a 10 amino acid tumor rejection antigen (pRL1) (Uenaka, A et al., 1994). Overlapping smORFs are smORFs that overlap with the refCDS in another frame. The start codons of overlapping ORFs strictly overlap with ref CDSs and can extend to the 3'UTR (Michel, A.M et al., 2012). Unlike uORFs majority of overlapping ORFs are smORFs. These smorfs are widely expressed suggesting their diverse roles. Example of this class of smorf is the gp75 melanoma antigen transcript that contains an overlapping smORF encoding a 24 amino acid peptide identified as a tumor rejection antigen (Wang, R.F. et al., 1996). Downstream smORFs are ORFs located in 3'UTRs. The coding potential of downstream smORFs is underestimated and poorly explored compared to that of other alternative smORFs (Mehta, A. et al., 2006). These smorfs contain translational regulatory elements and subcellular localization signals. Further contribute to eukaryotic transcript stability, tissue patterning, embryonic axis formation and mammalian spermatogenesis. The H60 histo-compatibility gene encoding the eight amino acid antigen LYL8, which is presented by MHC class I to the immune system is an example of this type of smorf(Malarkannan et al., 1998).

In summary, integration between omics sciences and bioinformatics has enabled biological investigation, discovery and annotation of several classes of smorf. This enterprise will help elucidate diversity and singularities of smorf and their role in evolutionary innovations in the hidden coding DNA world. Hundreds of new players are been discovered with diverse roles in transcription/translation control, tissue physiology, immunological surveillance, developmental biology, cancer, neuro-pathologies, and environmental responses. Such diverse biological roles implicate their important roles previously underestimated.

Synthesis, mechanisms and functions of smorf

There is compelling evidence for diverse RNA molecules involved in translation of peptide smORFs such as mRNAs (K. Bharathan Sruthi et al., 2022), introns of pre-mRNAs (Marta A Inchingolo 2023), lncRNAs (Pedro Patraquim et al., 2022), and even primary transcripts of miRNAs (Dominique Lauressergues et al., 2015) or rRNAs (Julie L Aspden et al., 2016). The diversity is summarized in Fig. 1. Cells employ several mechanisms to generate smorfs, which are based on the central dogma maxim that a gene with multiple exons can be transcribed and spliced to generate an mRNA, which is translated to a protein, determining its cellular function. However, in some cases, the mRNAs can additionally have regulatory functions independent of translation. These protein-coding RNAs are classified as cncRNAs (coding/non-coding RNAs). Further, the same gene can also generate long non-coding transcripts by alternative splicing. These lncRNAs can be genuine "non-coding RNAs" with regulatory functions, or they can harbor smORFs (small ORFs), which encode micro-peptides for example the TERRA stabilizer protein enables interaction of proteins with telomeres (Bersabe Wondimagegnhu et al., 2024). Finally, back-splicing can also generate circRNAs (circular RNAs) from several genes.

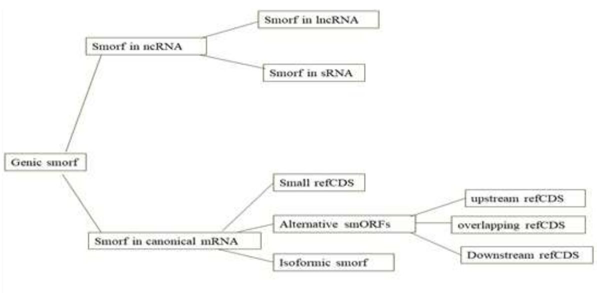
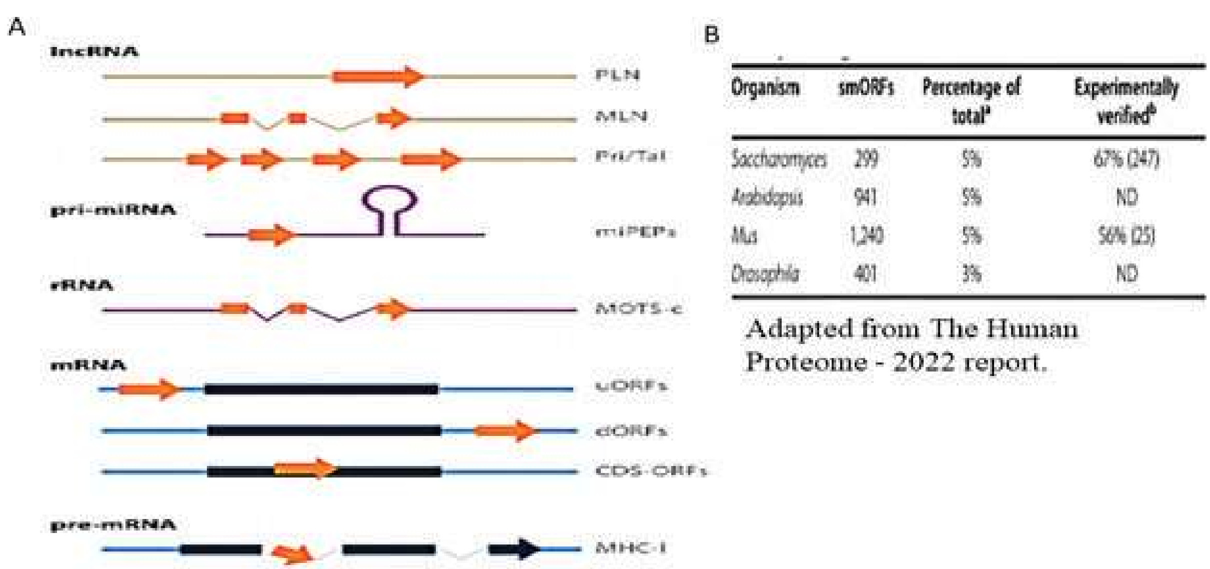


Fig. 1: Classification of Genie smorf

Recent research indicates that at least some circRNAs may be translated, despite the fact that they are typically stable non-coding RNAs with regulatory roles (reviewed in Sonam Dhamija & Manoj B. Menon 2018). Other mechanisms consist of Isoform smORFs are generated via large ORF transcriptional editing, which fragments large CDSs into smaller variants that can fall within the smorf length limits. Alternative transcription initiation via alternative promoters, alternative polyadenylation cleavage, alternative CDS translation mediated by downstream start codons, and the fragmentation of homologous pseudogenes (reviewed in Diego Guerra-Almeida et al., 2021).

Example includes the Sp1 zinc-finger protein that binds to the 3′-UTR of mRNAs (Jingwen Song 2022). smorf peptides are transcribed in the nucleus and can play regulatory roles within and outside the cell. Regulatory mechanism (function) of micro peptides consists of pepto that inhibit downstream coding sequence expressions by blocking ribosomes through direct or indirect small molecular activation switches. Absence of the N-terminal signal sequence in the micro peptides is suggestive of their location in the cytoplasm. Abundance of amino acids reported in Drosophila smorfs are phe, lys, arg, and ser respectively. Finally, the relative small size confers an essential role to these proteins such as regulator of several important biological functions (Keira R Hassel, et al., 2023). Recent evidence indicates that a large number of lncRNAs harbor smORFs and encode short micro peptides. The distinction between coding and noncoding transcriptomes is slowly being narrowed by the emergence of micro peptide-encoding smORFs. Fig. 2 am summarizes the mechanism involved in smorf generation and the diversity of RNA molecules involved.



Organism	smORFs	Percentage of total[a]	Experimentally verified[b]
Saccharomyces	299	5%	67% (247)
Arabidopsis	941	5%	ND
Mus	1,240	5%	56% (25)
Drosophila	401	3%	ND

Fig. 2: A Various RNA classes involved in generating smorf (Adapted from Diego Guerra-Almeida 2021)
B. Table of smORFs identified in various species

Numerous studies have documented the similarity of uORFs in the 5′mRNA leader sequence and their impact on translational regulation in the transcriptional machinery of eukaryotes (reviewed in Luis Enrique Cabrera-Quio et al., 2016). By adversely affecting the efficiency of protein translation, they impose this regulation, which is primarily responsible for differentiation and development. Next, a number of smorfs function as extracellular signaling molecules that resemble hormones and play important functions in the development of vertebrates and cellular stress (Rui Vitorino 2024). Numerous SEPs have been discovered to favorably or negatively regulate the calcium pump known as SERCA, which in turn controls the development and function of muscles (Emile G Magny et al., 2013).These hydrophobic peptides have properties that facilitate interactions with membranes. Additionally, studies smorf them to the function of cellular organelle's such as P-bodies involved in a variety of RNA metabolism processes, as well as endosomes and lysosomes (Akinobu Matsumoto, et al., 2017). DNA repair and transcription control are two examples of the nuclear roles of small peptides. Dysregulation of these proteins has been linked to a number of pathological conditions, including diabetes (Sheng Xu 2018), amyotrophic lateral sclerosis (ALS) (Yanan Pang et al., 2018), chagas disease (Carina Azevedo Oliveira Silva et al., 2024), cancer (Iűaki Merino-Valverde et al., 2020), and cardiovascular illnesses (Jia-Rui You et al., 2024). In conclusion, several lines of research demonstrate the significance of these micro peptides in various life forms and their varied functions, which include cellular, metabolic, and physiological functions. Further their association to number of illnesses and roles in genome control underscores their importance.

Evolutionary insights into smORF

Several genomes have 75–90% junk DNA, which is made up of non-coding RNAs from extensive transcription (Nils G Walter 2024; ENCODE Project Consortium, 2012). Additionally, they provide neutrally evolving smorf peptides. Genetic drift causes the DNA elements to accumulate mutations, and thus neutral and non-adaptive evolution occurs (Matthew Putnins and Ioannis P Androulakis 2021). "Selected effects" and "causal roles" are two theories put forth to explain the evolution of junk DNA (Doolittle and Brunet, 2017). The activity carried

Table 2a: Internet resources (databases/tools/prediction methods of smorf)

Repository	Type of smorf	Brief desctiption	url/Reference
sORFs.org	sORF repository	Small peptide validation	http://www.sorfs.org/ Olexiouk V et al., 2018
SmProt	sORF repository	Integrates conservation analyses	http:// bioin fo. ibp. ac. cn/ SmProt/ Hao Y et al., 2017
OpenProt	altORF resource	Protein isoforms. conservation analyses and functional domains	https:// www. openp rot. org/ Brunet MA et al., 2021
ARA-PEPs	sORF repository	Putative sORF-encoded peptides specifically in Arabidopsis thaliana	http:// www. biw. kuleu ven. be/ CSB/ ARA- PEPs Hazarika RR et al., 2017
PsORF	sORF repository	sORF across different plant species,	http:// psorf. whu. edu. cn/\ Chen Y et al., 2017
MetamORF	sORF repository	Unique sORFs in H. sapiens and M. musculus genomes	http:// metam orf. hb. univ- amu. fr/ Choteau SA, et al., 2021
nORFs.org	novel ORF (nORF) repository	Aggregated information	https:// norfs. org/ Neville MDC, et al., 2021

Table 2b: Internet resources (prediction methods)

Tool	Brief description	url	Reference
Coding NonCoding Identifying Tool (CNIT)	Distinguishes between coding and non-coding regions based on intrinsic sequence compositions	http://cnit.noncode.org/CNIT/	Guo JC et al., 2019
Coding Region Identification Tool Invoking Comparative Analysis (CRITICA)	Analyses nucleotide sequence composition and conservation at the amino acid level	http:// rdpwww. life. uiuc. edu/	Badger JH et al.,1999

Coding Potential Calculator (CPC)/ CPC2	Assess protein-coding potential based on important features(size, coverage, integrity)	http://cpc.cbi.pku.edu.cn/ http://cpc2.gao-lab.org/	Kong I et al.,2007 Kang Y J et al., 2017
Coding Potential Predictor (CPPred)	Predicts the coding potential of RNA transcript	http:// www. rnabi nding. com/ CPPred/	Tong X etal., 2019
PhyloCSF	Phylo CSF Determines a conserved protein-coding region based on formal statistical comparison of phylogenetic codon models	http:// compb io. mit. edu/	Lin M F et al., 2011
MicroPeptide Tool (MiPepid)	Identifies coding sORFs based on existing microproteins subpopulation set	https:// github. com/ MindAI/ MiPep id	Zhu M et al., 2019
sORF Finder	Identifies sORF with high coding potential based on nucleotide composition bias and potential functional constraint at the amino acid level	http:// evolver. psc. riken. jp/	Handa K at al., 2010
smORFunction	Provides function prediction of sORFs/ micro-proteins	https:// www. cuilab. cn/ smorf function/ home	Ji X et al., 2020
miPFinder	Identifies and evaluates microproteins func tionality using information on size, domain, protein interactions and evolutionary origin	https:// github. com/ DaStr aub/ miPFinder	Straub D et al., 2021
PhastCons	Based on conservation scoring and identifica tion of conserved elements	http:// compg en. cshl. edu/ phast/	Siepel A et al., 2005

out by a neutrally evolving element by chance is referred to as the "causal role." Natural selection based adaptive phenotypes are acquired through "selected effects." Under environmental stress, neutrally evolving smorf peptides may shift from "causal roles" to "selected effects," exposing their neutral phenotypes to natural selection and driving the emergence of novel coding genes. Few examples are the mitochondrial peptide code LINC00116 (Anastasia Chugunova et al., 2019) and the Streptomyces coelicolor Small ORF trpM (Alberto Vassallo et al., 2020), which accelerates growth and morphogenesis. Overlapping the reference CDS of canonical mRNAs or using alternate smorfs in UTRs are two other possible pathways. Here, smorfs are translated widely or by modifications that cause the cis-regulatory elements to be induced (Benoît Vanderperre et al., 2013). The alternative smorfs would either be fixed as alternative smorfs in the original transcripts or establish independent gene units by retro transposition if the "causal roles" carried out by neutrally developing smorf peptides become "selected effects" (Zhe Ji, et al., 2015). Proteins are synthesized by ribosomes through a process called translation. Several small sequences that were previously believed to be noncoding are pervasively translated, such as the Saccharomyces cerevisiae translatome employing ribosome profiling, iRibo, and high-powered selection inferences (Aaron Wacholder et al., 2023). Suggesting their contribution to the advantages of evolutionary fitness. Therefore, the study of smorf provides an unparalleled framework for comprehending the significance of genetic novelty in the creation of novel protein structures, functions, and phenotypes at many scales, ranging from the DNA sequence to integration into cellular networks.

METHODS OF DETECTION OF SMORF

Bioinformatic predictions of sORFs prove valuable due to lower cost. Table 2 a,b. summarizes the internet resources (databases/tools/prediction methods) for the smorf. The need to re-evaluate the coding potentials of sORFs entails use of several methods based on computational analysis, next-generation sequencing (NGS) and mass spectrometry (MS) and to identify the smORFs or their products. Further, difficulties include the smaller sizes (< 100 amino acids; < 20 kDa) which necessitate the adaptation of existing laboratory techniques. In the following paragraph few methods are briefly explained.

Ribosome profiling (RIBOSeq)-although the coding potentials of sORFs can be predicted via computational tools, this does not provide sufficient evidence about the transcription and translation locus of smORFs. Also the lack of evolutionary constraint renders the identification difficult. These short-comings can be overcome through ribosome

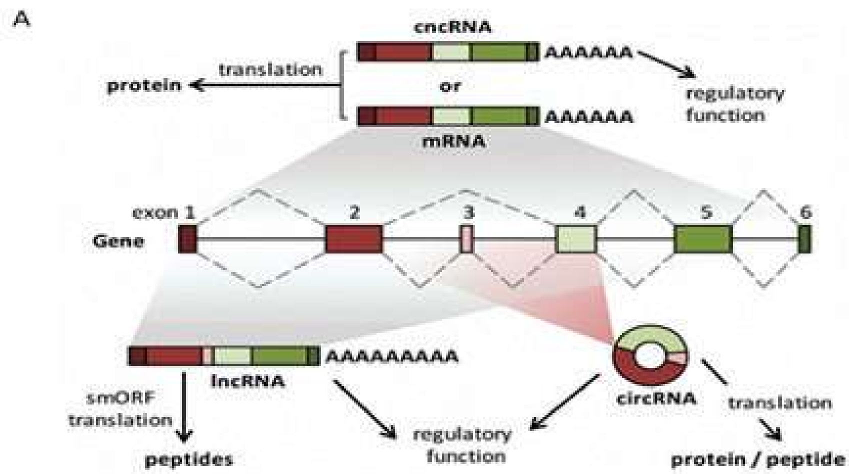


Fig. 3: A. Various transcription mechanisms generating small peptides (Adapted from- Serge Plaza 2017)

profiling (RIBO-seq) techniques(Mudge JM et al., 2021; McLysaght A, et al. 2016). RIBO-Seq has been applied to identify novel sORFs and to explore the protein-coding potentials of RNAs in various models organism. RIBO-seq is based on the principle of–Translatomics which involves measurement of cellular translational activity, which provides a snapshot of the translation process and regulation. It is a NGS based tool for measuring ribosome-protected mRNA fragments developed by Ingolia et al. (2009).Advantage of this method includes parallel isolation of multiple clusters of ribosomes-Poly-RIBO-Seq (Aspden et al. 2014).

Mass spectrometry (MS)-Based Proteomics

MS-based proteomics remains indispensable in proteomics for sequencing and quantifying proteins and peptides. MS strategy is well suited in smORF research since it involves direct analyses of microproteins with prior optimization and modifications. Other advantage of the method is the assessment phosphorylation of microproteins (López E et al., 2012). Size exclusion filter, acid precipitation and C8 reverse phase solid phase extraction and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) are methods

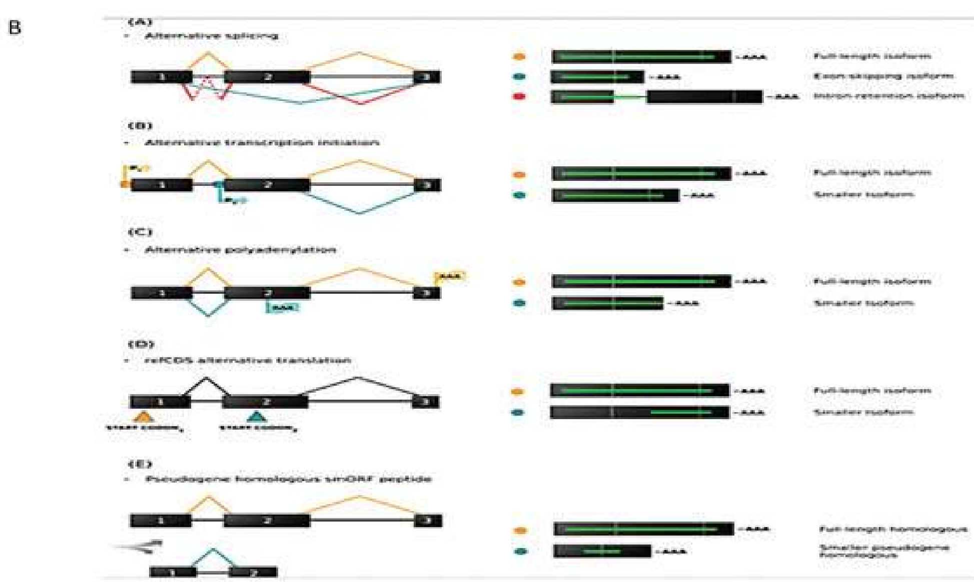


Fig. 3: B. Splicing, translation and pseudogene mediated mechanisms of generating small peptides (Adapated from Diego Guerra-Almeida 2021)

used to detect low-molecular-weight peptides.

Proteogenomics Approach

Proteogenomics is a comprehensive approach where MS data is coupled to genomic, transcriptomic or translatomic data providing an alternative for further validation of low-abundant microproteins (Low TY and Heck AJ. 2016).

Validation of microproteins and their biological functions

Micro proteins exert their cellular functions either by forming a complex with larger canonical proteins, or by acting autonomously. Hence a prerequisite for validation is to understand their modes of actions. Whilst the aforementioned methods are efficient at detecting microproteins, a validation step is required to elucidate the exact biological functions. One of the strategies is to analyze differential expression data of the transcripts or proteins. Example includes the Cao et al. (2020) study of 16 potential microproteins from leukemia cell lines. Second approach is to focus on specific genomic regions which are likely to possess high coding potentials such as the mtDNA genomic region which encodes MOTS-c, a microprotein Lee, Kim and Cohen (2016). Lastly, Genome-wide CRISPR/Cas9-based screens enable to assessment of the influence of microproteins on the phenotype (Chen J, et al., 2020). For direct observation of the targeted microproteins, loss-of-function (e.g., knockdown, knockout) or gain-of-function (e.g., over expression or activation) assays are tailored to deduce the resulting phenotype. The inference of the mutant phenotype from the wildtype helps infer the role of the microproteins in a biological process, and at also the concentrations that produce the effects. Stein et al. (2018) validated the function of a 56-amino-acid microprotein mitoregulin (Mtln). Several drawbacks and limitations of these methods are discussed in details and several new methods developed. Further advances in detection methods and algorithms are improved over the years are aiding the field. For a detailed reading the reader is referred to Fabiola Valdivia-Francia and Ataman Sendoel 2024.

Table 3: Insect Low molecular weight proteins isolated from various insect species

Description/Mol wt (Kda)	Function (GO-ontology)/ Cellular location	References	Insect species/Method used to identify
Odorant receptor	Odorant-binding/ Secreted	Uniprot	Heliconius melpomene rosina/MS
Heat shock protein 70	ATP-dependent protein folding chaperone/ Endoplasmic reticulum lumen	Rama et al., 2024	H.armegeria/MS
Diazepam-binding inhibitor	Fatty-acyl-CoA binding/ Mitochondria	Rama et al., 2024	H.armegeria/MS
Elongation factor 1-alpha, partial	GTP binding/cytoplasm	BS Chaitra et al., 2020	H.armegeria/MS
Seminal fluid protein HACP001	Serine-type endopeptidase /Enzyme	James R. Walters and Richard G. Harrison 2000	Danaus plexippus plexippus/MS
Bombyxin B-12	Hormone/ Secreted	BS Chaitra et al., 2020	H.armegeria/2DE-MS
Tubulin polyglutamylase TTLL4-like	Tubulin-glutamic acid ligase activity/ Enzyme	Xia Xu et al., 2022	H.armegeria/2DE Bombyx mori
Initiatorin	Serine endopeptidase/Enzyme	Sakakura M et al., 2022	Bombyx mori/ MS
Major facilitator superfamily (MFS) profile domain-containing protein	transmembrane transporter activity/ Enzyme	Francesca Scolari et al., 2016	Glossina morsitans / MS
Tubulin beta chain	Nucleotide-binding/ Constituent of Cytoskeleton	Francesca Scolari et al., 2016	Glossina morsitans / Mass spectrometry (LC/MS/MS)
Cytosol aminopeptidase	Manganese ion binding/ Enzyme	Ethan C. Degner et al., 2019	Aedes aegypti/ Mass spectrometry
Adipokinetic hormone 1 preprohormone	Hormone/ Secreted	Ethan C. Degner et al., 2019	Aedes aegypti/ Mass spectrometry

A list of Low molecular weight proteins identified in various insects through various methods listed along with functions and localization is listed in table-3. As evident from the table the diverse lower molecular weight and cellular location and functions of proteins call for different detection methods. Although few proteins are common to several genera of insects and species, however majority are unique to species. This observation can have relevance to the roles of smorf in their evolutions, also in transcription and translation. As described in the preceding sections several adaptive and evolutionary causes are attributed to this uniqueness. It is possible that the smorf enable such changes through various genetic mechanisms.

The physiological roles of low molecular proteins in insect reproduction Peptides and proteins are among the many macromolecules found in seminal fluid, the non-sperm portion of the ejaculate. Many protein types, including lectins, prohormone precursors, proteases/protease inhibitors, peptides, and antioxidants, are included in SFPs (L A Herndon, M F Wolfner et al., 2015; K Ravi Ram and Mariana F Wolfner 2007; Frank W Avila et al., 2011). Several unique features of the protein structure explain their biochemical/physiological roles. The relative molecular weights vary from 36 to 200–400 amino acids. The proteins are members of a class that includes hormones, endopeptidases, chaperones, and enzymes. Additionally, a variety of activities such as catalysis, inhibitor activity, protein folding, and protein binding are modulated by the peptides. Their occurrence in diverse insect taxa is an indication of their conservation and important roles in gametogenesis and reproduction. These proteins impact several pre-and post-mating responses (PMR). Cooperation and conflict between the sexes are both important aspects of sexual reproduction. Significant selection pressure also determines the rapid development of SFPs (Chapman et al., 2008; Steven A Ramm et al., 2009; Laura K Sirot et al., 2014). In insects, this results in reproductive isolation across diverse taxa—a striking pattern (Jolie A. Carlisle and Willie J. Swanson 2021). Many fresh SFP from various insects belonging to various taxa are being added to the list as a result of improvements in omics techniques, improved algorithms, and improved annotation techniques. Since many SFP are low-molecular-weight proteins, their roles in reproduction raises evolutionary questions. Research reveals distinct domain residues/regions (ECRs) in the proteins that are evolutionary-constrained and suggestive of their functions in physiological pathways associated to reproduction. Evolutionary mechanisms acting on these proteins are suggested to include duplication and micro-evolution processes (Mrinalini et al., 2021; Seirian Sumner et al., 2023). The variety of insect antimicrobial peptides, small heat shock proteins (Hsp27), serine protease inhibitors (SGPI-2), and fibrinogen-related proteins (FREPs) are examples of low molecular proteins. Table-4 provides a summary of these proteins in a variety of insects. They carry out a variety of tasks, including defensive, cellular, immunological, and physiological responses. Further, few proteins aid cell defense mechanisms such as venoms (example ant myrmeciitoxins) proteins and few are anti-freeze proteins (example Tenebrio molitor from the beetle Tenebrio molitor). Additionally, these functions imply that they are changing in accordance with the evolutionary timeline.

Table 4: Various classes of Insect small molecular proteins with functions isolated from diverse insects

Class of insect smorf/example	Insect	Functions	Reference
Antimicrobial Peptides/ Defensins	Spodoptera frugiperda	Antibacterial against Gram-positive and few gram-negative bacteria	Hui-Yu Yi et al., 2014
Small heat shock proteins/ Hsp27	Drosophila	Cellular stress	Mohamed Taha Moutaoufik & Robert M. Tanguay et al., 2021
Odorant-binding proteins/ Obp19d	Drosophila	Odoront binding	Jennifer S. Sun et al., 2018
Serine protease inhibitors/ (SGPI-2)	*Schistocerca gregaria*	Iprotease inhibitor	Rose-Anne Boigegrain 2000
Fibrinogen-Related Proteins (FREPs),	Anopheles gambiae	Immune response	Iwona Wojda et al., 2020
Antifreeze	Tenebrio molitor	Antifreeze	Midya US and Bandyopadhyay S. 2024
Venom Protein/ω-conotoxin	Rhynocoris iracundus	Defense	Nicolai Rügen et al., 2021

Evolution implications of Insect Low molecular weight protein

The rates by which reproductive proteins evolve compared to others proteins vary in animal kingdom (Patlar B and Civetta A. 2022). Further unique features in insects such as variation within species (Martin D Garlovsky and Yasir H Ahmed-Braimah 2023), different evolutionary rates are gender specific (Papachristos K 2024) and cast specific (Brittany L. Enzmann and Peter Nonacs 2018) raises few questions about the de novo gene birth mechanisms also on the fine scale genomic regulation. Several neutrally evolving peptides are facilitated by the translation of open reading frames in non-coding genomics sequences (Ruiz-Orera J and Albà MM. 2018). Studies suggest through certain genome modifications such as optimal coding such as structural stabilization, emergence of Kozak consensus; internal ribosome entry sites (IRES), coverage by enhancers and the elongation of coding smORFs involve these gene birth processes(Juan-Pablo Couso Pedro Patraquim 2017). Since sequence conservation is common to several smorf this proposition seems a true possibility since during evolution they will be fixed.

New findings in both model and non-model insects are expanding the collection of insect reproductive peptides. Examples include the RNA-encoded "micro peptide" (MSAmiP) in Drosophila melanogaster, which consists of nine or twenty amino acids. It is a nested gene in the HOX cluster that is expressed in secondary cells of the male accessory gland. Further, loss of function of this micro peptide causes defects in sperm competition implicating its role in reproduction. Studies in Drosophila also suggest pattern of conservation of these smORFs, synteny and their ratio of conservative versus non-conservative nucleotide changes (Emmanuel Ladoukakis *et al.*, 2011). Linrong Wan *et al.*, 2002 identified in silkworm using comprehensive methods and identified several smORFs which were differentially and tissue specific expressed.

Diverse functions of these peptides in insects include the conserved peptides mlpt/Ubr3/Svb module, an ancient developmental switch involved in embryonic patterning (Azza Dib *et al.*, 2021), the Pri smorf peptides mediators of Ecdysone signaling, the glass locus which is involved in eye development (Cornelia Fritsch *et al.*, 2019), polished rice (Pri) which has roles in maintaining adult stem cell niches (Emile G. Magny *et al.*, 2013). Together, the study findings are pointers to their varied roles in the cellular and developmental processes of insects.

CONCLUSION

From a genomic perspective the non-coding transcripts are now proposed with consensus are missing link in explaining the complexity achieved by mammals, despite having similar number of protein-coding genes compared to invertebrates. Studies comparing smORFs at the base of metazoan and non-bilaterian groups, cnidarians, placozoans and ctenophores, would be particularly interesting. From an evolutionary perspective comparison of the roles and conservation of smORFs during speciation and/or whole-genome du plication events could provide insights into the origin and functions. Roles for lncRNAs in complex processes unique to higher eukaryotes like neuronal development and cancer metastasis support this notion. With mounting evidences of human homologs in basal vertebrates, invertebrates and yeast coupled with advances in gene knock-out and in methods and high-resolution detection methods the discovery biology of smorf it being fueled at rapid pace (Peter F Renz, *et al.*, 2020). Genetic variation in smORFs is now proposed to contribute to human traits and diseases implicating that these variations could be used as disease marker. Further, discovery of homologs and novel smorf from diverse life forms will add to this goal. From an application perspective several small peptides provide alternative opportunities for drug therapies. For example, Humanin derivatives in Alzheimer's disease and dementia (Chrysoula-Evangelia Karachaliou and Evangelia Livaniou 2023), Daptomycin Use in Pediatric Patients (Hanna Persha *et al.*, 2024), Liraglutide used to treat diabetics (Assim A. Alfadda, *et al.*, 2024). A growing number of peptides are now are being investigated as potential tumor biomarkers. With the increase in antibiotic resistance antimicrobial micropeptides provide alternative treatment regime.

Insect proteins are increasingly used as bioactive peptides and insects are emerging as alternative sources of proteins for use in food applications. Insect proteins such as anti-freeze proteins have diverse applications in several industries. Several plant peptides enable resilience against abiotic and biotic stressors. In summary these scientific advances are indicators of the potential these proteins hold for various applications in human welfare and industry apart from addressing basic research questions in biochemistry and evolution.

REFERENCES

1. C. P. Ponting, P. L. Oliver, W. Reik, Evolution and functions of long noncoding RNAs.
2. Cell.2009. 136. 629–641.
3. Clément Immarigeon,Yohan Frei, Sofie Y. N. and Robert K. Maeda.Identification of a micropeptide and multiple secondary cell genes that modulate Drosophila male reproductive success. Proc Natl Acad Sci U S A. 2021 Apr 13;118(15):e2001897118. doi: 10.1073/pnas.2001897118 https://doi.org/10.1073/pnas.2001897118;
4. Emmanuel Ladoukakis, Vini Pereira, Emile G Magny, Adam Eyre-Walker & Juan Pablo Couso. Hundreds of putatively functional small open reading frames in Drosophila. Genome Biol. 2011. 12, R118.
5. Alvaro de Andres-Pablo, Antonin Morillon, Maxime Wery. LncRNAs, lost in translation or licence to regulate?.Curr. Genet. 2017.63.29–3.
6. Diego Guerra-Almeida, Diogo Antonio Tschoeke, Rodrigo Nunes-da-Fonseca.Understanding small ORF diversity through a comprehensive transcription feature classification.DNA Res. 2021 Jul 7;28(5):dsab007. doi: 10.1093/dnares/dsab007
7. Ruth Steinberg, Hans-Georg Koch. The largely unexplored biology of small proteins in pro- and eukaryotes.FEBS J. 2021 Dec;288(24):7002-7024. doi: 10.1111/febs.15845. Epub 2021 Jun 8.
8. Catherine A Makarewich. The hidden world of membrane microproteins.Exp Cell Res. 2020 Mar 15;388(2):111853. doi: 10.1016/j.yexcr.2020.111853. Epub 2020 Jan 21.
9. Nausica Arnoult, Adriana Correia, Jiao Ma, Anna Merlo, Sara Garcia-Gomez, Marija Maric, Marco Tognetti, Christopher W Benner, Simon J Boulton, Alan Saghatelian, Jan Karlseder.Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN.Nature. 2017 Sep 20;549(7673):548-552. doi: 10.1038/nature24023.
10. Y Wang, R Benezra, DA Sassoon. Id expression during mouse development: a role in morphogenesis.Dev Dyn. 1992 Jul;194(3):222-30. doi: 10.1002/aja.1001940307.
11. Kaushal Kumar Bhati, Anko Blaakmeer, Esther Botterweg Paredes, Ulla Dolde, Tenai Eguen, Shin-Young Hong, Vandasue Rodrigues, Daniel Straub, Bin Sun, Stephan Wenkel.Approaches to identify and characterize microProteins and their potential uses in biotechnology. Cell Mol Life Sci. 2018 Jul;75(14):2529-2536. doi: 10.1007/s00018-018-2818-8. Epub 2018 Apr 18.
12. Munira A Basrai, Victor E Velculescu, Kenneth W Kinzler, Philip Hieter.NORF5/HUG1 is a component of the MEC1-mediated checkpoint response to DNA damage and replication arrest in Saccharomyces cerevisiae. Mol Cell Biol. 1999 Oct;19(10):7041-9. doi: 10.1128/MCB.19.10.7041.
13. Jiang Zhu, Shiyi Wang, Cheng Wang, Zhi Wang, Gan Luo, Junhui Li, Yangyang Zhan, Dongbo Cai, Shouwen Chen.Microbial synthesis of bacitracin: Recent progress, challenges, and prospects.Synth Syst Biotechnol. 2023 Apr 10;8(2):314-322. doi: 10.1016/j.synbio.2023.03.009. eCollection 2023 Jun.
14. Annie Rathore, Qian Chu, Dan Tan, Thomas F Martinez, Cynthia J Donaldson, Jolene K Diedrich, John R Yates, Alan Saghatelian.MIEF1 Microprotein Regulates Mitochondrial Translation Biochemistry. 2018 Sep 25;57(38):5564-5575. doi: 10.1021/acs.biochem.8b00726. Epub 2018 Sep 14.
15. Yoo-Seok Hwang, Hyun-Shik Lee, Teddy Kamata, Kathleen Mood, Hee Jun Cho, Emily Winterbottom, Yon Ju Ji, Arvinder Singh, Ira O Daar.The Smurf ubiquitin ligases regulate tissue separation via antagonistic interactions with ephrinB1.Genes Dev. 2013 Mar 1;27(5):491-503. doi: 10.1101/gad.208355.112.
16. José I Pueyo, Emile G Magny, Christopher J Sampson, Unum Amin, Iwan R Evans, Sarah A Bishop, Juan P Couso.Hemotin, a Regulator of Phagocytosis Encoded by a Small ORF and Conserved across Metazoans.PLoS Biol. 2016 Mar 25;14(3):e1002395. doi: 10.1371/journal.pbio.1002395. eCollection 2016 Mar.
17. Sarah A Slavoff, Jinho Heo, Bogdan A Budnik, Leslyn A Hanakahi, Alan Saghatelian. Human Short Open Reading Frame (sORF)-encoded Polypeptide That Stimulates DNA End Joining.J Biol Chem. 2014 Apr 18; 289(16): 10950–10957.Published online 2014 Mar 7. doi: 10.1074/jbc.C113.533968.
18. Nadia G D'Lima, Jiao Ma, Lauren Winkler, Qian Chu, Ken H Loh, Elizabeth O Corpuz, Bogdan A Budnik, Jens Lykke-Andersen, Alan Saghatelian, Sarah A Slavoff.A human microprotein that interacts with the mRNA decapping complexNat Chem Biol. 2017 Feb;13(2):174-180. doi: 10.1038/nchembio.2249. Epub 2016 Dec 5.
19. Jose Ignacio Pueyo, Juan Pablo Couso.The 11-aminoacid long Tarsal-less peptides trigger a cell signal in Drosophila leg development.Dev Biol. 2008 Dec 15;324(2):192-201. doi: 10.1016/j.ydbio.2008.08.025. Epub 2008 Sep 5.
20. Andrea Pauli, Eivind Valen, Alexander F Schier.. Identifying (non-)coding RNAs and small peptides: challenges and opportunities. Bioessays. 2015 Jan;37(1):103-12. doi: 10.1002/bies.201400103. Epub 2014 Oct 24.
21. Sonam Dhamija & Manoj B. Menon.Non-coding transcript variants of protein-coding genes – what are they good for?, RNA Biology.2018.15:8.1025-1031, DOI:10.1080/15476286.2018.1511675.
22. Diego Guerra-Almeida, Diogo Antonio Tschoeke, Rodrigo Nunes-da-Fonseca Understanding small ORF diversity through a comprehensive

transcription feature classification. DNA Research. 2021. 28(5), 1-18. https://doi.org/10.1093/dnares/dsab007.

23. Couso, J. and Patraquim, P. Classification and function of small open reading frames, Nat. Rev. Mol. Cell Biol. 2017. 18. 575–89.

24. Ladoukakis, E., Pereira, V., Magny, E.G., Eyre-Walker, A. and Couso, J., Hundreds of putatively functional small open reading frames in Drosophila, Genome Biol. 2011.12. R118. 43.

25. Hu¨cker, S.M., Ardern, Z., Goldberg, T., *et al*., Discovery of nu merous novel small genes in the intergenic regions of the Escherichia coli O157:H7 Sakai genome, PLoS One. 2017.12 e0184119.

26. Friedman, R.C., Kalkhof, S., Doppelt-Azeroual, O., *et al*., Common and phylogenetically widespread coding for peptides by bacte rial small RNAs, BMC Genomics. 2017.18. 553.

27. Bronesky, D., Wu, Z., Marzi, S., *et al*. Staphylococcus aureus RNAIII and its regulon link quorum sensing, stress responses, metabolic adaptation, and regulation of of Virulence Gene Expression.2016. Annu Rev Microbiol. 2016 Sep 8:70:299-316

28. Yuji Kageyama, Takefumi Kondo, Yoshiko Hashimoto.Coding vs non-coding: translatability of short ORFs found in putative non-coding transcripts, Biochimie.2011.93, 1981–6.

29. Kageyama, Y., Kondo, T. and Hashimoto, Y., Coding vs non-coding: translatability of short ORFs found in putative non-coding transcripts, Biochimie. 2011.93.1981–6.

30. Lauressergues, D., Couzigou, J.-M., Clemente, H.S., *et al*. Primary transcripts of microRNAs encode regulatory peptides, Nature.2015.520. 90–3.

31. Pamudurti, N.R., Bartok, O., Jens, M., *et al*. Translation of CircRNAs, Mol. Cell. 2017.66. 9–21.e7.

32. Pauli, A., Valen, E. and Schier, A.F. Identifying (non-)coding RNAs and small peptides: challenges and opportunities, Bioessays. 2015. 37 103–12.

33. Chu, Q., Ma, J. and Saghatelian, A. Identification and characteri zation of sORF-encoded polypeptides, Crit. Rev. Biochem. 2015. Mol. Biol. 50.134–41.

34. Sasaki, K., Nakamura, Y., Maki, K., *et al*. Functional analysis of a dominant-negative DETS TEL/ETV6 isoform, Biochem. Biophys. Res. Commun. 2004.317.1128–37.

35. Ohno, S. Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence, Proc. Natl. Acad. Sci. U S A.1984. 81.2421–5.

36. Klemke, M., H., Kehlenbach, R. and Huttner, W.B. Two overlap ping reading frames in a single exon encode interacting proteins–a novel way of gene usage, Embo J. 2001. 20. 3849–60.

37. Vanderperre, B., Lucier, J.-F., Bissonnette, C., *et al*. Direct detection of alternative open reading frames translation products in human significantly expands the proteome, PLoS One. 2013.8. e70698.

38. Kochetov, A.V. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs, Bioessays.2008.30,.683–91.

39. Uenaka, A., Ono, T., Akisawa, T., Wada, H., Yasuda, T. and Nakayama, E. Identification of a unique antigen peptide pRL1 on BALB/c RL male 1 leukemia recognized by cytotoxic T lymphocytes and its relation to the Akt oncogene, J. Exp. Med. 1994.180.1599–607.

40. Michel, A.M., Choudhury, K.R., Firth, A.E., Ingolia, N.T., Atkins, J.F. and Baranov, P.V. Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res. 2012 Nov;22(11):2219-29. doi: 10.1101/gr.133249.111. Epub 2012 May 16.

41. Wang, R.F., Parkhurst, M.R., Kawakami, Y., Robbins, P.F. and Rosenberg, S.A. Utilization of an alternative open reading frame of a normal gene in generating a novel human cancer antigen, J. Exp. Med. 1996. 183. 1131–40.

42. Mehta, A., Trotta, C.R. and Peltz, S.W. Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells, Genes Dev. 2006. 20.939–53.

43. Malarkannan, S., Shih, P.P., Eden, P.A., *et al*. The molecular and functional characterization of a dominant minor H antigen, H60, J. Immunol. 1998.161. 3501–9.

44. K. Bharathan Sruthi,Athira Menon, Akash Pakash, Peppurath Vasudevan Soniya.Pervasive translation of small open reading frames in plant long non-coding RNAs.Front Plant Sci. 2022 Oct 24;13:975938. doi: 10.3389/fpls.2022.975938. eCollection 2022.

45. Marta A Inchingolo, Aurélie Diman, Maxime Adamczewski, Tom Humphreys, Pascale Jaquier-Guble, Joseph A Curran.TP53BP1, a dual-coding gene, uses promoter switching and translational reinitiation to express a smORF protein.iScience. 2023 Apr 27;26(5):106757. doi: 10.1016/j.isci.2023.106757. eCollection 2023 May 19.

46. Pedro Patraquim, Emile G. Magny, José I. Pueyo, Ana Isabel Platero & Juan Pablo Couso.Translation and natural selection of micropeptides from long non-canonical RNAs.Nature Communications.2022. volume 13, 6515

47. Dominique Lauressergues, Jean-Malo Couzigou, Hélène San Clemente, Yves Martinez, Christophe Dunand, Guillaume Bécard, Jean-Philippe Combier. Primary transcripts of microRNAs encode regulatory peptidesNature. 2015 Apr 2;520(7545):90-3. doi: 10.1038/nature14346. Epub 2015 Mar 25.

48. Julie L Aspden, Ying Chen Eyre-Walker, Rose J Phillips, Unum Amin, Muhammad Ali S Mumtaz, Michele Brocard, Juan-Pablo Couso.Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq.Elife. 2014 Aug 21:3:e03528. doi: 10.7554/eLife.03528.

49. Bersabe Wondimagegnhu,Wen Ma, Tapas Paul, Ting-Wei Liao, Chun Ying Lee, Samantha Sanford, Patricia L Opresko, Sua Myong.The molecular mechanism for TERRA recruitment and annealing to telomeres.Nucleic Acids Res. 2024 Sep 23;52(17):10490-10503. doi: 10.1093/nar/gkae732.
50. Jingwen Song, Syed Nabeel-Shah, Shuye Pu, Hyunmin Lee, Ulrich Braunschweig, Zuyao Ni, Nujhat Ahmed, Edyta Marcon, Guoqing Zhong,Debashish Ray, Kevin C H Ha, Xinghua Guo, Zhaolei Zhang, Timothy R Hughes, Benjamin J Blencowe,Jack F Greenblatt.Regulation of alternative polyadenylation by the C2H2-zinc-finger protein Sp1.Mol Cell. 2022 Sep 1;82(17):3135-3150.e9. doi: 10.1016/j.molcel.2022.06.031. Epub 2022 Jul 31.
51. Keira R Hassel, Omar Brito-Estrada, Catherine A Makarewich. Microproteins: Overlooked regulators of physiology and disease.iScience. 2023 Apr 29;26(6):106781. doi: 10.1016/j.isci.2023.106781. eCollection 2023 Jun 16.
52. Luis Enrique Cabrera-Quio, Sarah Herberg, Andrea Pauli.Decoding sORF translation - from small proteins to gene regulation. RNA Biol. 2016 Nov;13(11):1051-1059. doi: 10.1080/15476286.2016.1218589. Epub 2016 Aug 12.
53. Rui Vitorino, Sofia Guedes, Francisco Amado, Manuel Santos, Nobuyoshi Akimitsu.. The role of micropeptides in biology.Cell Mol Life Sci. 2021 Apr;78(7):3285-3298. doi: 10.1007/s00018-020-03740-3. Epub 2021 Jan 28.
54. Emile G Magny, Jose Ignacio Pueyo, Frances M G Pearl, Miguel Angel Cespedes, Jeremy E Niven, Sarah A Bishop, Juan Pablo CousoConserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames.Science. 2013 Sep 6;341(6150):1116-20. doi: 10.1126/science.1238802. Epub 2013 Aug 22.
55. Akinobu Matsumoto, Alessandra Pasut, Masaki Matsumoto, Riu Yamashita, Jacqueline Fung, Emanuele Monteleone, Alan Saghatelian, Keiichi I. Nakayama, John G. Clohessy & Pier Paolo Pandolfi. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature. 2017 Jan 12;541(7636):228-232. doi: 10.1038/nature21034. Epub 2016 Dec 26.
56. Sheng Xu, LuYu Zhou, Murugavel Ponnusamy, LiXia Zhang, YanHan Dong, YanHui Zhang, Qi Wang, Jing Liu, Kun Wang.A comprehensive review of circRNA: from purification and identification to disease marker potential.PeerJ. 2018 Aug 24:6:e5503. doi: 10.7717/peerj.5503. eCollection 2018.
57. Yanan Pang, Chuanbin Mao,Shanrong Liu. Encoding activities of non-coding RNAs.Theranostics. 2018 Mar 28;8(9):2496-2507. doi: 10.7150/thno.24677. eCollection 2018.
58. Carina Azevedo Oliveira Silva, Sandy da Silveira Alves, Bruno da Costa Rodrigues, Jonatha Anderson Fraga Egidio, Lupis Ribeiro, Carlos Logullo, Flavia Borges Mury, Daniele das Graças Santos, Taynan Portal, Cintia Monteiro-de-Barros, José Roberto da Silva, José Luciano Nepomuceno-Silva, Rodrigo Nunes-da-Fonseca.The mlpt smORF gene is essential for digestive physiology and molting during nymphal stages in the kissing bug Rhodnius prolixus.Insect Biochem Mol Biol. 2024 Sep:172:104154. doi: 10.1016/j.ibmb.2024.104154. Epub 2024 Jul 5.
59. Iñaki Merino-Valverde, Emanuela Greco, María Abad.The microproteome of cancer: From invisibility to relevance.Exp Cell Res. 2020 Jul 1;392(1):111997. doi: 10.1016/j.yexcr.2020.111997. Epub 2020 Apr 14.
60. Jia-Rui You, Zeng-Jin Wen, Jia-Wei Tian, Xiao-Bing Lv, Rong Li, Shu-Ping Li, Hui Xin, Pei-Feng Li, Yin-Feng Zhang, Rui Zhang Crosstalk between ubiquitin ligases and ncRNAs drives cardiovascular disease progression. Front Immunol. 2024; 15: 1335519. Published online 2024 Mar 7. doi: 10.3389/fimmu.2024.1335519.
61. Nils G Walter. Are non-protein coding RNAs junk or treasure?: An attempt to explain and reconcile opposing viewpoints of whether the human genome is mostly transcribed into non-functional or functional RNAsBioessays. 2024 Apr;46(4):e2300201. doi: 10.1002/bies.202300201. Epub 2024 Feb 13.
62. https://www.encodeproject.org/help/project-overview/
63. Matthew Putnins, Ioannis P Androulakis. Self-selection of evolutionary strategies: adaptive versus non-adaptive forces. Heliyon. 2021 May 15;7(5):e06997. doi: 10.1016/j.heliyon.2021.e06997. eCollection 2021 May.
64. W Ford Doolittle, Tyler D P Brunet. On causal roles and selected effects: our genome is mostly junk.BMC Biol. 2017 Dec 5;15(1):116. doi: 10.1186/s12915-017-0460-9.
65. Anastasia Chugunova, Elizaveta Loseva, Pavel Mazin, Aleksandra Mitina, Tsimafei Navalayeu, Dmitry Bilan, Polina Vishnyakova, Maria Marey, Anna Golovina, Marina Serebryakova, Philipp Pletnev, Maria Rubtsova, Waltraud Mair, Anna Vanyushkina, Philipp Khaitovich, Vsevolod Belousov, Mikhail Vysokikh, Petr Sergiev, Olga Dontsova.LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. Proc Natl Acad Sci U S A. 2019 Mar 12;116(11):4940-4945. doi: 10.1073/pnas.1809105116. Epub 2019 Feb 22.
66. Alberto Vassallo, Emilia Palazzotto, Giovanni Renzone, Luigi Botta, Teresa Faddetta, Andrea Scaloni, Anna Maria Puglia, Giuseppe Gallo.The Streptomyces coelicolor Small ORF trpM Stimulates Growth and Morphological Development and Exerts Opposite Effects on Actinorhodin and Calcium-Dependent Antibiotic Production. Front Microbiol. 2020 Feb 19:11:224. doi: 10.3389/

fmicb.2020.00224. eCollection 2020.

67. Benoît Vanderperre, Jean-François Lucier, Cyntia Bissonnette, Julie Motard, Guillaume Tremblay, Solène Vanderperre, Maxence Wisztorski, Michel Salzet, François-Michel Boisvert, Xavier RoucouDirect detection of alternative open reading frames translation products in human significantly expands the proteome.PLoS One. 2013 Aug 12;8(8):e70698. doi: 10.1371/journal.pone.0070698. eCollection 2013.

68. Zhe Ji, Ruisheng Song, Aviv Regev, Kevin Struhl. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. eLife.2015. 4:e08890. doi: 10.7554/eLife.08890

69. Aaron Wacholder, Saurin Bipin Parikh, Nelson Castilho Coelho, Omer Acar, Carly Houghton, Lin Chou, Anne-Ruxandra Carvunis..A vast evolutionarily transient translatome contributes to phenotype and fitness.Cell Syst. 2023 May 17;14(5):363-381.e8. doi: 10.1016/j.cels.2023.04.002. Epub 2023 May 9.

70. Mudge JM, Ruiz-Orera J, Prensner JR, Brunet MA, Gonzalez JM, Magrane M, et al. A community-driven roadmap to advance research on translated open reading frames detected by Ribo-seq. bioRxiv. 2021;2021.06.10.447896. http://biorxiv.org/content/early/2021/06/10/ 2021.06.10.447896.abstract

71. McLysaght A, Hurst LD. Open questions in the study of de novo genes: what, how and why. Nat Rev Genet. 2016;17:567–78. https://doi.org/10.1038/nrg.2016.78.

72. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genomewide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324:218.

73. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M, et al. Extensive translation of small open reading frames revealed by poly-ribo-seq. Elife. 2014;3:1–19.

74. López E, Wang X, Madero L, López-Pascual J, Latterich M. Functional phosphoproteomic mass spectrometry-based approaches. Clin Transl Med. 2012. https://doi.org/10.1186/2001- 1326-1- 20.

75. Low TY, Heck AJ. Reconciling proteomics with next generation sequencing. Curr Opin Chem Biol. 2016;30:14–20.

76. Cao X, Khitun A, Na Z, Dumitrescu DG, Kubica M, Olatunji E, et al. Comparative proteomic profiling of unannotated microproteins and alternative proteins in human cell lines. J Proteome Res Am Chem Soc.

77. 2020;19:3418–26.

78. Lee C, Kim KH, Cohen P. MOTS-c: a novel mitochondrial-derived peptide regulating muscle and fat metabolism. Free Radic Biol Med. 2016. https://doi.org/10.1016/j.freeradbiomed.2016.05.015.

79. Chen J, Brunner A-D, Cogan JZ, Nuñez JK, Fields AP, Adamson B, et al. Pervasive functional translation of non-canonical human open reading frames. Science (80-). 2020;367:1140–6.

80. Stein CS, Jadiya P, Zhang X, McLendon JM, Abouassaly GM, Witmer NH, et al. Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. Cell Rep. 2018;23:3710–37208.

81. Fabiola Valdivia-Francia, Ataman Sendoel.No country for old methods: New tools for studying microproteins.iScience. 2024 Jan 20;27(2):108972. doi: 10.1016/j.isci.2024.108972. eCollection 2024 Feb 16.

82. L A Herndon, M F Wolfner. A Drosophila seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating.Proc Natl Acad Sci U S A. 1995 Oct 24;92(22):10114-8. doi: 10.1073/pnas.92.22.10114.

83. Frank W Avila, Laura K Sirot, Brooke A LaFlamme, C Dustin Rubinstein, Mariana F Wolfner. Insect seminal fluid proteins: identification and function. Annu Rev Entomol. 2011; 56: 21-40. doi: 10.1146/annurev-ento-120709-144823.

84. K Ravi Ram, Mariana F Wolfner. Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. Integr Comp Biol. 2007 Sep;47(3):427-45. doi: 10.1093/icb/icm046. Epub 2007 Jun 1.

85. Tracey Chapman. Sexual Conflict: Mechanisms and Emerging Themes in Resistance Biology.Am Nat. 2018 Aug;192(2):217-229. doi: 10.1086/698169. Epub 2018 May 24.

86. Steven A Ramm, Lucy McDonald, Jane L Hurst, Robert J Beynon, Paula StockleyComparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. Mol Biol Evol. 2009 Jan;26(1):189-98. doi: 10.1093/molbev/msn237. Epub 2008 Oct 17.

87. Laura K Sirot, Alex Wong, Tracey Chapman, Mariana F Wolfner.Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions.Cold Spring Harb Perspect Biol. 2014 Dec 11;7(2):a017533. doi: 10.1101/cshperspect.a017533.

88. Jolie A Carlisle, Willie J Swanson.Molecular mechanisms and evolution of fertilization proteins.J Exp Zool B Mol Dev Evol. 2021 Dec;336(8):652-665. doi: 10.1002/jez.b.23004. Epub 2020 Oct 4.

89. Mrinalini, Cho Yeow Koh, Nalini Puniamoorthy. Rapid Genomic Evolution Drives the Diversification of Male Reproductive Genes in Dung Beetles. Genome Biol Evol. 2021 Aug; 13(8): evab172. Published online 2021 Jul 28. doi: 10.1093/gbe/evab172.

90. Seirian Sumner, Emeline Favreau, Katherine Geist, Amy L. Toth, Sandra M. Rehan.Molecular patterns

and processes in evolving sociality: lessons from insects. Philos Trans R Soc Lond B Biol Sci. 2023; April 10, 378(1874): 20220076. Published online February 20, 2023. doi: 10.1098/rstb.2022.0076.

91. Patlar B, Civetta A.Seminal fluid gene expression and reproductive fitness in Drosophila melanogaster. BMC Ecol Evol. 2022 Feb 23;22(1):20. doi: 10.1186/s12862-022-01975-1.

92. Martin D Garlovsky, Yasir H Ahmed-Braimah. Evolutionary Quantitative Proteomics of Reproductive Protein Divergence in Drosophila. Mol Cell Proteomics. 2023 Aug;22(8):100610. doi: 10.1016/j.mcpro.2023.100610. Epub 2023 Jun 28.

93. Papachristos K, Sayadi A, Arnqvist G.Comparative Genomic Analysis of the Pattern of Evolution of Male and Female Reproductive Proteins in Seed Beetles. Genome Biol Evol. 2024 Jul 3;16(7):evae143. doi: 10.1093/gbe/evae143.

94. Brittany L. Enzmann, Peter Nonacs. Optimists or realists? How ants allocate resources in making reproductive investments.J Anim Ecol. 2018 Jul;87(4):1126-1136. doi: 10.1111/1365-2656.12840. Epub 2018 May 29.

95. Ruiz-Orera J, Albà MM.Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation.Trends Genet. 2019 Mar;35(3):186-198. doi: 10.1016/j.tig.2018.12.003. Epub 2018 Dec 31.

96. Juan-Pablo Couso, Pedro Patraquim.Classification and function of small open reading frames.Nat Rev Mol Cell Biol. 2017 Sep;18(9):575-589. doi: 10.1038/nrm.2017.58. Epub 2017 Jul 12.

97. Emmanuel Ladoukakis, Vini Pereira, Emile G Magny, Adam Eyre-Walker & Juan Pablo Couso. Hundreds of putatively functional small open reading frames in Drosophila.Genome Biol. 2011 Nov 25;12(11):R118. doi: 10.1186/gb-2011-12-11-r118.

98. Azza Dib, Jennifer Zanet, Alexandra Mancheno-Ferris, Maylis Gallois, Damien Markus, Philippe Valenti, Simon Marques-Prieto, Serge Plaza, Yuji Kageyama, Hélène Chanut-Delalande, François Payre.Pri smORF Peptides Are Wide Mediators of Ecdysone Signaling, Contributing to Shape Spatiotemporal Responses.Front Genet. 2021 Aug 30:12:714152. doi: 10.3389/fgene.2021.714152. eCollection 2021.

99. Linrong Wan, Wenfu Xiao, Ziyan Huang, Anlian Zhou, Yaming Jiang, Bangxing Zou, Binbin Liu, Cao Deng, Youhong Zhang..Systematic identification of smORFs in domestic silkworm (Bombyx mori).Peer J. 2023 Jan 13;11:e14682. doi: 10.7717/peerj.14682. eCollection 2023.

100. Cornelia Fritsch, F Javier Bernardo-Garcia, Tim-Henning Humberg, Abhishek Kumar Mishra, Sara Miellet, Silvia Almeida, Michael V Frochaux, Bart Deplancke, Armin Huber, Simon G Sprecher. Multilevel regulation of the glass locus during Drosophila eye development.PLoS Genet. 2019 Jul 12;15(7):e1008269. doi: 10.1371/journal.pgen.1008269. eCollection 2019 Jul.

101. Emile G Magny, Jose Ignacio Pueyo, Frances M G Pearl, Miguel Angel Cespedes, Jeremy E Niven, Sarah A Bishop, Juan Pablo Couso. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames.Science. 2013 Sep 6;341(6150):1116-20. doi: 10.1126/science.1238802. Epub 2013 Aug 22.

102. Peter F Renz, Fabiola Valdivia-Francia, Ataman Sendoel..Some like it translated: small ORFs in the 5'UTR.Exp Cell Res. 2020 Nov 1;396(1):112229. doi: 10.1016/j.yexcr.2020.112229. Epub 2020 Aug 17.

103. Chrysoula-Evangelia Karachaliou and Evangelia Livaniou.Neuroprotective Action of Humanin and Humanin Analogues: Research Findings and Perspectives. Biology (Basel). 2023 Dec; 12(12): 1534.

104. Hanna Persha, Stephen A Thacker, Krutika Mediwala Hornback, Gustavo R Alvira-Arill, Richard Lueking, Taylor Morrisette. Real-World Clinical Characteristics and Outcomes with Daptomycin Use in Pediatric Patients: A Retrospective Case Series Antibiotics (Basel). 2024 Sep 2;13(9):833. doi: 10.3390/antibiotics13090833.

105. Assim A. Alfadda, Anas M. Abdel Rahman,Hicham Benabdel kamel,Reem AlMalki,Bashayr Alsuwayni,Abdulaziz Alhossan,Madhawi M. Aldhwayan,Ghalia N. Abdeen,Alexander Dimitri Miras and Afshan Masood. Metabolomic Effects of Liraglutide Therapy on the Plasma Metabolomic Profile of Patients with Obesity.Metabolites. 2024 Sep 17;14(9):500. doi: 10.3390/metabo14090500.