

The Developmental Gene Hypothesis for Punctuated Equilibrium: Combined Roles of Developmental Regulatory Genes and Transposable Elements

Emily L. Casanova* and Miriam K. Konkel

Theories of the genetics underlying punctuated equilibrium (PE) have been vague to date. Here the developmental gene hypothesis is proposed, which states that: 1) developmental regulatory (DevReg) genes are responsible for the orchestration of metazoan morphogenesis and their extreme conservation and mutation intolerance generates the equilibrium or stasis present throughout much of the fossil record and 2) the accumulation of regulatory elements and recombination within these same genes—often derived from transposable elements—drives punctuated bursts of morphological divergence and speciation across metazoa. This two-part hypothesis helps to explain the features that characterize PE, providing a theoretical genetic basis for the once-controversial theory.

1. A Genetic Hypothesis to Explain the Theory of Punctuated Equilibrium

Despite his support for phyletic gradualism, Darwin himself acknowledged the abrupt appearance of new species within the fossil record.^[1] And although this observation was partly due to the infancy of 19th century paleontology, more than a century and a half later we recognize that not all aspects of cladogenesis can be accounted for by a lack of detailed stratigraphical study and instead adhere to the theory of punctuated equilibrium.

Detractors of punctuated equilibrium have previously chided the theory as a mechanism with “no scientific use,” as they claimed it could not be tested at the genetic level.^[2] However, with the advent of large-scale genomic sequencing and extensive molecular and computational study of numerous genomes, our wealth of available data has grown substantially since the early

battles over Gould and Eldredge’s theory.^[3] Since that time we have been able to study not only transposable elements (TE) (i.e., “selfish DNA”) and their roles in molecular evolution, but also the subset of genes responsible for the regulation of morphogenesis reflected in the fossil record.

Here, we propose that there is a native genetic complement to TE insertions leading to features of punctuated equilibria in both the vertebrate and invertebrate fossil records. Specifically, this complement lies within the developmental regulatory (DevReg) genes responsible for morphogenesis and their unique mutational patterns, as well as the elements that regulate their

expression. The genes’ relative mutation intolerance suggests a means by which morphology is actively conserved even in the face of exaggerated TE activity.^[4] Yet they also exhibit a clear history of TE insertion that is strongly correlated to the presence of conserved noncoding elements (CNE) and changes to gene regulation and phenotype by acting as promoters, enhancers, repressors, terminators, insulators, and post-transcriptional effectors.^[4–7] In addition, TE-derived RNA may act as direct regulators of these important developmental genes, strongly reminiscent of Barbara McClintock’s “controlling elements.”^[7,8] Alongside the TE-Thrust hypothesis^[9] and the TE-epigenomic arms race proposed by Zeh et al.,^[10] as well as the large body of work spearheaded by Eric Davidson^[11,12] concerning the roles of the regulome in speciation, we believe the developmental gene hypothesis may help explain the long periods of active morphological stasis within the fossil record bookended by rare TE-associated mutational events that have long lasting effects on phenotype and have been tightly conserved over evolutionary time^[4,13] (Figure 1).

2. Developmental Genes and Their Regulome Are Integral to the Phyletic Period of Development

Since the morphological features of an animal are the product of its developmental process, and since the developmental process in each animal is encoded in its species-specific regulatory genome, then change in animal form during evolution is the consequence of change in genomic regulatory programs for development.^[11]

Although evolution of the eukaryotic exome has been relatively conserved across time, evolution of the regulome seems to

Dr. E. L. Casanova
 Department of Biomedical Sciences
 University of South Carolina School of Medicine at Greenville
 200A Patewood Dr., Greenville, SC 29615, USA
 E-mail: casanove@greenvillemed.sc.edu;
 emily.casanova@prismahealth.org; scienceoveracuppa@gmail.com

Dr. M. K. Konkel
 Department of Genetics
 Clemson Center for Human Genetics
 Biomedical Data Science and Informatics Program
 Clemson University
 105 Collings St., Clemson, SC 29631, USA

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/bies.201900173>

DOI: 10.1002/bies.201900173

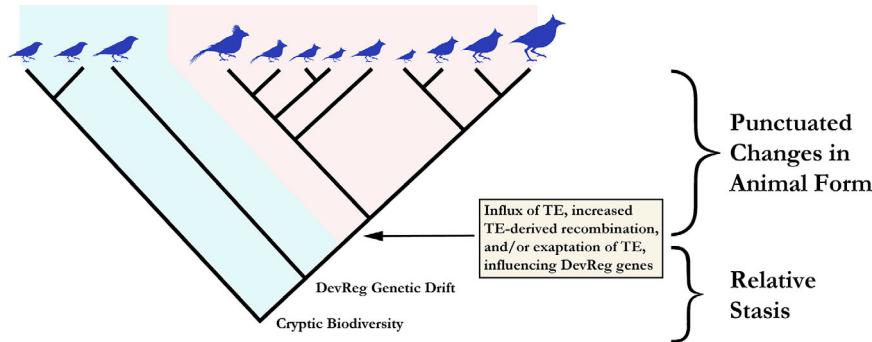


Figure 1. A schematic for the developmental gene hypothesis. Evolution of developmental regulatory (DevReg) genes is strongly conserved, though is not impervious to genetic drift. Changes primarily within non-DevReg genes may result in cryptic biodiversity (i.e., two species that cannot be morphologically differentiated), which at the genetic level are deemed separate species. Eventually, flurries of transposition events, exaptation of more ancient transposable elements (TE), and/or recombination as a result of TE-influenced genomic instability in or surrounding DevReg genes—together with environmental influences—may lead to a “punctuated” series of rapid morphological change. These changes in morphology are measurable within the fossil record and have given rise to the theory of punctuated equilibrium^[3] as well as the developmental gene hypothesis presented here.

account for considerable interspecies variation.^[11,13] The regulome is both a seat of rapid evolutionary development, as in the case of species-specific TE regulatory exaptation, and extreme conservation in the form of conserved (CNE) and ultraconserved noncoding elements (UCNE).^[14–16] UCNE networks are likewise strongly conserved across species in the form of gene regulatory blocks surrounding important developmental genes.^[17]

2.1. The Evolutionary Rate of Conserved Noncoding Elements Has Slowed

The rate of substitutions within protein-coding regions has been relatively constant across vertebrates, yet CNE have experienced significantly lower mutation rates since the dawn of amniotes.^[15] Interestingly, prior to this mutational retardation, tetrapod ancestors experienced a massive expansion of sequences leading to CNE, presumably related to the evolution of terrestrial life.^[15] Many of the UCNE are located in or near transcriptional and developmental genes, as are many of the younger less conserved noncoding elements unique to later phylogenetic branches, suggesting a strong directional selection for the retention and exaptation of potential noncoding elements in these regions.^[15,18] While many ancient coding and noncoding ultraconserved elements present in bony fish and their descendants occurred in roughly equal proportions targeting splice sites, intronic, and intergenic regions (and to a lesser degree coding sequences and untranslated regions), more recent elements in amniotes, theria, or eu-theria are derived primarily from introns and gene-poor sites.^[15]

Invertebrate genomes likewise exhibit conservation of non-coding elements surrounding key developmental genes, a feature they share with vertebrates. However, these CNE appear to be largely unique to each taxon and likely play an integral role in defining the general body plans particular to the major branches.^[19–21]

In contrast to the invertebrates we see today, however, it has been proposed that during the Cambrian period there was a much wider array of body plans, that is, the so-called “Cambrian explosion,” a concept which Gould made famous

in his 1989 book, *Wonderful Life: The Burgess Shale and the Nature of History*.^[22] It is now believed that rather than a single explosion, the Ediacaran–Cambrian interval experienced a series of significant radiations more akin to successive fireworks.^[23] Nevertheless, the variety in bauplan present in the Burgess Shale Lagerstätte, along with features of “interchangeable parts” of some of these strange creatures as suggested by Gould, indicates that CNE may have been experiencing a significant evolutionary flux during these early periods in metazoan evolution prior to entrenchment. Likewise, fluctuation in regulatory elements provides a theoretical means for the variety present during those early periods, as well as the few but highly conserved body plans and extreme CNE mutation intolerance we see in extant organisms today.

2.2. Developmental Regulatory Genes Share Unique Features

We have recently found that DevReg genes in humans, for example, maintain a particularly high relative density of intronic CNE compared to most other genes.^[4] These genes generally produce larger proteins but the genes themselves also tend to be quite long, which can be primarily attributed to intron size. Their sensitivity to variation is also significant and is evidenced by reduced numbers of common/rare single nucleotide variants and copy number variants, suggesting they may be strongly conserved.^[4,24] Interestingly, these genes also maintain more extensive protein–protein interaction networks, suggesting many of them may be foundational and function as molecular hubs.^[4,25] Data indicates that many of these genes are also targets of various forms of regulatory RNA derived from TE, adding yet another level of complexity to this broad developmental network.^[8,26,27]

2.3. Developmental Regulatory Genes Play a Critical Role during the Phylotypic Period of Development

Many DevReg genes are expressed during the phylotypic period of development, suggesting a reason for their conservation.^[28]

Box 1
Highlights

- The TE (transposable element)-thrust hypothesis^[9] and TE epigenomic arms race by Zeh et al.^[10] do not address the active role developmental regulatory genes play as a barrier to TE insertion and other mutations due to loss-of-function sensitivity.
- Developmental regulatory genes maintain significant density of intronic TE and conserved noncoding elements (CNE) in human.^[4]

- CNE are often derived from TE insertion events.^[5,13]
- Young CNE occur in the same regions as older CNE, indicating selective pressure for their aggregation in these regions.^[15,18]
- The developmental gene hypothesis proposes that patterns of resistance to and retention of TE within DevReg genes further supports the genetic basis of punctuated equilibrium.

Box 2
Outstanding Questions

- How has the evolution of developmental regulatory genes varied across species, genera, class, phyla, and even kingdoms?
- When comparing two closely related genomes, will accelerated evolution be consistently apparent in and near developmental regulatory genes when morphology is significantly divergent between the two species?

- Do older classes of developmental regulatory genes display a different evolutionary pattern than younger classes and are the latter more tightly conserved?
- How have transposable elements affected the evolving regulome of developmental regulatory genes across vertebrates and invertebrates?

This conservation is modeled in what Arthur^[29] called “the Fisher principle,” referring to the inverse relationship between the extent to which a mutation affects phenotype and the likelihood it will be selectively advantageous:

It seems reasonable to suppose that genes controlling early developmental decisions, such as which end of the embryo is anterior, will be subject, on average, to mutations with more major phenotypic effects than genes controlling later developmental processes, such as the production of mammalian hair... . If so, then under the Fisher

principle early genes will evolve more slowly than their later counterparts ... and, under certain assumptions ... this will mean that the earlier stages themselves will be more evolutionarily conserved.^[29]

Although Arthur^[29] was correct about weak genic conservation in late phenotypic stages, genes expressed in very early embryogenesis are in fact poorly conserved and small in size, an observation that explains significant variability amongst the zygotes of different species.^[30] Instead, genes expressed during the middle stages of embryogenesis (i.e., the phylotypic period)

Box 3
Glossary

- *Punctuated Equilibrium (PE):* A theory that posits the fossil record is characterized by episodes of rapid speciation bookended by broader periods of morphological stasis.
- *Developmental Regulatory (DevReg) Genes:* Genes involved in regulating morphogenesis and embryonic development.
- *Transposable Elements (TE):* Sequences of DNA that move about the genome and replicate either through transposition (cut-and-paste) or retrotransposition (copy-and-paste) mechanisms.
- *Ultraconserved Noncoding Elements (UCNE):* Ancient non-coding elements that are strongly conserved across verte-

- brates or older phyla and are presumably involved in regulation.
- *Conserved Noncoding Elements (CNE):* Noncoding elements involved in regulation that are strongly conserved across species (e.g., placentals or mammals) but more limited in span compared to ultraconserved noncoding elements.
- *Regulome:* The full complement of regulatory components in the cell.
- *Phylotypic Period:* The most conserved period of prenatal development during which an animal most closely resembles other species of its phylum.

are in fact more heavily conserved and fit the profiles of the long complex genes described here. This pattern of conservation is referred to as the “hourglass model,” in which genes involved in organogenesis share greatest overlap across diverse species (For example, see **Figure 2**).^[28] Various DevReg genes show the highest transcriptional expression during the phylotypic period in both mouse and human. They are typically expressed across multiple tissue types, potentially leading to shared changes across many organ systems when germline mutations do occur.^[31] This can be seen, for example, in many human-specific mutation events in DevReg genes that often lead not only to intellectual disability but multiple congenital anomalies.^[32] In addition, many genes associated with rare human genetic diseases share similar structural, functional, and conservation profiles as those reported by Casanova et al.,^[4] reinforcing the concept of their importance in development and human health.^[24] Although these are all extreme pathological examples, it is nevertheless easy to envision how subtler less detrimental changes to developmental gene function could lead to adaptive changes across organ systems over larger timescales, allowing evolution to act on specific aspects of morphology, not in isolation, but in concert.^[33]

Although there are considerable differences in physical form during the phylotypic stage across higher order taxa (e.g., insects, crustaceans, and vertebrates), within these taxa there is significant conservation during early organogenesis, which, as mentioned earlier, appears to be reflected in conservation of unique regulatory elements in and around developmental genes.^[19,20] Many CNE are *cis*-regulatory elements (e.g., enhancers) and are believed to reflect parallel evolution of alternative elements influencing expression of these genes and ultimately variations in body plan that differentiate the major taxa. Therefore, although most CNE are not conserved across vertebrates and invertebrates, networks that regulate these vital developmental genes are common to the metazoa.^[20,37]

In general, the phylotypic period is relatively conserved, yet it is during this stage in which minor progressive changes to the regulation of DevReg genes orchestrating it may lead to the inception of heterochronic tissue development and changes in morphology that ultimately differentiate species across the fossil record.^[11,38] In addition, while there is indeed notable developmental variability during this period between human and zebrafish embryos, for instance, similarities in overall morphology are also visibly apparent suggesting a conserved bauplan that is likely reflected in coding and noncoding ultraconserved elements within the genome.^[15,20,39]

2.4. Is Stasis an Active Phenomenon?

Gould and Eldredge^[40] once wrote:

As the most important change in research practice provoked by punctuated equilibrium, stasis has now exited from its closet of non-definition to become a subject of quantitative investigation in all major fossil groups ... Moreover, because species often maintain stability through such intense climatic changes as glacial cycling, stasis must be viewed as an active phenomenon, not a passive response to unaltered environments.

In this instance, the variation intolerance of DevReg genes provides an active safeguard against mutational events. Surprisingly, CNE within long genes such as DevReg genes are not mutation cold spots despite their functional sensitivity. Instead, such mutations are typically incompatible with life and often lead to miscarriage or serious functional impairment preventing further inheritance.^[41,42]

Nevertheless, developmental genes may be key sites for species-specific evolution. As an example from human and chimpanzee: genes with longer introns (e.g., DevReg genes) are sequentially more divergent than those with shorter introns (K_i = substitutions per intronic site), suggesting they are sites for accelerated evolution between our two species despite their overall mutation sensitivity.^[43,44]

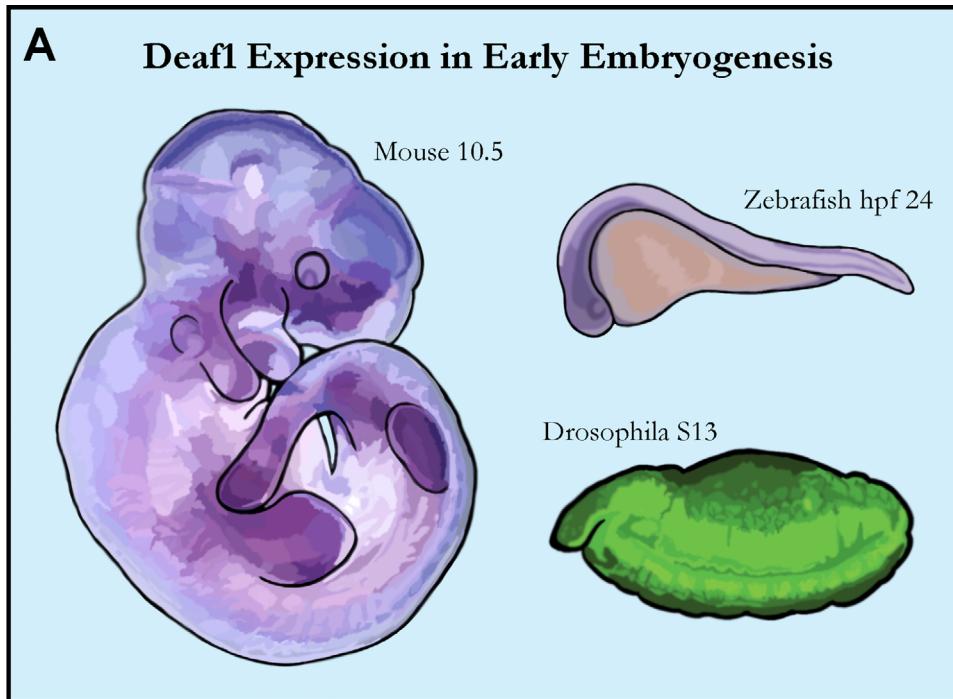
While genetic isolation and ultimately hybrid dysgenesis ensure speciation, the evolution of DevReg genes plays a fundamental role in the type of morphological divergence that originally inspired the Linnaean classification system.^[45] Of course, not all species can be divided according to morphological divergence (i.e., cryptic biodiversity) and, similarly, DevReg genes are not completely impervious to drift.^[46,47] However, following the recognition of this disparity, it has subsequently been demonstrated that morphological evolution correlates better with regulatory gene divergence rather than with overall rates of molecular evolution.^[48–50] Thus, the variation intolerance we see in DevReg genes appears to be responsible for the significant periods of morphological stasis present in the fossil record. Meanwhile, other gene groups evolve at a more rapid, and perhaps more constant, rate.

3. Transposons and the Evolving Regulome Drive Punctuated Changes in Animal Form

Since the publication of the theory of punctuated equilibrium, detractors have sometimes criticized its dependence on an imperfect fossil record and the lack of an underlying genetic theory.^[1] Oliver and Greene^[9] and Zeh et al.^[10] made significant strides by proposing genetic and epigenetic mechanisms that help explain the long periods of stasis bookended by periods of comparatively rapid change. Likewise, the body of work by Eric Davidson clearly outlines numerous regulatory mechanisms leading to morphological divergence across species.^[11,12] These works, however, do not recognize the active role DevReg genes play in maintaining morphological stasis, their propensity to collect TE-derived non-coding elements over time, or their tendency to be regulated by TE-derived RNA.^[8] All of these presumably underlie changes in gene function and morphology relevant to the patterns within the fossil record first recognized by Gould and Eldredge.^[3] The developmental gene hypothesis helps to fill in these important gaps.

3.1. Conserved Noncoding Elements Are Often Derived from Transposable Elements

CNE are key players in the evolution of DevReg genes and organismal morphology. What is more, DevReg genes in humans not only contain an increased density of CNE, they also have a higher density of TEs.^[4] In fact, intronic CNE and TE content across the



B BLASTp

Drosophila	Query_179519	1	MEQVDSSTELHLNRKDLAALAEDEVV	KEEVILESSHHHHHHHHOLDTKVRMVTSSSNDNSGSGGASGGTSGAGGGN	76
Mouse	Query_179521	1	MEDSDS-NAKQLGLAEEAAAVAAA[12]AEEPVLRSRDEDSEEDA	DSEAERTRRVTAVALMAAESGHMDMGEALPSPD	87
Zebrafish	Query_179520	1	MDATEPATKELELGDAESEASATG-	--EPVGS-----DTSEAEAVATHMTVMG---EAGTIDI-TESLPNPD	59
	Query_179519	77	GGGGV-----VSVPVSLPIPGMSMITGTTFNVITPDQLPPHF[11]	LSG-STVSMGNDL[24]NTTATNTIGLNLHDGSGS	175
	Query_179521	88	EAAAAaaFAEVTTTVANVGSADDNVFTTSVANAASISGHV	LSGRITALQIGDSL NTEKATLIVVHTDGSIV	159
	Query_179520	60	DAETA---FAEVTTAVTVGDVQSSDDSVFTSAVATATSIPEHV	LITGRATLQIGNTL STQKATLIVVHTDGSIV	128
	Query_179519	176	NNSHDSLATLEHAAGGASGVGGGGGGTGGSSGWSENPSQTQHNEVFq	IRCKTTCAELYRSKLGSGGRGRCVKYKDKWHTP	255
	Query_179521	160	ETTGILKGPAAPLTGPQSPPTPLAIPQGEKGCTKYNWNPDSVYSEL-P	-VRCRNISGTYLKSRILGSGGRCIKQGENWYSP	238
	Query_179520	129	DATSLKATGTPMTPGQPNTPLASGHDKDVSKYNWDPSPVYDNELP	-VRCRNTSGLLYKNRLGSGKGRCIKHNNSWYTP	207
	Query_179519	256	SEFHVCGRGSSKDWKRISIKYGGKSLSLIDEGTILTIPHATNCSTVCCDDEA	CE---SASGPVRLFPPYKRRKRNqt	329
	Query_179521	239	TEFEAMAGRASSKDWKRISIRYAGRPLQCLIQDGILNPHHAASCTAACCDDMT	-----LSGPVRLFPYKRRK---	306
	Query_179520	208	TEFEGMGRASSKDWKRISIRYAGRPLQCLIQERILNPHAASTCAACCDLIS[8]Qes18	MTGPVRLFPPYKRRK---	289
	Query_179519	330	DLDMESGPKR-----RNTHHSNNNSNTNNNTSGS-GANNCVDVTAAVAATASVVDEN	----NMFLSEENITSKDE	398
	Query_179521	307	ENELPTPVKX---DSPKNITLPPATAATTFTVTPSGQITTSGAL/FDRASRTVEATAVISESPAGQDVFAGATVQEAGVQ	-----383	383
	Query_179520	290	DNERAASPDNKkeiQSPKNNITLAPGA---TFTVSPSGQIISTGTLSDRSASGETASIISDSPAAPDVYTNTTV-----	-----360	360
	Query_179519	399	P WAALNDSDLDTST-TELVDQSQMNTYERETFVVNIND	GSSIAVLDTSQ---SMKNIEHVYCTMVKATNDP	464
	Query_179521	384	P(10)YPGYQDSD[10]JLPFShPKIV/LSLPALAVPPSTPTKAVSP	-----TVVSGLMSEHRHSWLYLEMVNSLLNTAQQL	471
	Query_179520	361	-----LTTLPALAVVPQQAVVQSKPppaGFLNLGLETGEQRTwLYLEEMANTLLSNVQQL	-----415	415
	Query_179519	465	KRMLINDMKQSFERRIEVL QKERDAAVSAMRVRQVHADID[9]NEIIASAKCANCNREALAECSLCRKTPYCSEFCQR	-----547	547
	Query_179521	472	KTLFEPQAQKQASSC-----REAAVTQARMQVDTERK-----EQSCVNCGREASECTGCHKVNYCSTFCQR	-----531	531
	Query_179520	416	KVLIQAQKQASQSGAHAT[7]SGRKEFQSQFSLTEEPEGKI	TEIIIKHTCVNCGREASCTGCHKVHYCSCFCQR	496
	Query_179519	548	KDWNAHQVECTR-NPQTTQ-----QVMLLIDDQS--	576	576
	Query_179521	532	KDWKDHQIVCGG-SASVTVqADDVHVEESVIEKVA	V 566	566
	Query_179520	497	KDWKEHQLNCCOpNTSVSIQ-EDTQME---LDKGKA	-----528	528

Figure 2. *Deaf1* is a developmental regulatory gene that is common to bilateria. A) Artist illustration of embryonic *Deaf1* expression representing in situ hybridization in mouse and zebrafish and immunofluorescence staining in fruit fly, compared across roughly equivalent stages [embryonic day (E) 10.5, hours post fertilization (hpf) 24, and stage (S) 13, respectively]. The widespread expression across numerous developing organs is apparent in all three species despite significant differences in morphology. Drawings based on Richardson et al.,^[34] Vulto-van Silfhout et al.,^[35] and Veraska et al.^[36] B) Amino acid sequence similarity of *Deaf1* protein across the three species. Mouse: zebrafish = 53.39% identity; mouse: fruit fly = 55.65% identity; zebrafish: fruit fly = 31.65% identity.

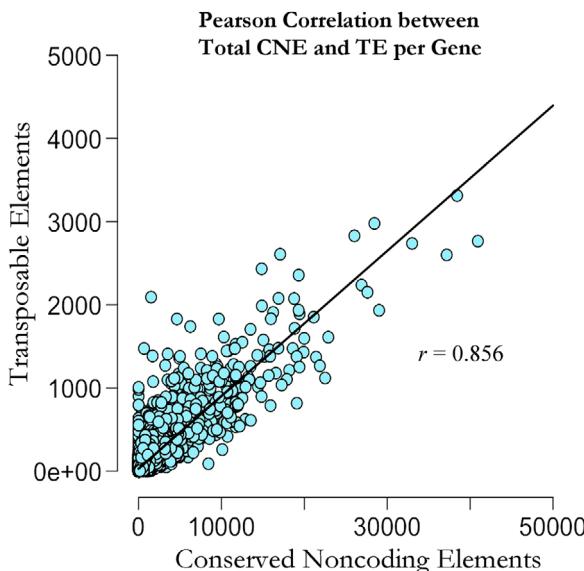


Figure 3. Correlation of total intronic transposable elements (TE) and conserved noncoding elements (CNE) per gene across the entire human genome. The strong correlation suggests a significant relationship between the two, potentially reflecting the TE derivation of many CNE (data published in Casanova et al.^[4]).

entire human genome correlates strongly with one another ($r = 0.856, p < 0.0001$) (Figure 3) (See Table 1 for relationship between CNE and TE content according to gene size.).^[4] This relationship is perhaps unsurprising since some CNE are derived from TE insertion and exaptation events.^[51] For instance, Lowe et al.^[52] reported that thousands of TE fragments in the human genome have been co-opted and have been under strong purifying selection over the last 100 million years. Indeed, at least 20% of regulatory sequences in humans are derived from mobile elements and because many ancient transposition events are no longer recognizable, an even wider breadth of TE-derived CNE may exist.^[51,53] Lowe and Haussler^[51] went on to report that, across 29 mammalian genomes, TE fragments are consistently exapted as gene regulatory sequences, especially transcription factor binding sites. Meanwhile, Villar et al.^[54] reported that, across 20 mammalian species, species-specific enhancers are often derived from ancient transposon-rich DNA, suggesting that in addition to TE insertion, genetic drift may play an important role in the development of new enhancer elements. A similarly slow phenomenon of “optimization” is noted in TE-derived exonization.^[55]

3.2. Transposon-Derived Regulatory Elements Are Key Drivers of Speciation

Within mammals, there are strong links between TE activity and speciation.^[56,57] As mentioned, TEs provide a prime source of potential regulatory material to the host genome and the primate lineage is an excellent example of such a relationship. As Zeh et al.^[10] review, the primate-specific retrotransposons known as *Alu* elements have been exceptionally successful in this lineage.^[58,59] For instance, the *AluJ* family arose prior to the divergence of haplorrhines and strepsirrhines approximately 65

million years ago. More recently, about 40–35 million years ago, the *AluS* family arose, followed by *AluY*. *AluS* was actively propagating during the radiation of tarsiers from platyrhines and before the radiation of new world monkeys (platyrhines) from catarrhines.^[60–62] In fact, the reconstruction of the *Alu* genetic “fossil record”¹ strongly supports that their rate of propagation was highest during this time period.^[64,65] *AluY* eventually became the leading family in catarrhines prior to the radiation of old world monkeys and apes.^[66] Finally, *AluYa5* and *AluYb8*—the two main *AluY* subfamilies currently active in humans—arose a few million years ago during the emergence of our own lineage, the hominins.^[58,63,67–69] Given the timing of events and the fact that *Alu* elements are associated with regulatory functions, such as transcription factor binding sites, adenosine-to-inosine (A-to-I) editing, and circular RNA, it is likely that this family of retrotransposons has played an integral role in the evolution of primate morphology, including our own.^[70–74]

TE insertions have also played an important role in the evolution of pregnancy, particularly in the decidualization of the connective tissue within the uterus. For instance, certain endogenous retroviruses (ERV) are highly expressed in mammalian uterus in a tissue-specific manner and, among other things, help to drive cell fusion within the trophoblast layers via action of the ERV-derived envelope glycoprotein, syncytin.^[75,76] Interestingly, the Mabuya lizard, which evolved approximately 25 million years ago (mya), is viviparous and has an unusually mammalian-like placenta that also expresses an ERV-derived envelope glycoprotein functionally identical to mammalian syncytin.^[77] Similarly, fossilized evidence of the Jurassic marine reptiles, ichthyosaurs, show they were also viviparous, indicating that live birth has evolved multiple times in the reptilian lineage and may commonly be linked with the exaptation of retroviruses (see Figure 4).

Thousands of other transposon-derived *cis*-regulatory elements have been identified that also regulate placental function, many of which have been exapted as hormone response elements (HRE).^[78] For instance, *Alu* elements have been shown to house high-affinity binding sites for the estrogen, thyroid, and retinoic acid receptors.^[79] SVA elements likewise appear to house HRE half-sites and also bind the glucocorticoid receptor.^[80] Therefore, evidence suggests that both TE-derived HRE and ERV-derived envelope glycoproteins have played important roles in the evolution of mammalian pregnancy.

1. Transposon-Mediated Recombination Can Drive Speciation

Aside from insertion and exaptation, TEs are also frequent sites for recombination due to their repetitive natures, resulting in high copy numbers.^[81–85] This provides additional means for escape from evolutionary stasis. Often, recombination results in deletion of genetic material but it can also lead to duplications or more complex forms of recombination such as translocation.^[86–89] For instance, compared to human, the chimpanzee genome shows evidence of over 600 *Alu*-mediated deletion events alone (panTro1, hg17).^[90] Meanwhile, the human genome contains close to 500 unique *Alu*-mediated deletions.^[91]

¹ Terminology used to refer to the genetic record of *Alu* insertions within the genome.^[63]

Table 1. Average protein length, total and relative densities of conserved noncoding elements (CNE) and transposable elements (TE), and transcript numbers per gene within the human genome broken down according to gene size. Note that protein length and transcript numbers increase consistently with gene size, although intron length is primarily responsible for overall gene size. CNE density increases with gene size but is somewhat variable and TE density peaks between 10 000 and 200 000 bp, then modestly decreases with gene size. The vast majority of genes in the human genome (91%) fall under 150 000 bp in size (see Casanova et al.,^[4] Supplementary Material 2, for full published data).

Gene length [bp]	N [genes]	Protein length [aa]	Total CNE [CNE/gene]	CNE density [CNE/gene normalized by gene length]	Total TE [TE/gene]	TE density [TE/gene normalized by gene length]	Number of Transcripts [per gene]
147–10 000	5802	318	39	6.4	4	0.7	3.2
10 001–25 000	4311	465	183	11.0	26	1.5	4.0
25 001–50 000	3395	603	411	11.5	60	1.7	4.4
50 001–100 000	2631	751	796	11.4	117	1.7	4.5
100 001–150 000	1155	964	1349	11.1	192	1.6	4.8
150 001–200 000	562	1078	1935	11.2	263	1.5	4.8
200 001–400 000	773	1146	3259	11.7	399	1.4	5.2
400 001–600 000	228	1212	5661	11.8	646	1.3	5.5
600 001–800 000	70	1297	8062	11.7	929	1.3	5.7
800 001–1 000 000	41	1257	11 250	12.5	1160	1.3	5.8
1 000 001–1 200 000	19	894	11 923	10.5	1421	1.3	6.1
1 200 001–1 400 000	11	999	16 130	12.7	1732	1.4	5.8
1 400 001+	17	1846	24 229	13.0	2325	1.3	6.6

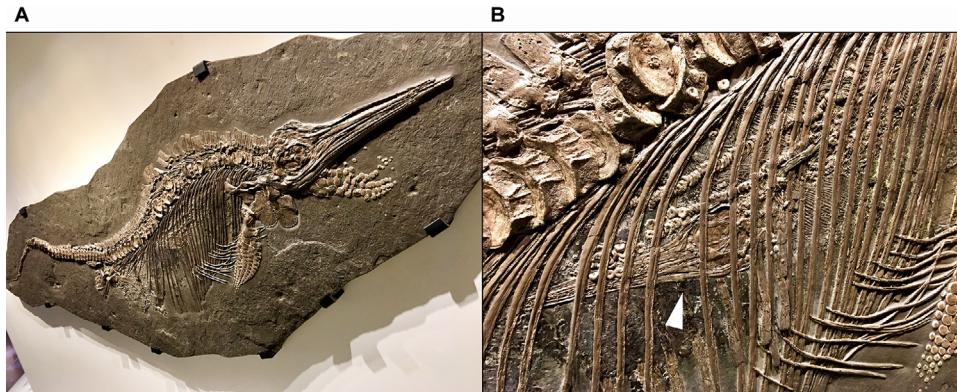


Figure 4. A) The ichthyosaur, *Stenopterygius*, known as “Jurassic Mom,” which died and was fossilized while pregnant with seven fetuses, shows clear evidence of viviparity. B) The skull of one of the fetuses can clearly be seen within the mother’s ribcage (indicated by white arrowhead) along with at least two other vertebral columns from other fetuses located superior to the skull (fossil housed at the Houston Museum of Natural Science within the Morian Hall of Paleontology in Houston, Texas).

In addition, scientists have found that recombination rates are increased around conserved noncoding regions in human as compared to chimpanzee, indicating accelerated evolution in those areas.^[92,93] There are also numerous examples of genes associated with human diseases whose insertions and deletions are the result of *Alu*-mediated recombination.^[82,94–100] Therefore, both the loss and addition of TE-associated genetic materials have been shown to play active roles in primate speciation and pathology.^[66,101]

Primates, particularly humans and the great apes, show evidence of higher rates of segmental duplications.^[102] Interestingly, evidence of segmental duplications has also been identified in other clades that have undergone rapid speciation events. One such example is the infraorder, cetacea, which first began diverging from artiodactyls over 50 mya. Neoceti, which is made up

of the extant parvorders, odontocetes and mystocetes, diverged from archaeocetes at least 36 mya.^[103] From the time of the partly aquatic whale ancestor, *Indohyus*, it took this lineage approximately 10–12 million years to become fully aquatic—an extraordinarily rapid series of speciation events.^[104,105]

Although scientists are unable to study the genomes of extinct lineages, all modern toothed and baleen whales sequenced to date show evidence of significant numbers of segmental duplications enriched in regions housing morphological genes.^[106] The presence of duplications in these regions suggests these variants have influenced cetacean morphology at least from the time of the basilosaurids, if not earlier, and may have played significant roles in this lineage’s return to the sea.

The genomes of many invertebrates, in contrast to those of mammalian vertebrates, show comparatively minimal evidence

of TE insertion into gene-rich areas, with some exceptions.^[107,108] Given the hypothesis presented here and the relative morphological conservation apparent in many invertebrate lines, a slower evolutionary pace may influence DevReg genes within these taxa.^[11] An intriguing and notable exception, however, appears to be the octopus, which has experienced significant TE expansion leading to large-scale genomic rearrangements and an enlargement of specific gene families influencing morphogenesis and central nervous system development in particular.^[109] Interestingly, octopuses are known for some object-oriented play and tool use, and are a model organism for the study of advanced cognitive abilities in invertebrates.^[110]

4. Conclusions and Outlook

While punctuated equilibrium as a theory has stood the test of time, until now there has been no clear genetic explanation for the trends present within the fossil record. Work by Oliver and Greene,^[9] Zeh et al.,^[10] and Davidson^[11] provide admirable starting points and, we think, vital pieces of the larger story.

The developmental gene hypothesis, however, proposes a clear and testable mechanism for stasis via the strong purifying selection acting upon DevReg genes. Likewise, a measurable record of transposable element insertion, exaptation, and recombination within these same genes provides a primary mechanism for bursts of adaptation and evidence of accelerated evolution in these genes across select species lends further support to this notion.

Future avenues of research may include paired approaches using cladistics and genetics, with a particular focus on evolutionary patterns of DevReg genes across extant species and their relationship to morphological divergence. In addition, the study of these genes in species that share cryptic biodiversity may provide additional contrast. With further investigation, we believe the developmental gene hypothesis will provide biologists a better understanding of observable trends within the fossil record and of evolution of this unique and important group of genes responsible for animal form.

Acknowledgements

This work was funded by National Institutes of Health grant (R01 HD-65279).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

conserved noncoding elements, developmental regulatory genes, macroevolution, punctuated equilibrium, regulome, transposable elements

Received: September 23, 2019

Revised: November 30, 2019

Published online:

- [1] S. J. Gould, N. Eldredge, *Syst. Zool.* **1986**, *35*, 143.
- [2] D. Penny, *Syst. Zool.* **1985**, *34*, 201.
- [3] S. J. Gould, N. Eldredge, in *Models in Paleobiology* (Ed: T. J. M. Schopf), Freeman Cooper, San Francisco, CA **1972**, pp. 82–115.
- [4] E. L. Casanova, A. E. Switala, S. Dandamudi, A. R. Hickman, J. Vandenbrink, J. L. Sharp, F. A. Feltus, M. F. Casanova, *Autism Res.* **2019**, *12*, 860.
- [5] J. Pontis, E. Planet, S. Offner, P. Turelli, J. Duc, A. Coudray, T. W. Theunissen, R. Jaenisch, D. Trono, *Cell Stem Cell* **2019**, *24*, 724.
- [6] E. B. Chuong, N. C. Elde, C. Feschotte, *Nat. Rev. Genet.* **2017**, *18*, 71.
- [7] N. C. Comfort, *J. Hist. Biol.* **1999**, *32*, 133.
- [8] R. N. Platt, 2nd, M. W. Vandewege, C. Kern, C. J. Schmidt, F. G. Hoffmann, D. A. Ray, *Mol. Biol. Evol.* **2014**, *31*, 1536.
- [9] K. R. Oliver, W. K. Greene, *Mobile DNA* **2011**, *2*, 8.
- [10] D. W. Zeh, J. A. Zeh, Y. Ishida, *BioEssays* **2009**, *31*, 715.
- [11] E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*, Elsevier, Amsterdam **2006**.
- [12] R. J. Britten, E. H. Davidson, *Q. Rev. Biol.* **1971**, *46*, 111.
- [13] G. Elgar, T. Vavouri, *Trends Genet.* **2008**, *24*, 344.
- [14] V. Sundaram, Y. Cheng, Z. Ma, D. Li, X. Xing, P. Edge, M. P. Snyder, T. Wang, *Genome Res.* **2014**, *24*, 1963.
- [15] S. Stephen, M. Pheasant, I. V. Makunin, J. S. Mattick, *Mol. Biol. Evol.* **2008**, *25*, 402.
- [16] K. Stefflova, D. Thybert, M. D. Wilson, I. Streeter, J. Alekscic, P. Karagianni, A. Brazma, D. J. Adams, I. Talianidis, J. C. Marioni, P. Flicek, D. T. Odom, *Cell* **2013**, *154*, 530.
- [17] S. Dimitrieva, P. Bucher, *Bioinformatics* **2012**, *28*, i395.
- [18] M. Sironi, G. Menozzi, G. P. Comi, R. Cagliani, N. Bresolin, U. Pozzoli, *Hum. Mol. Genet.* **2005**, *14*, 2533.
- [19] F. Galis, B. Sinervo, *Contrib. Zool.* **2002**, *71*, 101.
- [20] T. Vavouri, K. Walter, W. R. Gilks, B. Lehner, G. Elgar, *Genome Biol.* **2007**, *8*, R15.
- [21] C. J. Tabin, S. B. Carroll, G. Panganiban, *Am. Zool.* **1999**, *39*, 650.
- [22] S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History*, W. W. Norton & Co., New York **1989**.
- [23] R. Wood, A. G. Liu, F. Bowyer, P. R. Wilby, F. S Dunn, C. G. Kenchington, J. F. H. Cuthill, E. G. Mitchell, A. Penny, *Nat. Ecol. Evol.* **2019**, *3*, 528.
- [24] N. López-Bigas, C. A. Ouzounis, *Nucleic Acids Res.* **2004**, *32*, 3108.
- [25] D. Ekman, S. Light, A. Björkland, A. Elofsson, *Genome Biol.* **2006**, *7*, R45.
- [26] A. Kapusta, Z. Kronenberg, V. J. Lynch, X. Zhuo, L. Ramsay, L., G. Bourque, M. Yandell, C. Feschotte, *PLoS Genet.* **2013**, *9*, e1003470.
- [27] J. Piriyapongsa, L. Mariño-Ramírez, L. I. K. Jordan, *Genetics* **2007**, *176*, 1323.
- [28] N. Irie, S. Kuratani, *Nat. Commun.* **2011**, *2*, 248.
- [29] W. Arthur, *The Origin of Animal Body Plans: A Study in Evolutionary Developmental Biology*, Cambridge University Press, Cambridge **1997**.
- [30] P. Heyn, M. Kircher, A. Dahl, J. Kelso, P. Tomancak, A. T. Kalinka, K. M. Neugebauer, *Cell Rep.* **2014**, *6*, 285.
- [31] H. Fang, Y. Yang, C. Li, S. Fu, Z. Yang, G. Jin, K. Wang, J. Zhang, Y. Jin, *Dev. Cell* **2010**, *19*, 174.
- [32] E. L. Casanova, Z. Gerstner, J. L. Sharp, M. F. Casanova, F. A. Feltus, *Front. Psychiatry* **2018**, *9*, e535.
- [33] L. A. Pennacchio, N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, E. M. Rubin, *Nature* **2006**, *444*, 499.
- [34] L. Richardson, S. Venkataraman, P. Stevenson, Y. Yang, J. Moss, L. Graham, N. Burton, B. Hill, J. Rao, R. A. Baldock, C. Armit, *Nucleic Acids Res.* **2014**, *42*, D835.
- [35] A. T. Vulto-van Silfhout, S. Rajamanickam, P. J. Jensik, S. Vergult, N. de Rocker, K. J. Newhall, R. Raghavan, S. N. Reardon, K. Jarrett, T.

- McIntyre, J. Bulinski, S. L. Ownby, J. I. Huggenvik, G. S. McKnight, G. M. Rose, X. Cai, A. Willaert, C. Zweier, S. Endele, J. de Ligt, B. W. van Bon, D. Lugtenberg, P. F. de Vries, J. A. Veltman, H. van Bokhoven, H. G. Brunner, A. Rauch, A. P. de Brouwer, G. L. Carvill, A. Hoischen, H. C. Mefford, E. E. Eichler, L. E. L. M. Vissers, B. Menten, M. W. Collard, B. B. A. de Vries, *Am. J. Hum. Genet.* **2014**, *94*, 649.
- [36] A. Veraska, J. Kennison, W. McGinnis, *Genesis* **2002**, *33*, 67.
- [37] E. C. Raff, E. M. Popodi, J. S. Kauffman, B. J. Sly, F. R. Turner, V. B. Morris, R. A. Raff, *Evol. Dev.* **2003**, *5*, 478.
- [38] M. K. Richardson, *Dev. Biol.* **1995**, *172*, 412.
- [39] R. J. Richards, *Biol. Philos.* **2009**, *24*, 147.
- [40] S. J. Gould, N. Eldredge, *Nature* **1993**, *366*, 223.
- [41] J. M. Comerón, M. Kreitman, *Genetics* **2000**, *156*, 1175.
- [42] J. A. Drake, C. Bird, J. Nemesh, D. J. Thomas, C. Newton-Cheh, A. Reymond, L. Excoffier, H. Attar, S. E. Antonarakis, E. T. Dermitzakis, J. N. Hirschhorn, *Nat. Genet.* **2006**, *38*, 223.
- [43] E. Gazave, T. Marqués-Bonet, O. Fernando, B. Charlesworth, A. Navarro, *Genome Biol.* **2007**, *8*, R21.
- [44] M. J. Hubisz, K. S. Pollard, *Curr. Opin. Genet. Dev.* **2014**, *29*, 15.
- [45] D. C. Presgraves, *Nat. Rev. Genet.* **2010**, *11*, 175.
- [46] M. Pfenninger, K. Schwenk, *BMC Evol. Biol.* **2007**, *7*, 121.
- [47] J. R. True, E. S. Haag, *Evol. Dev.* **2001**, *3*, 109.
- [48] M. C. King, A. C. Wilson, *Science* **1975**, *188*, 107.
- [49] L. M. Cherty, S. M. Case, A. C. Wilson, *Science* **1978**, *200*, 209.
- [50] M. Barrier, R. H. Robichaux, M. D. Purugganan, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10208.
- [51] C. Lowe, D. Haussler, *PLoS One* **2012**, *7*, e43128.
- [52] C. B. Lowe, G. Bejerano, D. Haussler, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8005.
- [53] G. Bejerano, C. B. Lowe, N. Ahituv, B. King, A. Siepel, S. R. Salama, E. M. Rubin, W. J. Kent, D. Haussler, *Nature* **2006**, *441*, 87.
- [54] D. Villar, C. S. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. Turner, M. F. Bertelsen, E. P. Murchison, P. Flückeck, D. T. Odom, *Cell* **2015**, *160*, 554.
- [55] J. Schmitz, J. Brosius, *Biochimie* **2011**, *93*, 1928.
- [56] M. Ricci, V. Peona, E. Guichard, C. Taccioli, A. Boattini, *J. Mol. Evol.* **2018**, *86*, 303.
- [57] H. H. Kazazian, *Science* **2004**, *303*, 1626.
- [58] M. A. Batzer, P. L. Deininger, *Nat. Rev. Genet.* **2002**, *3*, 370.
- [59] J. A. Bailey, G. Liu, E. E. Eichler, *Am. J. Hum. Genet.* **2003**, *73*, 823.
- [60] D. A. Ray, M. A. Batzer, *BMC Evol. Biol.* **2005**, *5*, 51.
- [61] G. E. Liu, C. Alkan, L. Jiang, S. Zhao, E. E. Eichler, *Genome Res.* **2009**, *19*, 876.
- [62] The Marmoset Genome Sequencing and Analysis Consortium, *Nat. Genet.* **2014**, *46*, 850.
- [63] D. J. Hedges, P. A. Callinan, R. Cordaux, J. Xing, E. Barnes, M. A. Batzer, *Genome Res.* **2004**, *14*, 1068.
- [64] M. K. Konkel, J. A. Walker, M. A. Batzer, *Evol. Anthropol.* **2010**, *19*, 236.
- [65] P. Deininger, *Genome Biol.* **2011**, *12*, 236.
- [66] K. Han, M. K. Konkel, J. Xing, H. Wang, J. Lee, T. Meyer, C. T. Huang, E. Sandifer, K. Hebert, E. W. Barnes, R. Hubley, W. Miller, A. F. Smit, B. Ullmer, M. A. Batzer, *Science* **2007**, *316*, 238.
- [67] E. A. Bennet, H. Keller, R. E. Mills, S. Schmidt, J. V. Moran, O. Weichenrieder, S. E. Devine, *Genome Res.* **2008**, *18*, 1875.
- [68] M. K. Konkel, J. A. Walker, A. B. Hotard, M. C. Ranck, C. C. Fontenont, J. Storer, C. Stewart, G. T. Marth, 1000 Genomes Consortium, M. A. Batzer, *Genome Biol. Evol.* **2015**, *7*, 2608.
- [69] International Human Genome Sequencing Consortium, *Nature* **2001**, *409*, 860.
- [70] P. Polak, E. Domany, *BMC Genomics* **2006**, *7*, 133.
- [71] A. Athanasiadis, A. Rich, S. Maas, *PLoS Biol.* **2004**, *2*, e391.
- [72] W. R. Jeck, J. A. Sorrentino, K. Wang, M. K. Slevin, C. E. Burd, J. Liu, W. F. Marzluff, N. E. Sharpless, *RNA* **2013**, *19*, 141.
- [73] J. Häslar, K. Strub, *Nucleic Acids Res.* **2006**, *34*, 5491.
- [74] N. Paz-Yaacov, E. Y. Levanon, E. Nevo, Y. Kinar, A. Harmelin, J. Jacob-Hirsch, N. Amariglio, E. Eisenberg, G. Rechavi, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12174.
- [75] J. Sugimoto, D. J. Schust, *Reprod. Sci.* **2009**, *16*, 1023.
- [76] S. Mi, X. Lee, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X. Y. Tang, P. Edouard, S. Howes, J. C. Keith, Jr., J. M. McCoy, *Nature* **2000**, *403*, 785.
- [77] G. Cornelis, M. Funk, C. Vernochet, F. Leal, O. A. Tarazona, G. Meurice, O. Heidmann, A. Dupressoir, A. Miralles, M. P. Ramirez-Pinilla, T. Heidmann, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E10991.
- [78] V. J. Lynch, M. C. Nnamani, A. Kapusta, K. Brayer, S. L. Plaza, E. C. Mazur, D. Emera, S. Z. Sheikh, F. Grützner, S. Bauersachs, A. Graf, S. L. Young, J. D. Lieb, F. J. DeMayo, C. Feschotte, G. P. Wagner, *Cell Rep.* **2015**, *10*, 551.
- [79] V. Babich, N. Aksenen, V. Alexeenko, S. L. Oei, G. Buchlow, N. Tomilin, *Gene* **1999**, *239*, 341.
- [80] D. C. Hancks, H. H. Kazazian, Jr., *Semin. Cancer Biol.* **2010**, *20*, 234.
- [81] J. McFadden, G. Knowles, *J. Theor. Biol.* **1997**, *186*, 441.
- [82] M. K. Konkel, M. A. Batzer, *Semin. Cancer Biol.* **2010**, *20*, 211.
- [83] R. E. Mills, C. T. Luttig, C. E. Larkins, A. Beauchamp, C. Tsui, W. S. Pittard, S. E. Devine, *Genome Res.* **2006**, *16*, 1182.
- [84] J. Jurka, O. Kohany, A. Pavlicek, V. V. Kapitonov, M. Jurka, *Cytogenet. Genome Res.* **2005**, *110*, 117.
- [85] A. R. Cardoso, M. Oliveira, A. Amorim, L. Azevedo, *Hum. Genomics* **2016**, *10*, 30.
- [86] T. Ashizawa, R. D. Wells, in *Genetic Instabilities & Neurological Diseases*, 2nd ed. (Eds: R. D. Wells, T. Ashizawa), Academic Press, Amsterdam **2006**, pp. 3–17.
- [87] S. Cerbin, N. Jiang, *Curr. Opin. Genet. Dev.* **2018**, *49*, 63.
- [88] R. E. Mills, K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. Lam, J. Leng, R. Li, Y. Li, C. Y. Lin, R. Luo, et al., *Nature* **2011**, *470*, 59.
- [89] J. Jurka, O. Kohany, A. Pavlicek, V. V. Kapitonov, M. V. Jurka, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1268.
- [90] K. Han, J. Lee, T. J. Meyer, J. Wang, S. K. Sen, D. Srikanta, P. Liang, M. A. Batzer, *PLoS Genet.* **2007**, *3*, e184.
- [91] S. K. Sen, K. Han, J. Wang, J. Lee, H. Wang, P. A. Callinan, M. Dyer, R. Cordaux, P. Liang, M. A. Batzer, *Am. J. Hum. Genet.* **2006**, *79*, 41.
- [92] J. Freudenberg, Y. H. Fu, L. J. Prácek, *Bioinformatics* **2007**, *23*, 1441.
- [93] J. A. Capra, G. D. Erwin, G. McKingsey, J. L. R. Rubenstein, K. S Pollard, *Philos. Trans. R. Soc., B* **2013**, *368*, 20130025.
- [94] S. De Brakeler, J. De Grève, W. Lissens, E. Teugels, *Hum. Mutat.* **2013**, *34*, 785.
- [95] E. M. Brooks, R. F. Branda, J. A. Nicklas, J. P. O'Neill, *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2001**, *476*, 43.
- [96] J. O'Neil, J. Tchinda, A. Gutierrez, L. Moreau, R. S. Maser, K. K. Wong, W. Li, K. McKenna, X. S. Liu, B. Feng, D. Neuberg, L. Silverman, D. J. DeAngelo, J. L. Kutok, R. Rothstein, R. A. DePinho, L. Chin, C. Lee, A. T. Look, *J. Exp. Med.* **2007**, *204*, 3059.
- [97] V. P. Belancio, D. J. Hedges, P. Deininger, *Genome Res.* **2008**, *18*, 343.
- [98] R. Cordaux, M. A. Batzer, *Nat. Rev. Genet.* **2009**, *10*, 691.
- [99] D. C. Hancks, H. H. Kazazian, Jr., *Mobile DNA* **2016**, *7*, 9.
- [100] H. H. Kazazian, Jr., V. Moran, *N. Engl. J. Med.* **2017**, *377*, 361.
- [101] H. T. Lam, X. J. Mu, A. M. Stütz, A. Tanzer, P. D. Cayting, M. Snyder, P. M. Kim, J. O. Korbel, M. B. Gerstein, *Nat. Biotechnol.* **2010**, *28*, 47.
- [102] T. Marques-Bonet, S. Girirajan, E. E. Eichler, *Trends Genet.* **2009**, *25*, 443.

- [103] J. H. Geisler, *Curr. Biol.* **2017**, *27*, R596.
- [104] J. G. Thewissen, L. N. Cooper, M. T. Clementz, S. Bajpai, B. N. Tiwari, *Nature* **2007**, *450*, 1190.
- [105] J. C. Uyeda, T. F. Handsen, S. J. Arnold, J. Pienaar, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 15908.
- [106] M. Tolls, J. Robbins, A. E. Webb, L. F. K. Kuderna, A. F. Caulin, J. D. Garcia, M. Bérubé, N. Pourmand, T. Marques-Bonet, M. J. O'Connell, P. J. Palsbøll, C. C. Maley, *Mol. Biol. Evol.* **2019**, *36*, 1746.
- [107] N. Sela, E. Kim, G. Ast, *Genome Biol.* **2010**, *11*, R59.
- [108] G. E. Chisholm, T. G. Cooper, *J. Bacteriol.* **1992**, *174*, 2548.
- [109] C. B. Albertin, O. Simakov, T. Mitros, Z. Y. Wang, J. R. Pungor, E. Edsinger-Gonzales, S. Brenner, C. W. Ragsdale, D. S. Rokhsar, *Nature* **2015**, *524*, 220.
- [110] M. J. Kuba, R. A. Byme, D. V. Meisel, J. A. Mather, *J. Comp. Psychol.* **2006**, *120*, 184.