



Evolution of the Insulin Gene: Changes in Gene Number, Sequence, and Processing

David M. Irwin^{1,2*}

¹ Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada, ² Banting and Best Diabetes Centre, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:

Jeff M. P. Holly,
University of Bristol, United Kingdom

Reviewed by:

Briony Forbes,
Flinders University, Australia
Sara V. Good,
University of Winnipeg, Canada

*Correspondence:

David M. Irwin
david.irwin@utoronto.ca

Specialty section:

This article was submitted to
Molecular and Structural
Endocrinology,
a section of the journal
Frontiers in Endocrinology

Received: 04 January 2021

Accepted: 01 March 2021

Published: 02 April 2021

Citation:

Irwin DM (2021) Evolution of the Insulin Gene: Changes in Gene Number, Sequence, and Processing. *Front. Endocrinol.* 12:649255. doi: 10.3389/fendo.2021.649255

Insulin has not only made major contributions to the field of clinical medicine but has also played central roles in the advancement of fundamental molecular biology, including evolution. Insulin is essential for the health of vertebrate species, yet its function has been modified in species-specific manners. With the advent of genome sequencing, large numbers of insulin coding sequences have been identified in genomes of diverse vertebrates and have revealed unexpected changes in the numbers of genes within genomes and in their sequence that likely impact biological function. The presence of multiple insulin genes within a genome potentially allows specialization of an insulin gene. Discovery of changes in proteolytic processing suggests that the typical two-chain hormone structure is not necessary for all of insulin's biological activities.

Keywords: insulin, gene duplication, evolution, adaptive evolution, vertebrates, proteolytic processing, virus

INTRODUCTION

Insulin is well-characterized as a key regulator of blood glucose levels in vertebrates (1). Insulin related peptides have also been identified in a number of other metazoan species and have been shown to contribute to various aspects of physiology in these species (2–4). The discovery of insulin 100 years ago, led to a revolution in clinical medicine, as it allowed an effective treatment for diabetes (5). Since its discovery, the treatment of diabetes using insulin, derivatives of insulin, and other peptides has and continues to evolve (6). There still is no cure. In addition to its critical role in the history of clinical medicine, insulin has played key roles in the development of many revolutionary technologies that are now commonplace in molecular biology, including protein sequencing (7) and the deduction of the three-dimensional structures of proteins (8). A key discovery made with insulin, but with important implications for many other bioactive peptides, is the role of proteolytic processing in regulating its biological action (9, 10). Since the sequencing of human insulin more than 60 years ago (7), a large number of insulin protein sequences have been determined due to its importance in medicine, as well as its small size and relative ease at isolation (11, 12). Over the past 20 years, as we entered the genomic era, an increasing number of insulin sequences have been predicted from the complete genome sequences of organisms. Genomic sequences have led to improvements of our understanding of not only human genetics and disease (13, 14), but also nearly all other areas of biology (15). The new insulin sequences identified from genome sequences have revealed an increased diversity in the number of insulin genes within

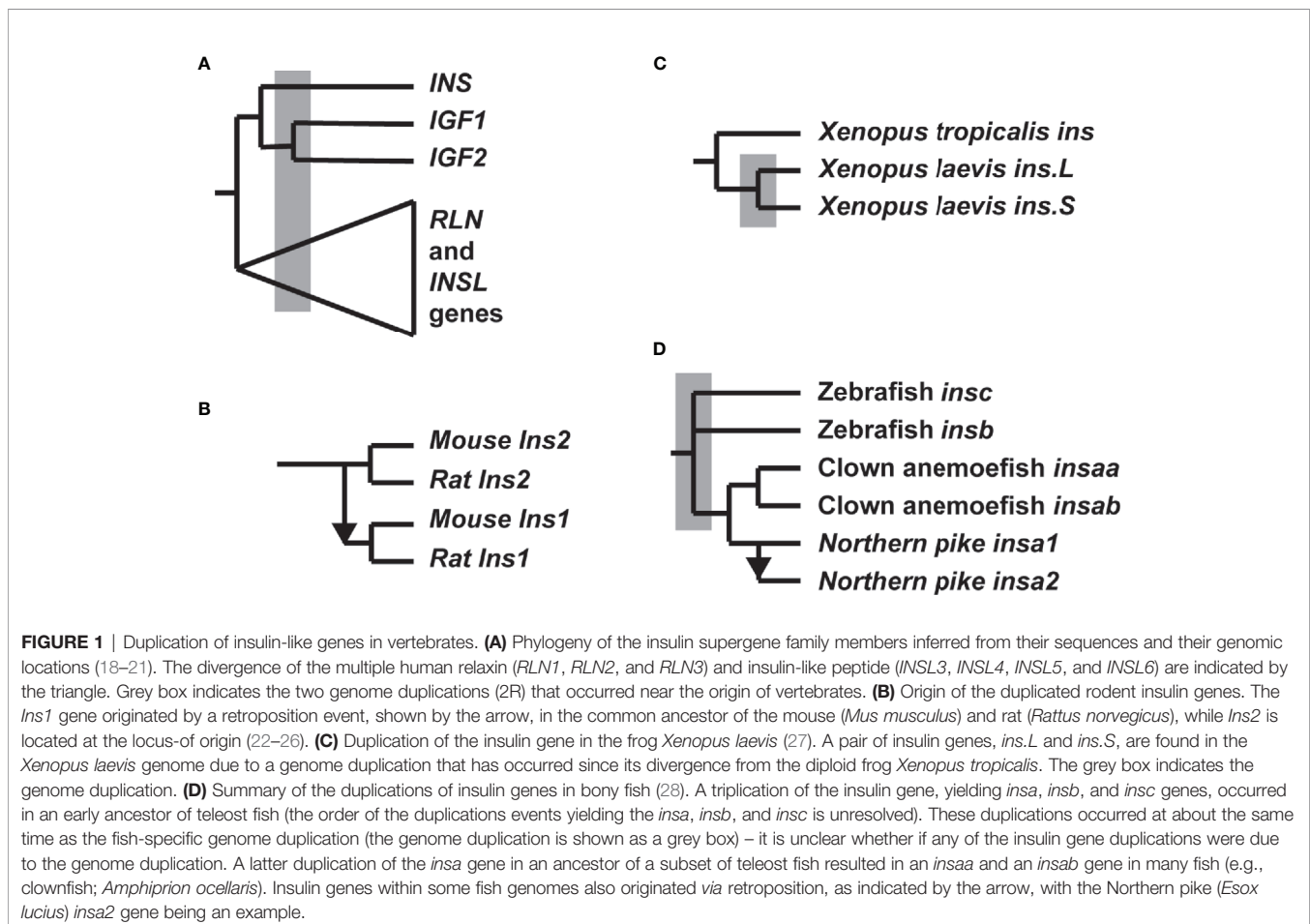
species and has revealed that changes in the proteolysis processing of the proinsulin precursor likely contributes to the diversity of the biological actions of insulin.

SUPERFAMILY OF INSULIN-LIKE GENES

While insulin was first identified in mammals, it soon became evident that peptides with sequences similar to insulin can be found in diverse multicellular animals, including many non-vertebrate species such as insects and worms (2–4). In many of these species, the insulin-like peptides were found through directed efforts to identify peptides with similarity to insulin, but increasingly, they are now being reported from searches of genome sequences. Multiple insulin-like genes have been characterized in the genomes of many non-vertebrate species that are due to lineage-specific gene duplication events (3, 4). A parallel set of duplications of insulin-like genes has also occurred within vertebrates. In addition to the insulin gene (*INS*), nine other genes encoding peptides with similarity to insulin both in their primary sequences and secondary structures have been identified in the human genome, including 2 insulin-like growth factors (*IGF1* and *IGF2*), 4 insulin-like factors (*INSL3*, *INSL4*, *INSL5*, and *INSL6*), and three relaxins (*RLN1*, *RLN2*, and *RLN3*)

(16, 17). While the relationships among these genes cannot be fully resolved by phylogenetic analysis, due to their short protein lengths, information derived from their locations within the genome has helped to largely resolve the order and timing of the gene duplication events that generated this gene family (18–21). As summarized in **Figure 1A**, these studies suggest that the initial gene duplication event that originated this gene family separated an ancestor for the insulin and insulin-like growth factor (*IGF1* and *IGF2*) genes from an ancestor of the insulin-like (*INSL*) and relaxin (*RXN*) genes. This duplication was then followed by a duplication that separated the insulin gene from an ancestor of the insulin-like growth factor (*IGF1* and *IGF2*) genes. Both of these gene duplication events occurred before the two genome duplications that are associated with the origin of vertebrates (18).

The human insulin gene (*INS*) is a small gene of 1,425 base pairs located on chromosome 11 and is composed of 3 exons separated by two introns (29). The first exon of this gene is composed entirely of 5' untranslated sequence, with all of the coding region, which encodes the 110 amino acid long proinsulin precursor protein, distributed across exons 2 and 3 (29). The N-terminal portion of the proinsulin precursor protein sequence is a signal peptide that allows secretion from pancreatic beta cells and is removed by signal peptidase to yield proinsulin (10, 30).



Further processing by prohormone convertase enzymes releases the internal C-peptide to produce the two-chain insulin hormone, composed of A- and B-chains linked by disulphide bridges (10, 30). The proteolytic processing of proinsulin to yield a two-chain insulin hormone contrasts with that of the insulin-like growth factors, where both are single chain peptides that retain the C-peptide-like sequences (31, 32). Insulin genes from diverse species representing multiple classes of vertebrates, including fish, frogs, and birds, have been characterized that share with the human gene a similar three exon (two coding) gene structure, and encode homologous protein sequences (30, 33). In most vertebrates only a single insulin gene has been found, however, multiple copies of this gene, including some with differing gene structures, have been found in some species.

DUPLICATED INSULIN GENES

Rats (*Rattus norvegicus*) and mice (*Mus musculus*) were the first vertebrates found to each have two insulin proteins, which were subsequently found to be encoded by a pair of genes (22). While the insulin 2 genes (*Ins2*) have a gene structure similar to that of the human gene (three exons and two introns), the insulin 1 gene (*Ins1*) was found to be composed of only 2 exons, with all of the coding sequence contained in the second exon (23). These two genes have a relatively recent origin (see **Figure 1B**), in the common ancestor of mice and rats (22–26). Further study of these genes indicated that the *Ins1* gene originated from an aberrant mRNA transcript that was initiated about 500 bases upstream of the normal mRNA start site and partially processed to remove only intron 2 sequences before being reverse transcribed and inserted into the genome (23, 24, 26). Despite the differences in the structures of these two genes, both are equally expressed (22, 24). Since the mouse and rat *Ins1* genes only have ~500 bases of 5' flanking sequence homologous to the *Ins* genes of other mammals, this indicates that only a limited amount of sequence is needed for efficient beta-cell-specific expression of the insulin gene. But why do mice and rats have two insulin genes that have identical expression patterns, while other species can survive with a single gene? This might suggest that the two genes differ in function, perhaps specializing function at different sites or times, thus generating a need to retain both genes. While some evidence for differences in the selective constraints acting on the two insulin genes has been detected (25), no evidence for different functions have been found, thus a convincing explanation has not been reached. Intriguingly, duplicated insulin genes that originated by retroposition, like the rodent *Ins1* gene, have also been described in several species of bony fish through three independent origins (28). While most of the retroposed insulin-like sequences in fish appeared to be pseudogenes, the gene sequence of a potentially retroprocessed insulin gene in the Northern pike (*Esox lucius*) has an intact coding sequence (**Figure 1C**) that is potentially functional (28). Further study is needed to determine whether this retroposed Northern pike insulin gene is expressed and if its encoded protein has a physiological function. These observations also indicate that the vertebrate insulin gene is expressed in the

germ cells of a number of species, thus allowing it to be retroprocessed and integrated into the genome allowing to be passed on to the next generation, raising questions about the possible function of insulin in germ line cells.

Duplicated copies of the insulin gene have also been found in several other vertebrate species (**Figures 1C, D**), where these genes retain the three exon and two intron gene structure and potentially have large amounts of flanking sequences that would allow their continued expression. The frog *Xenopus laevis*, which experienced a recent genome duplication (**Figure 1D**), was the first published example (34). Unlike the duplicated rodent insulin genes, the two *Xenopus laevis* insulin genes were found to display differing developmental gene expression patterns that might suggest diverging functions (34), and a reason why both are retained, however they also had overlapping gene expression in the adult pancreas (35). A genome duplication, the fish-specific whole genome duplication (3R), was experienced by an ancestor of teleost fish (36), thus the discovery of a second insulin gene in several fish genome sequences (37) was not a surprise. Characterization of the two zebrafish (*Danio rerio*) insulin genes (*insa* and *insb*) provided evidence that the two genes had distinct expression patterns and potentially differing functions (38), supporting a hypothesis that fish genes have diverged in function, potentially subfunctionalizing so that each now is responsible for a subset of the ancestral functions of insulin. More recently, a third insulin gene (*insc*) has been found in some, but not all, fish (**Figure 1C**), with all three of these genes originating in an early teleost, thus it is unclear which, if any, originated through the fish-specific whole genome duplication (28). The role of this third insulin gene in fish physiology is unknown. Does it also possess a subset of ancestral insulin functions, or has it gained new function? In addition to these three types of insulin genes found in fish, additional lineage-specific duplications of insulin genes were found, including some species [e.g., carp (*Cyprinus carpio*)] that are associated with additional genome duplications on these lineages (28). In addition to these lineage-specific duplications due to genome duplications, a duplication of the *insa* insulin gene, resulting in the *insaa* and *insab* genes, occurred early in the diversification of teleost fish (**Figure 1C**) yielding a large number of species with these gene duplicate. Intriguingly, most of the proteins encoded by the *insab* genes have amino acid substitutions that are predicted to impair proteolytically processing that generates the typical two-chain insulin hormone (28). While these sequences retain signal peptides, which would allow secretion, and cysteine residues that allow disulphide bridge formation, this raises the possibility that they yield an unprocessed insulin-like protein that retains an insulin-like protein structure and has an unknown function (28).

EVOLUTION OF INSULIN SEQUENCES

In addition to changes in the numbers and structure of insulin genes, sequences of insulin genes have also changed. Typically, genes evolve at a near steady rate, but occasionally they display

episodes of more rapid change, which are hypothesized to signal a change in gene function. Studies of mammalian insulins have provided support for this hypothesis. Insulin sequences from the guinea pig (*Cavia porcellus*) and relatives (rodents of the suborder Hystricomorpha) are well known for having insulin sequences with highly divergent sequences (39, 40). The biological activities of these insulins also differ, acting more as a growth factor than as a metabolic hormone (41, 42). These changes, in sequence and function, have been accompanied with an acceleration of the rate of evolution of the insulin protein sequence in the guinea pig and relatives (40, 43, 44). Similar, but less dramatic, episodes of accelerated evolution of insulin sequences have been observed in some species of New World monkeys (44, 45), species that have insulin hormones with lower potency (46).

CHANGES IN PROTEOLYTIC PROCESSING

Recent surveys of fish and mammalian insulin coding sequences have identified several species that have accumulated increased amounts of sequence change (28, 44). However, in contrast to the sequences from the rodent suborder of Hystricomorpha and New World monkeys, the striking changes in these sequences were at sites involved in proteolytic processing. Insulin is composed of two peptide chains linked by disulphide bonds, with both peptide chains generated from a single precursor protein (9, 10). Studies on insulin emphasized the importance of proteolytic processing in the generation of bioactive peptides (2, 9, 10, 47). The insulin precursor undergoes two types of proteolytic processing to generate a functional hormone: 1) removal of its signal peptide, which is necessary for secretion, and 2) removal of the C-peptide to generate the two-

chain molecule linked by disulphide bonds (10, 48). Evolutionary changes have occurred at all of these proteolytic sites. Changes in the sites of proteolytic processing of the insulin precursor had previously been observed, especially at the signal peptidase cleavage site and at the B-chain/C-peptide processing site (12). While some earlier studies have suggested altered proteolytic processing of proinsulin in some species of the most divergent classes of vertebrates, jawless and cartilaginous fish (49, 50), recent studies of insulin gene sequences obtained from the genomic sequences of some species of bony fish and mammals predict amino acid replacements that abolish proteolytic processing at both the B-chain/C-peptide and the C-peptide/A-chain processing sites of insulin (28, 44). Almost all of the proinsulin sequences predicted by teleost fish *ins* genes encode sequences that can be processed into two-chain insulin hormones, except for those encoded by *insab* genes (28). Most proinsulin sequences encoded by the *insab* genes contain amino acid substitutions at their B-chain/C-peptide processing site that likely prevent proteolytic cleavage, with many of them also having substitutions that should impair processing at the C-peptide/A-chain site (Figure 2), thus leading to improperly processed insulin molecules (28). Similarly, insulin gene sequences from two bats (*Myotis brandtii* and *M. lucifugus*) predict substitutions at both the A-chain/C-peptide and C-peptide/B-chain processing sites that should prevent processing (Figure 2), while the insulin sequences from several species of Afrotheria [e.g., armadillo (*Oryzomys azeri*)] likely have altered C-peptide/A-chain processing (Figure 2) (44). Thus, it is likely that the insulin genes of many vertebrate species generate a final protein product that is composed of either a single protein chain or are two-chain protein molecules that have an A- or B-chains that is extended to include the complete C-peptide sequence (28, 44). Intriguingly, all of these insulin sequences with altered protein processing retain conserved cysteine residues that are

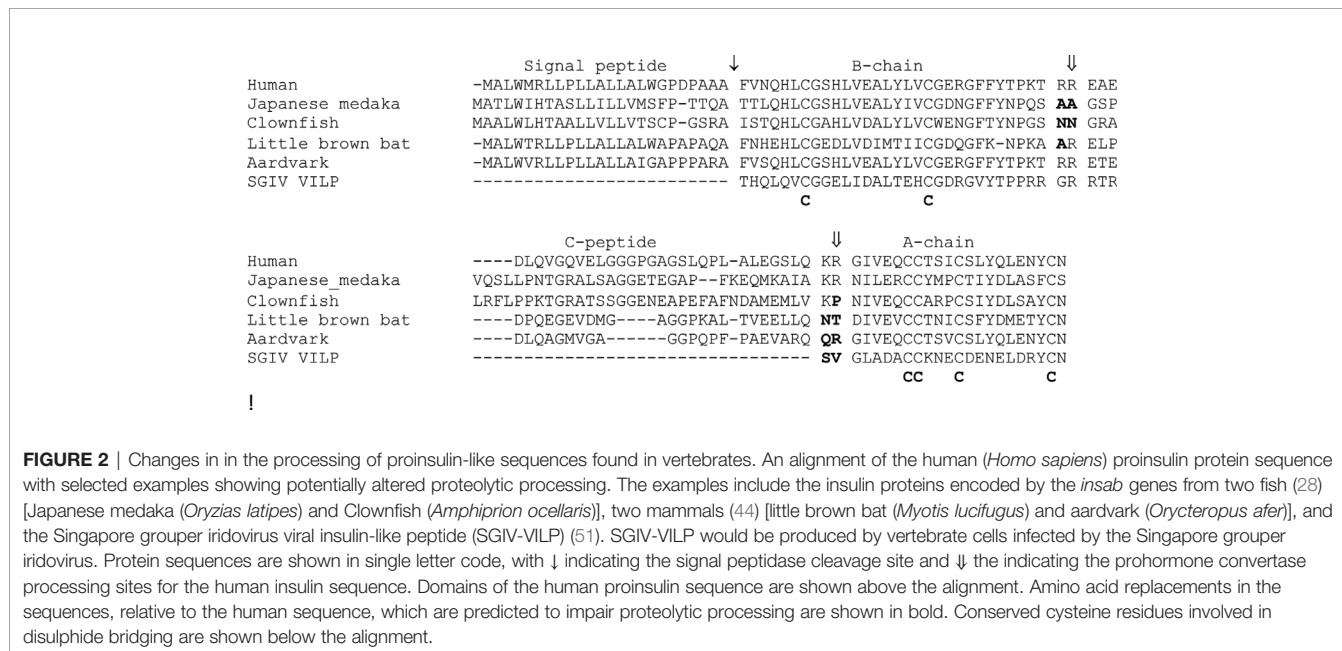


FIGURE 2 | Changes in the processing of proinsulin-like sequences found in vertebrates. An alignment of the human (*Homo sapiens*) proinsulin protein sequence with selected examples showing potentially altered proteolytic processing. The examples include the insulin proteins encoded by the *insab* genes from two fish (28) [Japanese medaka (*Oryzias latipes*) and Clownfish (*Amphiprion ocellaris*)], two mammals (44) [little brown bat (*Myotis lucifugus*) and armadillo (*Oryzomys azeri*)], and the Singapore grouper iridovirus viral insulin-like peptide (SGIV-VILP) (51). SGIV-VILP would be produced by vertebrate cells infected by the Singapore grouper iridovirus. Protein sequences are shown in single letter code, with ↓ indicating the signal peptidase cleavage site and ↓ indicating the prohormone convertase processing sites for the human insulin sequence. Domains of the human proinsulin sequence are shown above the alignment. Amino acid replacements in the sequences, relative to the human sequence, which are predicted to impair proteolytic processing are shown in bold. Conserved cysteine residues involved in disulphide bridging are shown below the alignment.

important for disulphide bridging, thus, these proteins potentially retain three-dimensional structures that are similar to insulin and are also biologically functional.

VIRAL INSULIN-LIKE PEPTIDES

Most studies on the function of insulin assume that this peptide is of endogenous origin, or from relatively closely related species. Indeed, humans have been treated with insulin from several other mammalian sources (5). Recent studies analyzing the sequences of viral genomes have revealed that some viruses that infect vertebrates could be another source of insulin-like peptides, with these peptides having the potential to affect physiology and pathophysiology (51). Altindis et al. (51) identified four viruses, which infect fish, whose genomes predict peptides with similarity with insulin, which they called viral insulin-like peptides (VILPs). While these new VILPs sequences share similarity with insulin, differences exist at the regions corresponding to the B-chain/C-peptide and C-peptide/A-chain processing sites (Figure 2), thus, and might not generate two-chain molecules (51). However, like the incompletely processed vertebrate insulin sequences described above, the VILPs share the conserved cysteine residues involved in disulphide bridging and are requisite to the maintain the 3D structure (51). Some VILPs have been shown to bind to insulin receptors and regulate glucose metabolism in mice (51, 52), indicating that it may function in its proinsulin form, and that they might have a pathophysiological role beneficial to the viruses. Insulin-like molecules are used as toxins by cone snails (53), thus the use of an insulin-like peptide by a virus in pathophysiology should not be a surprise. It has long been known that full length proinsulin can bind and activate the

insulin receptor (54), as do the single-chain IGF1 and IGF2 hormones with their specific receptors, thus, the incompletely processed insulin molecules encoded by genes in fish, mammals, or viruses could impact physiology or pathophysiology.

PERSPECTIVES AND FUTURE DIRECTIONS

The biology of insulin as well as the evolution of insulin have been studied for many years (2, 9, 12) yet new discoveries and insights have been gained from the analysis of the rapidly increasing amount of genomic data. Studies into the evolution of the genes for the insulin receptor, enzymes involved in producing the mature hormone and downstream signaling partners should also improve our understanding of the biology of insulin. Genomic sequences, together with improved bioinformatic search algorithms, allow unbiased searches for sequences with similarity to insulin (or your favorite protein) in genomes, revealing sequences that might not have been found in more directed searches for bioactive peptides. As we complete more genomes and microbiomes, it is certain that we will discover more insulin-like sequences with novel aspects to their sequences, structures, and functions. However, sequence will not tell us function. Experimental work is still needed to identify the functions of these novel insulin-like sequences, which may uncover new roles for insulin in biology.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Röder PV, Wu B, Liu Y, Han W. Pancreatic regulation of glucose homeostasis. *Exp Mol Med* (2016) 48:e219. doi: 10.1038/emmm.2016.6
- Chan SJ, Steiner DF. Insulin through the ages: Phylogeny of a growth promoting and metabolic regulatory hormone. *Amer Zool* (2000) 40:213–22. doi: 10.1093/icb/40.2.213
- Wu Q, Brown MR. Signaling and function of insulin-like peptides in insects. *Annu Rev Entomol* (2006) 51:1–24. doi: 10.1146/annurev.ento.51.110104.151011
- Cherif-Feidel M, Heude Berthelin C, Adeline B, Rivière G, Favrel P, Kellner K. Molecular evolution and functional characterization of insulin related peptides in molluscs: Contributions of *Crassostrea gigas* genomic and transcriptomic-wide screening. *Gen Comp Endocrinol* (2019) 271:15–29. doi: 10.1016/j.ygcen.2018.10.019
- Vasiljević J, Torkko JM, Knoch KP, Solimena M. The making of insulin in health and disease. *Diabetologia* (2020) 63:1981–9. doi: 10.1007/s00125-020-05192-7
- Vecchio I, Tornali C, Bragazzi NL, Martini M. The discovery of insulin: An important milestone in the history of medicine. *Front Endocrinol* (2018) 9:613. doi: 10.3389/fendo.2018.00613
- Ryle AP, Sanger F, Smith LF, Kitai R. The disulphide bonds of insulin. *Biochem J* (1955) 60:541–56. doi: 10.1042/bj0600541
- Blundell TL, Cutfield JF, Dodson EJ, Dodson GG, Hodgkin DC, Mercola DA. The crystal structure of rhombohedral 2 zinc insulin. *Cold Spring Harb Symp Quant Biol* (1972) 36:233–41. doi: 10.1101/sqb.1972.036.01.031
- Steiner DF, Cunningham D, Spigelman L, Aten B. Insulin biosynthesis: evidence for a precursor. *Science* (1967) 157:697–700. doi: 10.1126/science.157.3789.697
- Steiner DF. On the discovery of precursor processing. *Methods Mol Biol* (2011) 768:3–11. doi: 10.1007/978-1-61779-204-5_1
- Conlon JM. Molecular evolution of insulin in non-mammalian vertebrates. *Amer Zool* (2000) 40:200–12. doi: 10.1093/icb/40.2.200
- Conlon JM. Evolution of the insulin molecule: insights into structure-activity and phylogenetic relationships. *Peptides* (2001) 22:1183–93. doi: 10.1016/s0196-9781(01)00423-5
- Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell* (2019) 177:70–84. doi: 10.1016/j.cell.2019.02.032
- Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurler ME, et al. A brief history of human disease genetics. (2020). *Nature* 577:179–89. doi: 10.1038/s41586-019-1879-7
- Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth BioGenome Project: Sequencing life for the future of life. *Proc Natl Acad Sci USA* (2018) 115:4325–33. doi: 10.1073/pnas.1720115115
- Shabanpoor F, Separovic F, Wade JD. The human insulin superfamily of polypeptide hormones. *Vitam Horm* (2009) 80:1–31. doi: 10.1016/S0083-6729(08)00601-8
- Bathgate RA, Halls ML, van der Westhuizen ET, Callander GE, Kocan M, Summers RJ. Relaxin family peptides and their receptors. *Physiol Rev* (2013) 93:405–80. doi: 10.1152/physrev.00001.2012

18. Patton SJ, Luke GN, Holland PW. Complex history of a chromosomal paralogy region: insights from amphioxus aromatic amino acid hydroxylase genes and insulin-related genes. *Mol Biol Evol* (1998) 15:1373–80. doi: 10.1093/oxfordjournals.molbev.a025865
19. Yegorov S, Good S. Using paleogenomics to study the evolution of gene families: origin and duplication history of the relaxin family hormones and their receptors. *PLoS One* (2012) 7:e32923. doi: 10.1371/journal.pone.0032923
20. Bathgate RAD, Kocan M, Scott DJ, Hossain MA, Good SV, Yegorov S, et al. The relaxin receptor as a therapeutic target - perspectives from evolution and drug targeting. *Pharmacol Ther* (2018) 187:114–32. doi: 10.1016/j.pharmthera.2018.02.008
21. Perillo M, Arnone MI. Characterization of insulin-like peptides (ILPs) in the sea urchin *Strongylocentrotus purpuratus*: insights on the evolution of the insulin family. *Gen Comp Endocrinol* (2014) 205:68–79. doi: 10.1016/j.ygcen.2014.06.014
22. Lomedico P, Rosenthal N, Efstratidis A, Gilbert W, Kolodner R, Tizard R. The structure and evolution of the two nonallelic rat preproinsulin genes. *Cell* (1979) 18:545–58. doi: 10.1016/0092-8674(79)90071-0
23. Soares MB, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, et al. RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Mol Cell Biol* (1985) 5:2090–103. doi: 10.1128/mcb.5.8.2090
24. Wentworth BM, Schaefer IM, Villa-Komaroff L, Chirgwin JM. Characterization of the two nonallelic genes encoding mouse preproinsulin. *J Mol Evol* (1986) 23:305–12. doi: 10.1007/BF02100639
25. Shiao MS, Liao BY, Long M, Yu HT. Adaptive evolution of the insulin two-gene system in mouse. *Genetics* (2008) 178:1683–91. doi: 10.1534/genetics.108.087023
26. Irwin DM. Variation in the rates of evolution of the insulin and glucagon hormone and receptor genes in rodents. *Gene* (2020a) 728:144296. doi: 10.1016/j.gene.2019.144296
27. Shuldiner AR, Phillips S, Roberts CT Jr, LeRoith D, Roth J. *Xenopus laevis* contains two nonallelic preproinsulin genes. cDNA cloning and evolutionary perspective. *J Biol Chem* (1989) 264:9428–32. doi: 10.1016/S0021-9258(18)60549-6
28. Irwin DM. Duplication and diversification of insulin genes in ray-finned fish. *Zool Res* (2019) 40:185–97. doi: 10.24272/zj.issn.2095-8137.2018.052
29. Bell GI, Pictet RL, Rutter WJ, Cordell B, Tischer E, Goodman HM. Sequence of the human insulin gene. *Nature* (1980) 284:26–32. doi: 10.1038/284026a0
30. Steiner DF, Chan SJ, Welsh JM, Kwok SC. Structure and evolution of the insulin gene. *Annu Rev Genet* (1985) 19:463–84. doi: 10.1146/annurev.ge.19.120185.002335
31. Rotwein P. Insulinlike growth factor 1 gene variation in vertebrates. *Endocrinol* (2018) 159:2288–305. doi: 10.1210/en.2018-00259
32. Rotwein P. The insulin-like growth factor 2 gene and locus in nonmammalian vertebrates: Organizational simplicity with duplication but limited divergence in fish. *J Biol Chem* (2018) 293:15912–32. doi: 10.1074/jbc.RA118.004861
33. Hay CW, Docherty K. Comparative analysis of insulin gene promoters: implications for diabetes research. *Diabetes* (2006) 55:3201–13. doi: 10.2337/db06-0788
34. Shuldiner AR, de Pablo F, Moore CA, Roth J. Two nonallelic insulin genes in *Xenopus laevis* are expressed differentially during neurulation in prepancreatic embryos. *Proc Natl Acad Sci USA* (1991) 88:7679–83. doi: 10.1073/pnas.88.17.7679
35. Celi FS, Tanner K, Roth AK, Roth AE, Shuldiner AR. The two nonallelic *Xenopus* insulin genes are expressed coordinately in the adult pancreas. *Gen Comp Endocrinol* (1994) 95:169–77. doi: 10.1006/gcen.1994.1114
36. Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* (2005) 27:937–45. doi: 10.1002/bies.20293
37. Irwin DM. A second insulin gene in fish genomes. *Gen Comp Endocrinol* (2004) 135:150–58. doi: 10.1016/j.ygcen.2003.08.004
38. Papasani MR, Robison BD, Hardy RW, Hill RA. Early developmental expression of two insulins in zebrafish (*Danio rerio*). *Physiol Genomics* (2008) 27:79–85. doi: 10.1152/physiolgenomics.00012.2006
39. Watt VM. Sequence and evolution of guinea pig preproinsulin DNA. *J Biol Chem* (1985) 260:10926–9. doi: 10.1016/S0021-9258(17)39123-8
40. Beintema JJ, Campagne RN. Molecular evolution of rodent insulins. *Mol Biol Evol* (1987) 4:10–8. doi: 10.1093/oxfordjournals.molbev.a040424
41. King GL, Kahn CR. Non-parallel evolution of metabolic and growth-promoting functions of insulin. *Nature* (1981) 292:644–6. doi: 10.1038/292644a0
42. King GL, Kahn CR, Heldin CH. Sharing of biological effect and receptors between guinea pig insulin and platelet-derived growth factor. *Proc Natl Acad Sci USA* (1983) 80:1308–12. doi: 10.1073/pnas.80.5.1308
43. Seino S, Blackstone CD, Chan SJ, Whittaker J, Bell GI, Steiner DF. Appalachian spring: variations on ancient gastro-entero-pancreatic themes in New World mammals. *Horm Metab Res* (1988) 20:430–5. doi: 10.1055/s-2007-1010853
44. Irwin DM. Evolution of the mammalian insulin (*Ins*) gene; Changes in proteolytic processing. *Peptides* (2020b) 135:170435. doi: 10.1016/j.peptides.2020.170435
45. Wallis M. Coevolution of insulin-like growth factors, insulin and their receptors and binding proteins in New World Monkeys. *Growth Horm IGF Res* (2015) 25:158–67. doi: 10.1016/j.ghir.2015.06.004
46. Seino S, Steiner DF, Bell GI. Sequence of a New World primate insulin having low biological potency and immunoreactivity. *Proc Natl Acad Sci USA* (1987) 84:7423–7. doi: 10.1073/pnas.84.21.7423
47. De Meyts P. Insulin and its receptor: structure, function and evolution. *Bioessays* (2004) 26:1351–62. doi: 10.1002/bies.20151
48. Liu M, Wright J, Guo H, Xiong Y, Arvan P. Proinsulin entry and transit through the endoplasmic reticulum in pancreatic beta cells. *Vitam Horm* (2014) 95:35–62. doi: 10.1016/B978-0-12-800174-5.00002-8
49. Conlon JM, Göke R, Andrews PC, Thim L. Multiple molecular forms of insulin and glucagon-like peptide from the Pacific ratfish (*Hydrolagus colliciei*). *Gen Comp Endocrinol* (1989) 73:136–46. doi: 10.1016/0016-6480(89)90064-6
50. Conlon JM, Nielsen PF, Youson JH, Potter IC. Proinsulin and somatostatin from the islet organ of the southern-hemisphere lamprey *Geotria australis*. *Gen Comp Endocrinol* (1995) 100:413–22. doi: 10.1006/gcen.1995.1172
51. Altindis E, Cai W, Sakaguchi M, Zhang F, GuoXiao W, Liu F, et al. Viral insulin-like peptides activate human insulin and IGF-1 receptor signaling: A paradigm shift for host-microbe interactions. *Proc Natl Acad Sci USA* (2018) 115:2461–6. doi: 10.1073/pnas.1721117115
52. Chrudinová M, Moreau F, Noh HL, Pániková T, Žáková L, Friedline RH, et al. Characterization of viral insulins reveals white adipose tissue-specific effects in mice. *Mol Metab* (2020) 44:101121. doi: 10.1016/j.molmet.2020.101121
53. Ahorokomeye P, Disotuar MM, Gajewiak J, Karanth S, Watkins M, Robinson SD, et al. Fish-hunting cone snail venoms are a rich source of minimized ligands of the vertebrate insulin receptor. *Elife* (2019) 8:e41574. doi: 10.7554/eLife.41574
54. Freychet P. The interactions of proinsulin with insulin receptors on the plasma membrane of the liver. *J Clin Invest* (1974) 54:1020–31. doi: 10.1172/JCI107845

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Irwin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.