New Genes Born-In or Invading Vertebrate Genomes

Carlos Herrera-Úbeda and Jordi Garcia-Fernàndez*

Department of Genetics, Microbiology and Statistics, Faculty of Biology, and Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Spain

Which is the origin of genes is a fundamental question in Biology, indeed a question older than the discovery of genes itself. For more than a century, it was uneven to think in origins other than duplication and divergence from a previous gene. In recent years, however, the intersection of genetics, embryonic development, and bioinformatics, has brought to light that *de novo* generation from non-genic DNA, horizontal gene transfer and, noticeably, virus and transposon invasions, have shaped current genomes, by integrating those newcomers into old gene networks, helping to shape morphological and physiological innovations. We here summarized some of the recent research in the field, mostly in the vertebrate lineage with a focus on protein-coding novelties, showing that the placenta, the adaptive immune system, or the highly developed neocortex, among other innovations, are linked to *de novo* gene creation or domestication of virus and transposons. We provocatively suggest that the high tolerance to virus infections by bats may also be related to previous virus and transposon invasions in the bat lineage.

Keywords: transposon-domestication, virus-invasion, Bex-Tceal, gene_birth, junk_DNA

INTRODUCTION: NEARLY TWO CENTURIES OF HONEYMOONS AND DIVORCES OF SCIENTIFIC FIELDS

That Charles Darwin changed the world with his book on the origin of the species (Darwin, 1859), and quaked human thinking with its "Descent of man" (Darwin, 1871) is beyond doubt. However, he did not know what a gene was. Now, we know what a gene is, but we do not really understand where genes come from. What Darwin was aware, however, was of the relevance of Embryology and, in fact, he based his evolutionary theory in part on the old observations of embryos similarities, regardless of adult similarities, of Karl Enst von Baer (1828). In Darwin words: "community of embryonic structure reveals community of descent," (Darwin, 1859). As a landmark example that validated the evolution/development marriage was the discovery by the Russian embryologist Kowalewski of a notochord in the urochordates (Kowalewski, 1866), noticeably even to non-specialist readers: "One could hardly open a scientific Journal or any popular essay on Natural History without meeting some allusion to the Ascidians as our ancestors" (Agassiz, 1874). Also, Haeckel and his biogenetic law (Haeckel, 1866) and his famous statement "ontogeny recapitulates phylogeny" became very popular. Therefore, at the end of the XIX century, the evolution of species was linked to developmental biology and developmental biology to cell behavior but, however, nothing about the evolution of genes, the ultimate actors of living beings.

The beginning of the XX century was dramatic for Evolution. Genes were (re)discovered, but the nascent new science, genetics, was a very disturbing third party to the evolution/development marriage and thus for comparative biology: genes were believed to have nothing to do with

embryonic development (Bateson, 1922). It is true that mutation rates, allele frequencies, vast amounts of mathematics, and exquisite research in wild animal and plant population insufflate and developed a well-established evolutionary thinking, named Evolutionary Synthesis or Modern Synthesis (Huxley, 1942). Later, several Extended Synthesis that included a variety of new concepts in addition to classic Mendelian Genetics were added (Eldredge and Gould, 1972; Mayr and Provine, 1980). However, where new genes came from was not a central issue for the new or extended synthesis. Still, after Watson and Crick (1953), it was believed that genes from flies and mouse, as an example, had to be extremely different and mainly unrelated. In addition, it was thought that the human genome had to include many more genes than the fly genome, as it was obvious in terms of body plan complexity (from a human point of view).

This "old" view of evolution was shaken in 1978 by a major article that merited a Nobel prize 20 years later, the discovery of the Hox cluster (Lewis, 1978). In 1984 the homeobox was identified molecularly (McGinnis et al., 1984; Scott and Weiner, 1984) and Hox genes from mouse and human, intriguingly similar to the fly genes, discovered. The homeobox, named the "Rosetta Stone" of Biology (Garcia-Fernàndez, 2005) was the launching point of a new discipline and, indeed, the start of the reconciliation of Evolution and Embryology. With Genetics as the major actor of the so-called New-new Synthesis, New extended synthesis or Evo-Devo, the rationale was simple: if developmental genes regulate development, development regulates morphology and physiology and evolution relies on morphology and physiology, then understanding the evolution of developmental genes is the crux to understanding evolution (Baguñà and Garcia-Fernàndez, 2003). Interestingly indeed, the central period of embryonic development, the so-called phylotypic stage, was the stage where crucial (and highly conserved) developmental genes were expressed, suggesting that only earlier and final periods of development were highly evolvable. This was named the hourglass model that related development and evolution (Raff, 1996; Wagner, 2014) and the extended Hox family that is expressed there as the "zootype" (Slack et al., 1993).

In this new field, myriads of reports highlighted the conservation of genes, gene networks among large evolutionary distances and overall number of protein-coding genes in metazoans, creating a paradox at the beginning of the XX century, especially in the animal kingdom: if animals share all this, why are they so different? The solution to the paradox may well rely upon changes of gene regulation in a broad sense, from changes in *cis*-regulation, post-transcriptional regulation, or effects of gene duplication and an ever-growing list of new regulatory elements, with microRNAs, genome topology, epigenomics, long non-coding RNAs or gene editing as the latest players (Heimberg et al., 2008; Wang and Chang, 2011; Irimia et al., 2012; Cavalli and Misteli, 2013; Fromm et al., 2015; Liscovitch-Brauer et al., 2017; Deline et al., 2018). However, not much attention was given to a kind of change in genome functioning which may have an unforeseeable impact in evolution: the birth of new genes.

For half a century, most scientists believed that new protein-coding genes arise as a result of mutations in existing protein-coding genes. It was considered nearly impossible for anything as complex as a functional new protein to arise from scratch (Jacob, 1977). This was puzzling to the understanding of the evolution of metazoans, especially for those major morphological transitions that constitute evolutionary key points bound to radical innovations (Simakov and Kawashima, 2017), from the origins of multicellularity (Stanley, 1973) to the mammal placenta (Roberts et al., 2016). Fortunately, and thanks to the present affordability of new high throughput technologies, an explosion of transcriptomic and genomic data from key phylogenetic species have allowed us to widen our understanding of how such novelties have arisen (Van Oss and Carvunis, 2019).

These disruptive novelties are often accompanied by the apparition of new genes that integrate into the current gene networks (Lavialle et al., 2013; Parrish and Tomonaga, 2018; Zhang et al., 2019; Navas-Pérez et al., 2020; Pascual-Carreras et al., 2020). These new genes (**Figure 1**) can of course be paralogs caused by duplication of genes already present in the genome which could allow the duplicated gene to change its sequence or its regulation (Ohno, 1970; Putnam et al., 2008; Jimenez-Delgado et al., 2009). But they can be fully *de novo* genes, which have no evident homology with any other gene within the studied species or their close relatives: the also known as orphan or taxonomically restricted genes (Singh and Syrkin Wurtele, 2020). We classify here those protein-coding new players found in old genomes from its origin and summarized the knowledge of their impact, with a special focus on its involvement in vertebrate evolution, highlighting the role of virus, virus-related elements, and transposons, when invading vertebrate and mammalian old genomes.

TWO SIDES OF THE SAME COIN: NEW GENES FROM SCRATCH AND NEW GENES FROM ALREADY EXISTING GENIC REGIONS

Although taxonomically restricted genes are a very heterogeneous group, they can be classified according to their origin (**Figure 2**). The true *de novo* genes emerged from ancestrally non-genic regions (wrongly called "junk DNA" in the past). These new genes can arise by mechanisms still poorly understood that involve a genomic region gaining both, transcriptional activity (maybe by transcriptional leaking) and an ORF (open reading frame) in either order (although the ORF is not necessary for non-coding genes). That would seem to be the case of the gene *blitzschnell* found in planarians (Pascual-Carreras et al., 2020) although with the current information we can not infer if the ORF originated first, or instead it was the transcriptional activity. Another mechanism that could incorporate non-genic material to the gene repertoire would be overprinting. In this process, a new ORF is created overlapping an existing one but in a different frame, resulting in two or more genes with overlapping coordinates but with substantially different amino acid sequences (Delaye et al., 2008). Finally, non-genic material can be added to an already existing gene through exonization, in which new exons are generated by random

FIGURE 1 | Schematics of mechanisms for generating *de novo* genes.
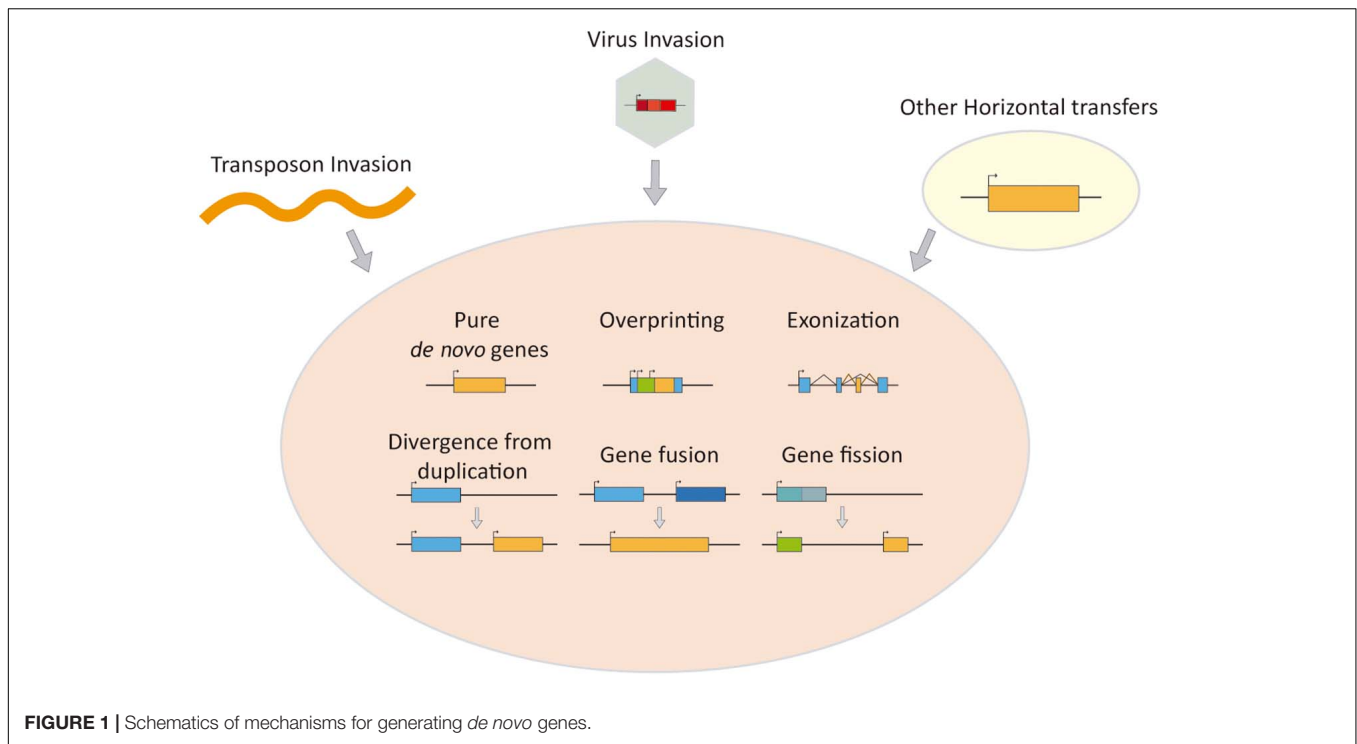
mutations in non-genic DNA, like most alternative exons regulated by NOVA (Irimia et al., 2011), a splicing factor known to be responsible for the inclusion of previously non-extant exons using cryptic splice sites near to the Nova-binding motifs. Or the well know example of the extra exon of TRPV1 (similar to the one present in Laurasiatheria mammals) that has been co-opted in vampire bats to detect hot spots on warm-blooded prey (Gracheva et al., 2011).
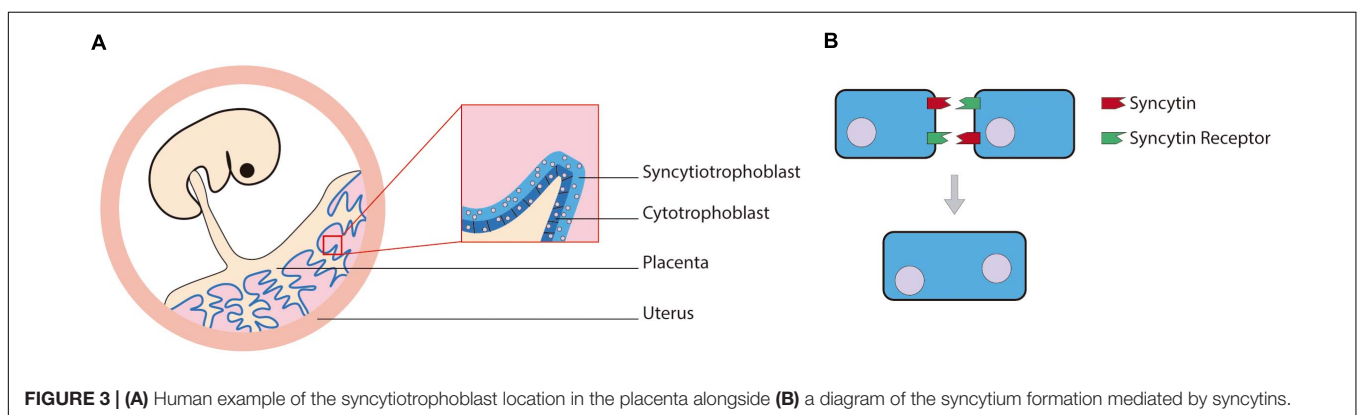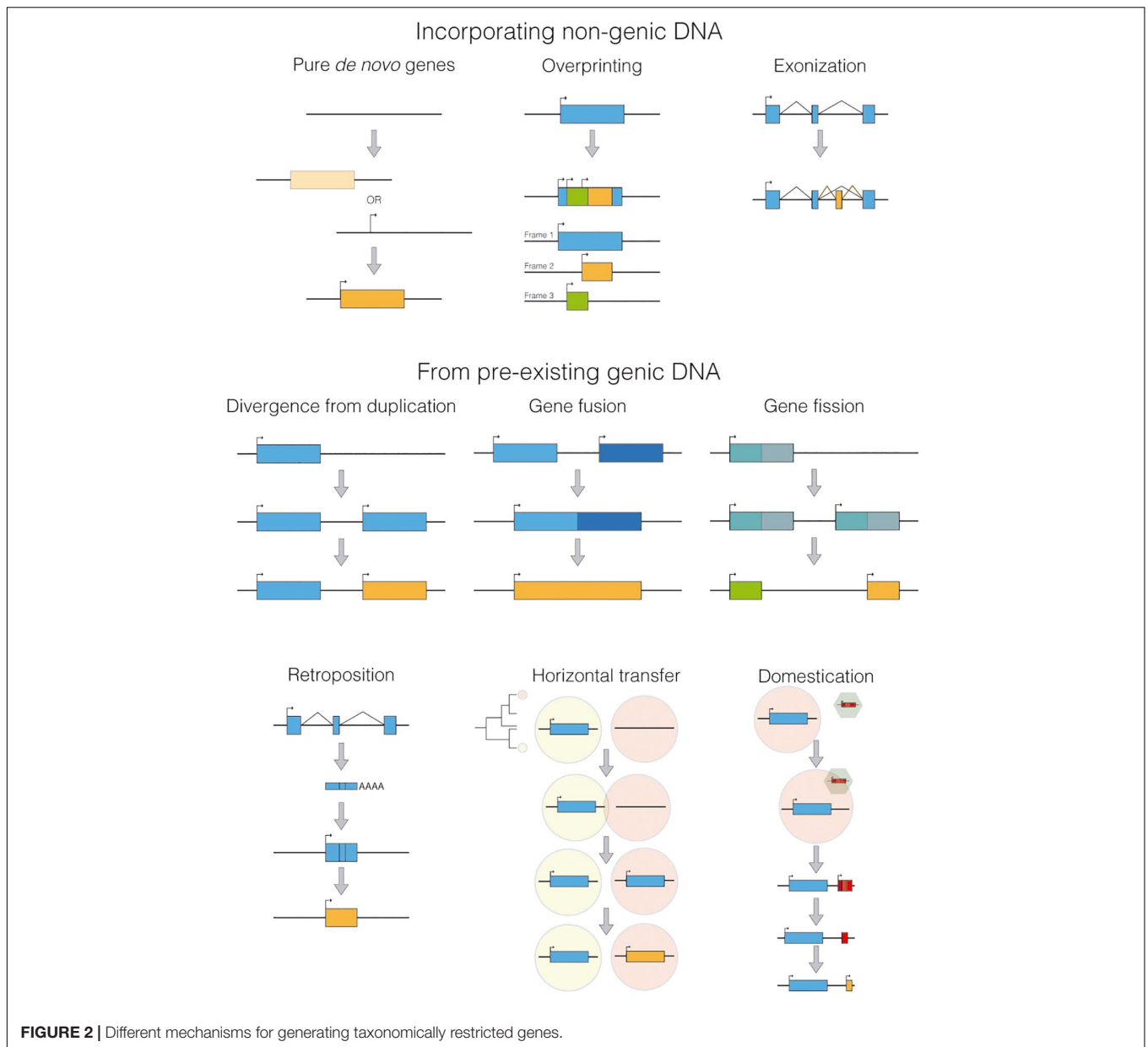
Meanwhile, other taxonomically restricted genes can emerge from several other mechanisms (**Figure 2**), such as extreme divergence from a previous duplication, gene fusion, gene fission, horizontal gene transfer, retroposition or domestication of viral or transposable elements. Remarkably, the process of domestication has been reported to be responsible for disruptive innovations, as independently evolving genes integrate into a new gene framework.

DOMESTICATED RETROVIRAL PROTEINS THROUGH EVOLUTION

Endogenous viral elements are remnants from the integration of retroviruses into the genome and are quite abundant in the animal genomes (Katzourakis and Gifford, 2010). For example, the human genome is formed approximately by 8% of endogenous retroviral sequences (Lander et al., 2001). Throughout evolution, most of the genes from these sequences lose their function, but some of them are captured and "domesticated" in a process called exaptation. Among the examples of domesticated viral genes, the group that has most clearly influenced the evolution of mammals are syncytins

(Lavialle et al., 2013). These captured viral proteins are the product of an envelope gene of retroviruses ancestrally endogenized. The envelope glycoprotein (Env) is crucial in the process of viral entry in enveloped viruses and induces fusion of the virion envelope with the cell plasmatic membrane (Sha et al., 2000). Within the human genome several Env genes can be found, but only two of them have a placental-specific expression and induce the formation of syncytia (**Figure 3**). Being the placenta such a defining organ in placental mammals, the syncytins responsible for its development could be expected to be orthologs in the different species, but that is not the case (Lavialle et al., 2013). Primate and mouse syncytins are not syntenic, and there is evidence pointing to independent capture events in the ancestors of each clade, as well as in the Scincidae genus Mabuya (Cornelis et al., 2017). In fact, in mammals, the different capture events can be linked with the four different main types of placental structures. In the same way, the differences between the lizard placenta present in the genus Mabuya and the mammal placenta can be traced to a completely different capture of Env genes (Cornelis et al., 2017).

Similarly, Arc genes mediate intercellular communication and synaptic plasticity via extracellular vesicles (Parrish and Tomonaga, 2018), and are homologous to the Group-specific antigen (Gag) polyproteins. In retroviruses, capsids are necessary for cellular infection and their assembly is mainly mediated by Gag (Rose et al., 2020). The similarities between Arc and Gag are not restricted to sequence, as Arc is able to spontaneously assemble into a structure that resembles to a capsid (Pastuzyn et al., 2018). In fact, Arc not only forms these capsid-like structures but also encapsulates any mRNA present during their formation. These processes allow the traffic of RNA molecules

FIGURE 2 | Different mechanisms for generating taxonomically restricted genes.

FIGURE 3 | (A) Human example of the syncytiotrophoblast location in the placenta alongside **(B)** a diagram of the syncytium formation mediated by syncytins.

between nervous system cells (Ashley et al., 2018). Regarding the capture and evolution of this viral protein, phylogenetic analyses showed at least two independent capture events that took place in the ancestors of tetrapods and in that of schizophorans (Pastuzyn et al., 2018). In both cases, the co-option of Arc led to similar functions of RNA trafficking in the nervous system. In both lineages the closest hit was a Ty3/gypsy retrotransposon, but tetrapod Arc grouped with the fish Ty3/gypsy, while fly Arc grouped with the insect Ty3/Gypsy, indicating that while sharing significant homology with the Gag protein, it seems to have originated from the Ty3/gypsy retrotransposons from each lineage.

TRANSPOSON-DERIVED NOVELTIES IN VERTEBRATE EVOLUTION

As with the endogenous viral elements, transposable elements can be a source of disruptive innovations. The process of transposition can place sequences near new promoters or generate new fusion proteins. In the evolution of the adaptive immune system of vertebrates, the domestication of the RAG (recombination-activating genes) transposon was instrumental for the V(D)J recombination system, which is a process that makes possible the diversity of antibodies and T cell receptors present in the vertebrate adaptive immune system (Zhang et al., 2019). The current model, supported by the presence of ProtoRAG in the pre-vertebrate amphioxus, is that an ancestral Transib transposon with RAG1-like ORF and terminal inverted repeats, similar to the recombination signal sequences present in V(D)J, captured a RAG2-like ORF to form the ancestral RAG transposon. This event, which took place in an early deuterostome (Carmona and Schatz, 2017), was followed in jawed vertebrates by the insertion of the RAG transposon into a gene encoding an immunoglobulin-domain receptor, among other changes that suppressed RAG transposition activity to finally constitute the V(D)J recombination system.

Another example of a domesticated transposable element is the paired box (PAX) family (Paixão-Côrtes et al., 2015). These homeotic genes were discovered in 1986 (Bopp et al., 1986) and have been proved to be master regulators of the development in metazoans (Dahl et al., 1997). There are several PAX genes, and they were thought to have a monophyletic origin which was initially set at the beginning of metazoan diversification (Hoshiyama et al., 2007). However, PAX-like genes were found in protozoans placing the ancestral PAX down to the pre-metazoan era (Wang et al., 2010). The origin of the ancestral PAX gene is characterized by the domestication of a Tc1/mariner transposon, an ancient and widespread transposon family present in metazoans as well as in plants and protozoans (Garcia-Fernàndez et al., 1993). The Tc1/mariner transposase is similar to the PAX DNA-binding paired domain and its capture was probably posterior to the formation of the other two characteristics domains.

The Pax family represents not the only transcription factor formed by fusion of a transposase domain with another gene. The host-transposase fusion (HTF) genes are a group of genes

that arose most probably from exon shuffling (Cosby et al., 2021), where a transposable element landed within an intron of an existing gene. Once there, the splicing machinery used the splice acceptor/donor sites pre-existing in the integrated transposon. The resemblances in the origin of several HTFs suggest then that DNA transposons are prone to be captured via alternative splicing. In fact, in tetrapod lineages, 106 distinct HTFs have been identified recently from 106 independent events (Cosby et al., 2021).

WHEN CAN NEW GENES INTEGRATE INTO EXISTING NETWORKS?

As we have seen, these kinds of newborn genes are not as rare as initially thought and can shape the evolution of entire groups of animals. But in order to generate a disruptive novelty, a new gene not only has to be born but to be integrated into an already existing gene network. Genes mainly involved in the phylotypic stage, the stage of development shared by all members of the phyla (Duboule, 1994) (the elongated neurula stage in the case of vertebrates) are ancient genes forming highly conserved gene networks (Irie and Sehara-Fujisawa, 2007), where the slightest variation could wreak havoc in the most crucial stages of development. Oppositely, gene networks acting in very early or very late stages of development tend to show more variation and have less conserved or even new genes implicated (Irie and Kuratani, 2011). In this context, the "hourglass model" (Drost et al., 2017), recently also found at the gene regulatory level (Liu et al., 2020), suggest that if a new gene is born, it may preferentially end up functioning at very early (e.g., placenta) or very late stages of the life cycle (e.g., adaptive immunity, synaptic plasticity, body size regulation).

A good example besides the ones already mentioned would be the aforementioned taxonomically restricted gene family, blitzschnell (Pascual-Carreras et al., 2020). Found only within the order Tricladida (planarians), this family is composed of 11 genes and four pseudogenes. It can be further divided into five subfamilies, with one of them organized in a cluster formed by tandem duplication events. Three of the subfamilies are coding and have been reported to regulate the growth/degrowth according to nutrient intake. Thus, these *de novo* genes have been integrated into an evolutionary conserved metabolic network, the insulin/Akt/mTOR network responsible for growth (Saxton and Sabatini, 2017) and other mechanisms at a late stage of the life cycle.

A WHOLE NEW CLUSTER FROM A DOMESTICATED TRANSPOSABLE ELEMENT: IMPACT INTO THE EVOLUTION OF THE EUTHERIAN BRAIN

Recently Navas-Pérez et al. (2020) presented evidence of the expansion of a domestication event into the Bex/Tceal multigenic family, constituting a cluster of 14 genes in the X chromosome of

the ancestor of eutherians, after the divergence of the marsupial-placental clades. The domestication event proposed (**Figure 4**) consists of six steps: (i) A proto-BGW motif (Winter and Ponting, 2005) existed upstream of the alpha-galactosidase (Gla) promoter (P α) in the X chromosome of the ancestor of eutherians and metatherians; (ii) In the eutherian lineage, a retrotranscribed Hnrnph1 was inserted next to the BGW motif and upstream of Gla, creating the Hnrnph2 retrogene; (iii) The region containing the co-opted BGW motif and Hnrnph2 suffered a duplication and HAL1b and L1ME-like retrotransposons were inserted in the vicinity; (iv) The BGW motif and the ORF created by the insertion of the retrotransposons conformed the proto-Bex/Tceal with the YY1 binding site from HAL1b preserved; (v) The BGW motif and the YY1 binding site of a Bex/Tceal gene duplicated upstream of a retrocopy of the Armc10 gene, giving rise to the ArmcX ancestral gene; and (vi) The Bex/Tceal and ArmcX gene families expanded forming the BGW cluster before the diversification of the placental lineage.

Regarding their expression, according to the available data from adult organs, most genes present a tissue-enriched pattern, with the brain be the organ showing the highest expression levels in most of the paralogs. Navas-Pérez et al. (2020) also reported an analysis of the expression via *in situ* hybridization during mouse development, where they observed that especially the Bex genes were highly and widely expressed with Bex3 strongly expressed in the nervous system.

Functional analyses of Bex/Tceal genes have been performed *in vitro*, where Bex3 for example was linked to neuronal physiology, and also through mutant mouse lines. The homozygous mutant lines showed subtle facial differences to the naked eye due to cranial morphology aberrances. The cerebellum and brain showed a reduction in size a fact that may be linked to the behavioral defects observed in the mutant mice. Particularly, the mutants showed impairment in social interactions, nest building, working memory, and object recognition memory. This mutant phenotype coupled with the reported physical interaction of BEX3 with TSC1 (Yasui et al., 2007) could mean that Bex3 is preventing the TSC1/2 complex from interacting with mTORC2, inhibiting this pathway. Thus, Bex3 would be fine-tuning the regulation of the mTOR pathway and its deregulation was suggested to be related to autism spectrum disorder in humans (Ganesan et al., 2019).

FIGURE 4 | Proposed steps for the Bex/Tceal cluster formation.
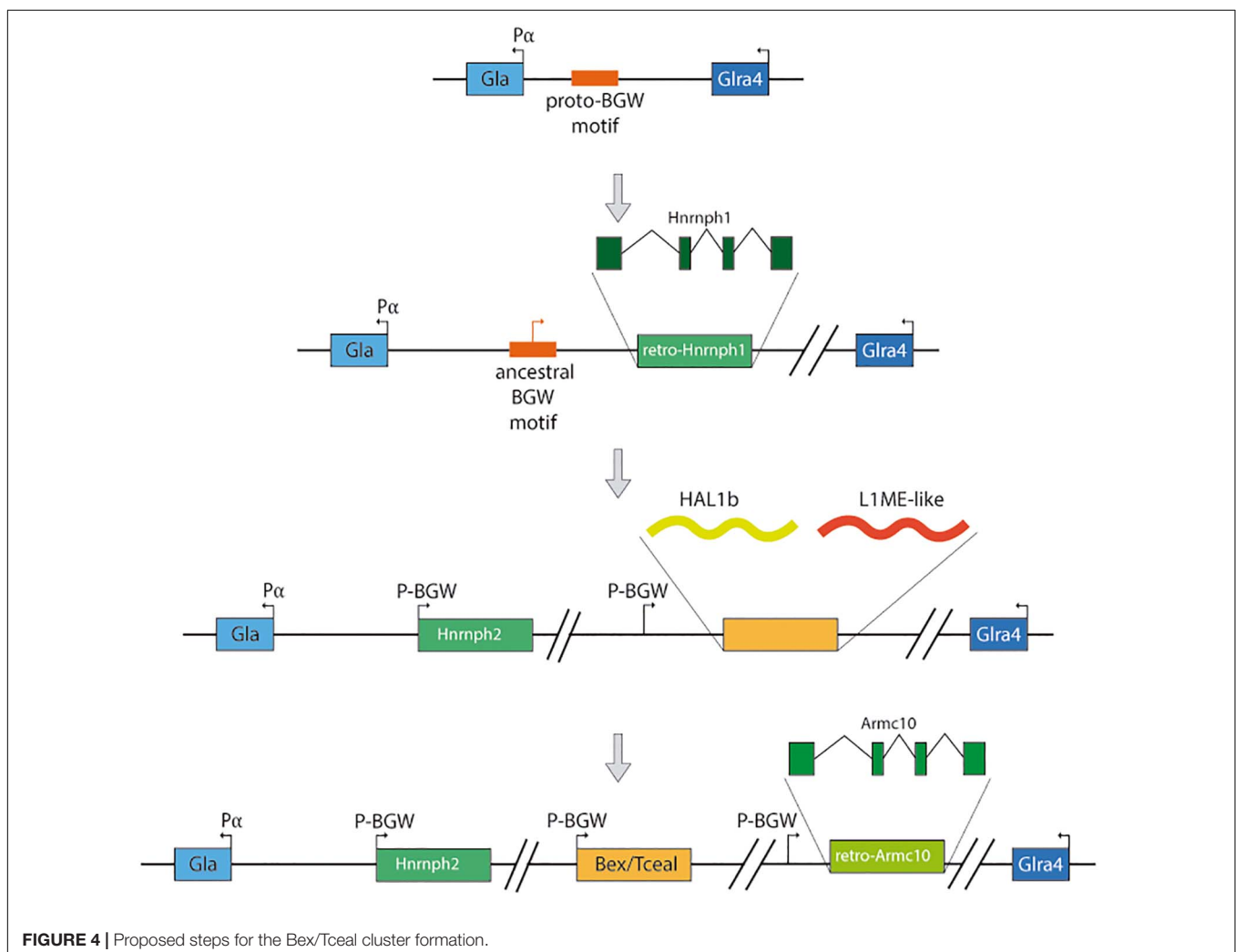
Some of the structural features present in BEX and TCEAL families have been detected in the ancestral transposon HAL1b (Navas-Pérez et al., 2020), which indicates that they were preserved along the domestication process., and a positive selection signature can be found in particular cases, suggesting than ancestral genes went through an adaptative process before the diversification of placental mammals. The differences in the neocortex complexity between eutherian and non-eutherian mammals (Cheung et al., 2010) may thus be linked to genomic novelties that emerged during this transition affecting neural development, and maybe, to the appearance of the Bex/Tceal cluster.

The ability to affect neural proliferation, however, seems to have been acquired at some point during the formation of the cluster and not be intrinsic to the eutherian Bex/Tceal ancestor. This was proved electroporating a synthetically reconstructed version of the ancestral Bex/Tceal protogene, as well as the murine Bex3 and Tceal7 into the neural tube of chicken embryos to serve as a non-eutherian vertebrate environment. Expression of Bex3 and Tceal7 generated a noticeable increase in cell proliferation in the embryonic neural tube, similar to what had been reported in mammalian cell cultures (Calvo et al., 2015). Meanwhile, the reconstructed protogene was not able to trigger cellular proliferation, which could mean that this ability was gained afterward its formation.

FINAL REMARKS: NEW GENES, OLD GENOMES, INNOVATIONS, ADAPTATIONS AND, MAYBE, VIRAL FIGHTING SOLUTIONS

Here, we have reviewed a series of cases in which new born genes have been instrumental for the emergence of novelties, some of them shaping a whole taxon. The *de novo* genes, irrespectively if they have been gained by horizontal transfer event or have been formed within the genome, are an engine of evolution, providing new tools for the regulatory networks of extant genomes. Therefore, disruptive novelties may emerge when new genes integrate into old genomes. In the particular case of vertebrates, it is clear that new exons and whole genes were born by recruitment of non-coding DNA and domestication of transposons and virus, that continuously invaded the genome

trough vertebrate evolution. Particularly noticeable are some periods linked to the appearance of remarkable innovations, such as the origin of the adaptative immune system, the origin of the placenta, or the deployment of a well-developed neural cortex. Further, in the time of COVID pandemics by SARS-CoV-2, it is tempting to speculate, and timely to investigate, if one of the most successful lineages of mammals, the Chiroptera, is particularly remarkable with regards to *de novo* genes. Bats are particularly more tolerant to viral infections than most mammals, including humans. As somewhat expected for their high resistance to viral infections, the bat genome shows gene expansions and deletions related to the immune system gene network (Jebb et al., 2020) but intriguingly also increasing numbers and high diversity of endogenous viral elements[1] and extremely variable numbers and types of transposon remnants, often bat family or bat species-specific (Jebb et al., 2020). All this together is suggestive of a high level of recent events of virus and transposon invasions in the bat lineage. Whether those invasions helped, in fact, the deployment of virus resistance, in an unexplored but exciting similarity to the original function of the CRISPR/Cas9 system in bacteria, the primitive procaryotic acquired adaptative immunity system, is something that remains to be investigated, with the ambition to learn how to fight, in humans, virus related diseases.

AUTHOR CONTRIBUTIONS

CH-Ú and JG-F: conceptualization, writing, review, and editing. JG-F funding acquisition. Both authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

---

[1] https://genome-public.pks.mpg.de

REFERENCES

Agassiz, L. (1874). *Evolution and Permanence of Type*. United Kingdom: Universidad de Oxford.

Ashley, J., Cordy, B., Lucia, D., Fradkin, L. G., Budnik, V., and Thomson, T. (2018). Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. *Cell* 172, 262–274.e11. doi: 10.1016/j.cell.2017.12.022

Baguñà, J., and Garcia-Fernàndez, J. (2003). Evo-Devo: the Long and Winding Road. *Int. J. Dev. Biol.* 47, 705–713. doi: 10.1387/ijdb.14756346

Bateson, W. (1922). Evolutionary faith and modern doubts. *Nature* 109, 553–556. doi: 10.1038/109553a0

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene paired and

in functionally related genes of Drosophila. *Cell* 47, 1033–1040. doi: 10.1016/0092-8674(86)90818-4

Calvo, L., Anta, B., López-Benito, S., Martín-Rodriguez, C., Lee, F. S., Pérez, P., et al. (2015). Bex3 dimerization regulates NGF-dependent neuronal survival and differentiation by enhancing trkA gene transcription. *J. Neurosci.* 35, 7190–7202. doi: 10.1523/JNEUROSCI.4646-14.2015

Carmona, L. M., and Schatz, D. G. (2017). New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J.* 284, 1590–1605. doi: 10.1111/febs.13990

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. *Nat. Struct. Mol. Biol.* 20, 290–299. doi: 10.1038/nsmb.2474

Cheung, A. F. P., Kondo, S., Abdel-Mannan, O., Chodroff, R. A., Sirey, T. M., Bluy, L. E., et al. (2010). The subventricular zone is the developmental milestone of

a 6-layered neocortex: comparisons in metatherian and eutherian mammals. *Cereb. Cortex* 20, 1071–1081. doi: 10.1093/cercor/bhp168

Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O. A., Meurice, G., et al. (2017). An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proc. Natl. Acad. Sci. U. S. A.* 114, E10991–E11000. doi: 10.1073/pnas.1714590114

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., et al. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* 371:eabc6405. doi: 10.1126/science.abc6405

Dahl, E., Koseki, H., and Balling, R. (1997). Pax genes and organogenesis. *BioEssays* 19, 755–765. doi: 10.1002/bies.950190905

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex (1st ed.).* London: John Murray.

Delaye, L., DeLuna, A., Lazcano, A., and Becerra, A. (2008). The origin of a novel gene through overprinting in *Escherichia coli. BMC Evol. Biol.* 8:31. doi: 10.1186/1471-2148-8-31

Deline, B., Greenwood, J. M., Clark, J. W., Puttick, M. N., Peterson, K. J., and Donoghue, P. C. J. (2018). Evolution of metazoan morphological disparity. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8909–E8918. doi: 10.1073/pnas.1810575115

Drost, H.-G., Janitza, P., Grosse, I., and Quint, M. (2017). Cross-kingdom comparison of the developmental hourglass. *Curr. Opin. Genet. Dev.* 45, 69–75. doi: 10.1016/j.gde.2017.03.003

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* 1994, 135–142. doi: 10.1242/dev.1994. Supplement.135

Eldredge, N., and Gould, S. (1972). "Punctuated Equilibria: an Alternative to Phyletic Gradualism," In *Models in Paleobiology*, ed. T.J.M. Schopf (San Francisco: Freeman, Cooper and Co), 82–115.

Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., King, B. L., et al. (2015). A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annu. Rev. Genet.* 49, 213–242. doi: 10.1146/annurev-genet-120213-092023

Ganesan, H., Balasubramanian, V., Iyer, M., Venugopal, A., Subramaniam, M. D., Cho, S. G., et al. (2019). mTOR signalling pathway - A root cause for idiopathic autism?. *BMB Rep.* 52, 424–433. doi: 10.5483/BMBRep.2019.52.7.137

Garcia-Fernàndez, J. (2005). The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* 6, 881–892. doi: 10.1038/nrg1723

Garcia-Fernàndez, J., Marfany, G., Baguñà, J., and Saló, E. (1993). Infiltration of mariner elements. *Nature* 364:109. doi: 10.1038/364109a0

Gracheva, E. O., Cordero-Morales, J. F., González-Carcacía, J. A., Ingolia, N. T., Manno, C., Aranguren, C. I., et al. (2011). Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* 476, 88–92. doi: 10.1038/nature10245

Haeckel, E. (1866). *Generelle Morphologie der Organismen*. Berlin: G. Reimer, 1036. doi: 10.1515/9783110848281

Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. J., and Peterson, K. J. (2008). MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acade. Sci. U. S. A.* 105, 2946–2950. doi: 10.1073/pnas.0712259105

Hoshiyama, D., Iwabe, N., and Miyata, T. (2007). Evolution of the gene families forming the Pax/Six regulatory network: isolation of genes from primitive animals and molecular phylogenetic analyses. *FEBS Lett.* 581, 1639–1643. doi: 10.1016/j.febslet.2007.03.027

Huxley, J. (1942). *Evolution, the Modern Synthesis*. Pennsylvania: Universidad Estatal de Pensilvania.

Irie, N., and Kuratani, S. (2011). Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* 2, 1–8. doi: 10.1038/ncomms1248

Irie, N., and Sehara-Fujisawa, A. (2007). The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. *BMC Biol.* 5:1. doi: 10.1186/1741-7007-5-1

Irimia, M., Denuc, A., Burguera, D., Somorjai, I., Martin-Duran, J. M., Genikhovich, G., et al. (2011). Stepwise assembly of the Nova-regulated alternative splicing network in the vertebrate brain. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5319–5324. doi: 10.1073/pnas.1012333108

Irimia, M., Denuc, A., Ferran, J. L., Pernaute, B., Puelles, L., Roy, S. W., et al. (2012). Evolutionarily conserved A-to-I editing increases protein stability of the alternative splicing factor Nova1. *RNA Biol.* 9, 12–21. doi: 10.4161/rna.9.1.18387

Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134

Jebb, D., Huang, Z., Pippel, M., Hughes, G. M., Lavrichenko, K., Devanna, P., et al. (2020). Six reference-quality genomes reveal evolution of bat adaptations. *Nature* 583, 578–584. doi: 10.1038/s41586-020-2486-3

Jimenez-Delgado, S., Pascual-Anaya, J., and Garcia-Fernandez, J. (2009). Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. *Brief. Funct. Genomic. Proteomic.* 8, 266–275. doi: 10.1093/bfgp/elp029

Katzourakis, A., and Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet.* 6:1001191. doi: 10.1371/journal.pgen.1001191

Kowalewski, A. O. (1866). Entwicklungsgeschichte der einfachen Ascidien. *Mem I'Acad St Petersbourg Ser.* 7, 1–9.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., et al. (2013). Paleovirology of "syncytins", retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20120507. doi: 10.1098/rstb.2012.0507

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* 276, 565–570. doi: 10.1038/276565a0

Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., Ziv, T., et al. (2017). Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* 169, 191–202.e11. doi: 10.1016/j.cell.2017.03.025

Liu, J., Frochaux, M., Gardeux, V., Deplancke, B., and Robinson-Rechavi, M. (2020). Inter-embryo gene expression variability recapitulates the hourglass pattern of evo-devo. *BMC Biol.* 18:129. doi: 10.1186/s12915-020-00842-z

Mayr, E., and Provine, W. B. (1980). *The Evolutionary Synthesis?: Perspectives On The Unification Of Biology*. Cambridge: Harvard University Press.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* 308, 428–433. doi: 10.1038/308428a0

Navas-Pérez, E., Vicente-García, C., Mirra, S., Burguera, D., Fernàndez-Castillo, N., Ferrán, J. L., et al. (2020). Characterization of an eutherian gene cluster generated after transposon domestication identifies Bex3 as relevant for advanced neurological functions. *Genome Biol.* 21:267. doi: 10.1186/s13059-020-02172-3

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer. doi: 10.1007/978-3-642-86659-3

Paixão-Côrtes, V. R., Salzano, F. M., and Bortolini, M. C. (2015). Origins and evolvability of the PAX family. *Semin. Cell Dev. Biol.* 44, 64–74. doi: 10.1016/j.semcdb.2015.08.014

Parrish, N. F., and Tomonaga, K. (2018). A Viral (Arc)hive for Metazoan Memory. *Cell* 172, 8–10. doi: 10.1016/j.cell.2017.12.029

Pascual-Carreras, E., Marin-Barba, M., Herrera-Úbeda, C., Font-Martín, D., Eckelt, K., de Sousa, N., et al. (2020). Planarian cell number depends on Blitzschnell, a novel gene family that balances cell proliferation and cell death. *Development* 147:dev.184044. doi: 10.1242/dev.184044

Pastuzyn, E. D., Day, C. E., Kearns, R. B., Kyrke-Smith, M., Taibi, A. V., McCormick, J., et al. (2018). The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* 172, 275–288.e18. doi: 10.1016/j.cell.2017.12.024

Putnam, N. H., Butts, T., Ferrier, D. E. K., Furlong, R. F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–1071. doi: 10.1038/nature06967

Raff, R. A. (1996). *The Shape of Life*. Chicago: University of Chicago Press. doi: 10.7208/chicago/9780226256573.001.0001

Roberts, R. M., Green, J. A., and Schulz, L. C. (2016). The evolution of the placenta. *Reproduction* 152, R179–R189. doi: 10.1530/REP-16-0325

Rose, K. M., Hirsch, V. M., and Bouamr, F. (2020). Budding of a retrovirus: Some assemblies required. *Viruses* 12:1188. doi: 10.3390/v12101188

Saxton, R. A., and Sabatini, D. M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. *Cell* 168, 960–976. doi: 10.1016/j.cell.2017.02.004

Scott, M. P., and Weiner, A. J. (1984). Structural relationships among genes that control development: Sequence homology between the antennapedia, ultrabithorax, and fushi tarazu loci of Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 81, 4115–4119. doi: 10.1073/pnas.81.13.4115

Sha, M., Lee, X., Li, X., Li, X., Veldman, G. M., Finnerty, H., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi: 10.1038/35001608

Simakov, O., and Kawashima, T. (2017). Independent evolution of genomic characters during major metazoan transitions. *Dev. Biol.* 427, 179–192. doi: 10.1016/j.ydbio.2016.11.012

Singh, U., and Syrkin Wurtele, E. (2020). How new genes are born. *Elife* 9:e55136. doi: 10.7554/eLife.55136

Slack, J. M. W., Holland, P. W. H., and Graham, C. F. (1993). The zootype and the phylotypic stage. *Nature* 361, 490–492. doi: 10.1038/361490a0

Stanley, S. M. (1973). An Ecological Theory for the Sudden Origin of Multicellular Life in the Late Precambrian. *Proc. Natl. Acad. Sci.U. S. A.* 70, 1486–1489. doi: 10.1073/pnas.70.5.1486

Van Oss, S. B., and Carvunis, A.-R. R. (2019). De novo gene birth. *PLoS Genet.* 15:e1008160. doi: 10.1371/journal.pgen.1008160

von Baer, K. E. (1828). *Über Entwickelungsgeschichte der Thiere. Beobachtung und Reflexion.* Königsberg: Bei den Gebrüdern Bornträger.

Wagner, G. P. (2014). *Homology, Genes, and Evolutionary Innovation.* Princeton: Princeton University Press. doi: 10.1515/9781400851461

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, Y. T., Pan, Y. J., Cho, C. C., Lin, B. C., Su, L. H., Huang, Y. C., et al. (2010). A novel pax-like protein involved in transcriptional activation of cyst wall protein genes in Giardia lamblia. *J. Biol. Chem.* 285, 32213–32226. doi: 10.1074/jbc.M110.156620

Watson, J. D., and Crick, F. H. C. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171, 737–738. doi: 10.1038/171737a0

Winter, E. E., and Ponting, C. P. (2005). Mammalian BEX, WEX and GASP genes: coding and non-coding chimaerism sustained by gene conversion events. *BMC Evol. Biol.* 5:54. doi: 10.1186/1471-2148-5-54

Yasui, S., Tsuzaki, K., Ninomiya, H., Floricel, F., Asano, Y., Maki, H., et al. (2007). The TSC1 gene product hamartin interacts with NADE. *Mol. Cell. Neurosci.* 35, 100–108. doi: 10.1016/J.MCN.2007.02.007

Zhang, Y., Cheng, T. C., Huang, G., Lu, Q., Surleac, M. D., Mandell, J. D., et al. (2019). Transposon molecular domestication and the evolution of the RAG recombinase. *Nature* 569, 79–84. doi: 10.1038/s41586-019-1093-7