nature genetics

Supplementary information https://doi.org/10.1038/s41588-024-02059-0

Functional innovation through new genes as a general evolutionary process

Supplementary Figures and Tables

Supplementary Figure 1. The distribution of gene numbers and genome sizes across 6,794 extant species.

Supplementary Figure 2. Origination of new genes across life-origination scale evolutionary timescales. The two panels illustrate the origination of new genes on a long timescale (life-origination scale, ~4,250Mya) in humans[1] and Arabidopsis[2].

Supplementary Figure 3. New genes with documented essential phenotypes and measured selective forces. A. The proportion (α) of amino acid substitutions in a group of fixed new gene duplicates under positive selection that has evolved essential phenotypes[3] recalculated from the data[4,5] (L and S are the phenotypes of lethality and sterility detected by the knockdown experiments in the two studies). This suggests that a significant number of these substitutions are likely contributing to the development of essential functions of essential new genes. B. A genomic knockdown experiment in D. melanogaster revealed rapid evolution of functional essentiality in new genes[6]. For example, 317 genes which originated between 69 Mya and 62 Mya were tested if gene knock-down can cause lethal phenotype, eventually, 93 out of 317 genes show lethal/semi-lethal phenotype after KD. The ages of genes from GenTree[7]. "avg." means the ratio of total lethal gene number (210 genes) to total tested genes number (752 genes).

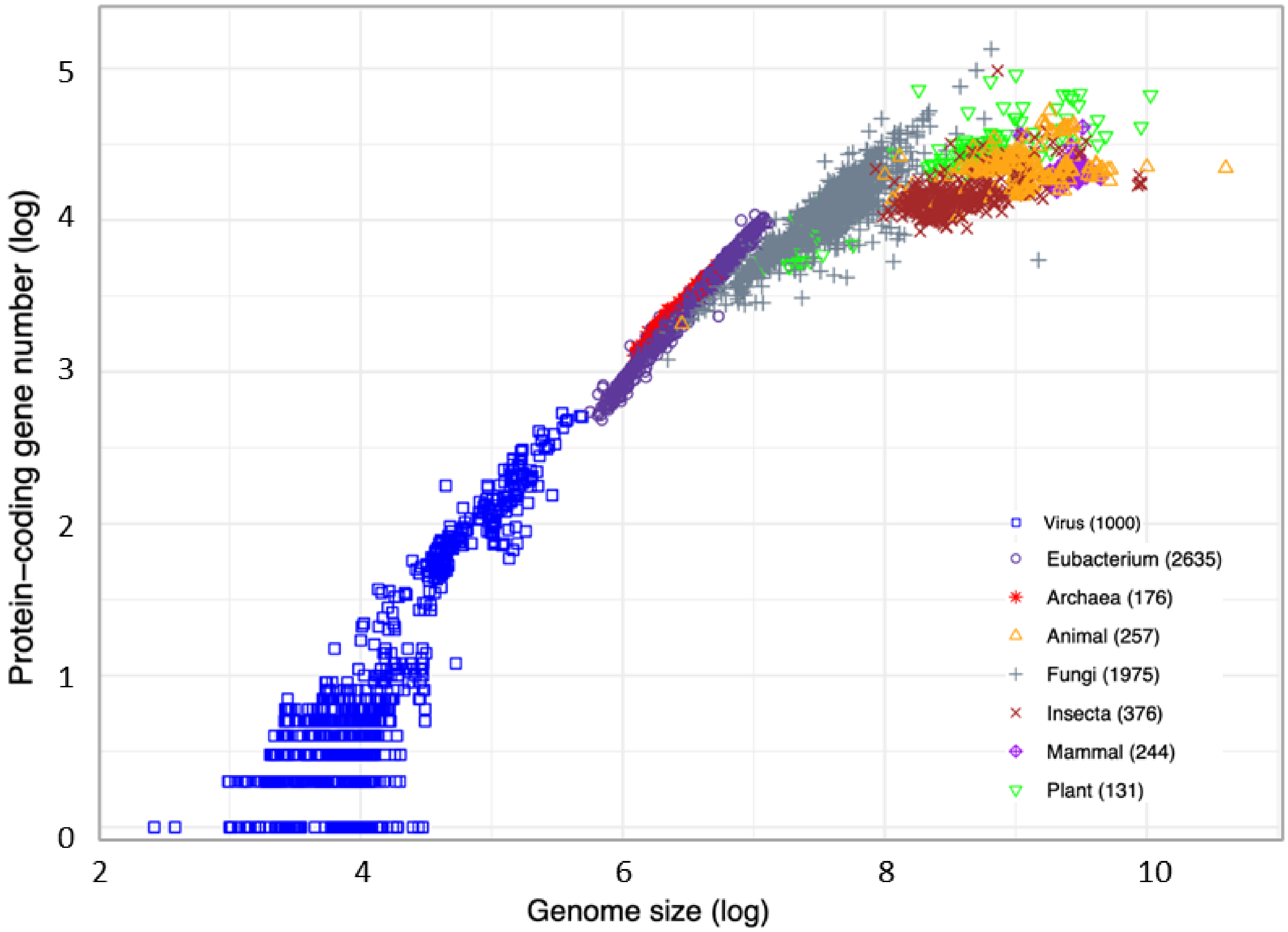
Supplementary Figure 4. The evolutionary fates of new gene via duplication and measured selective forces. The selective force of duplicate polymorphisms in various genomic regions in Drosophila: Complete gene duplicate; Intergenic region duplicate; exonic duplicate; Intronic duplicate; All duplicates; duplicates on X chromosome; Autosomal duplicate[8].

Supplementary Table 1. Fourteen molecular mechanisms of new gene structure formation.

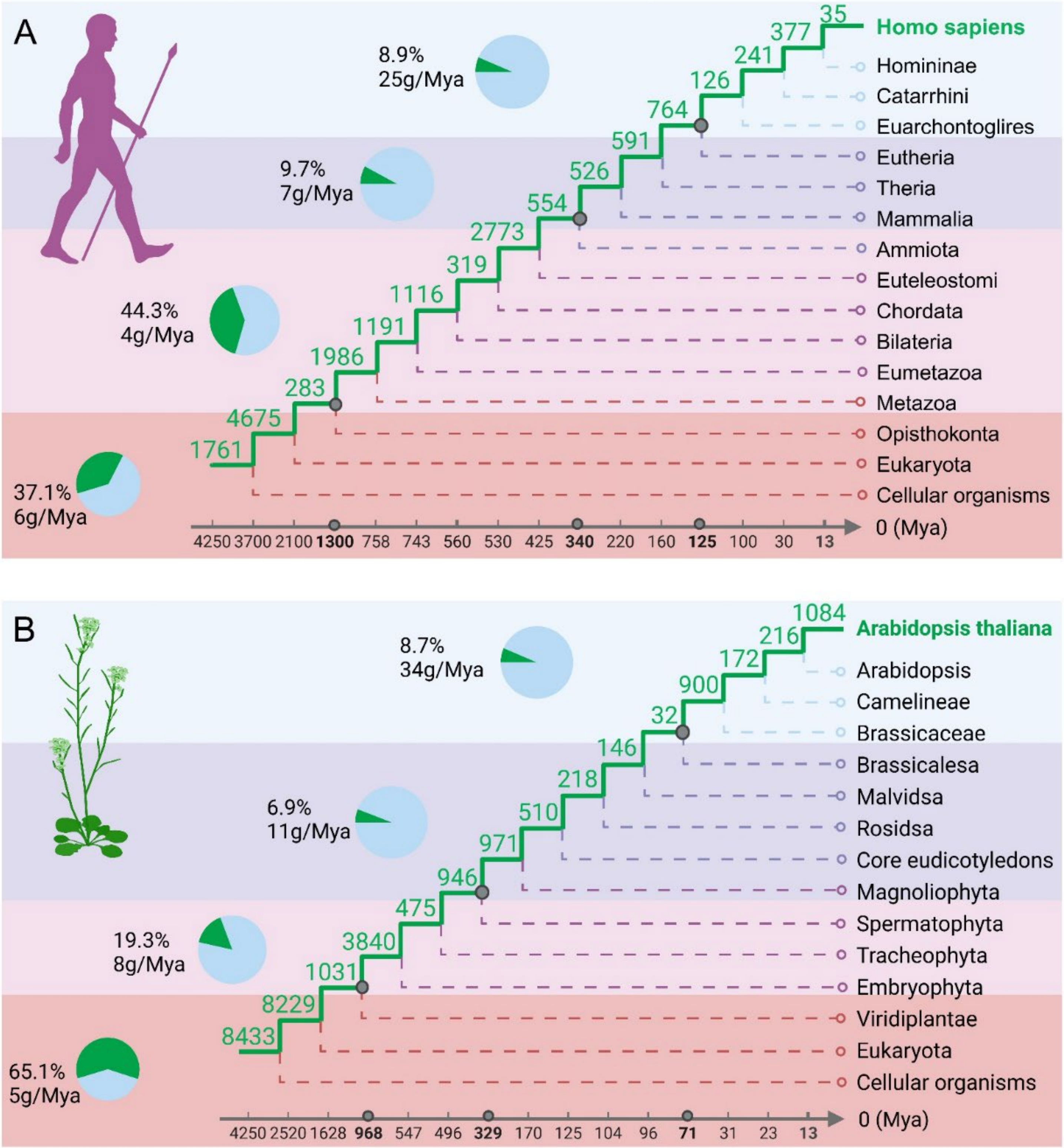
Supplementary Table 2. Lineage-specific origination rates of new genes across species.

Supplementary Table 3 . References for new genes with documented phenotypes and functions.

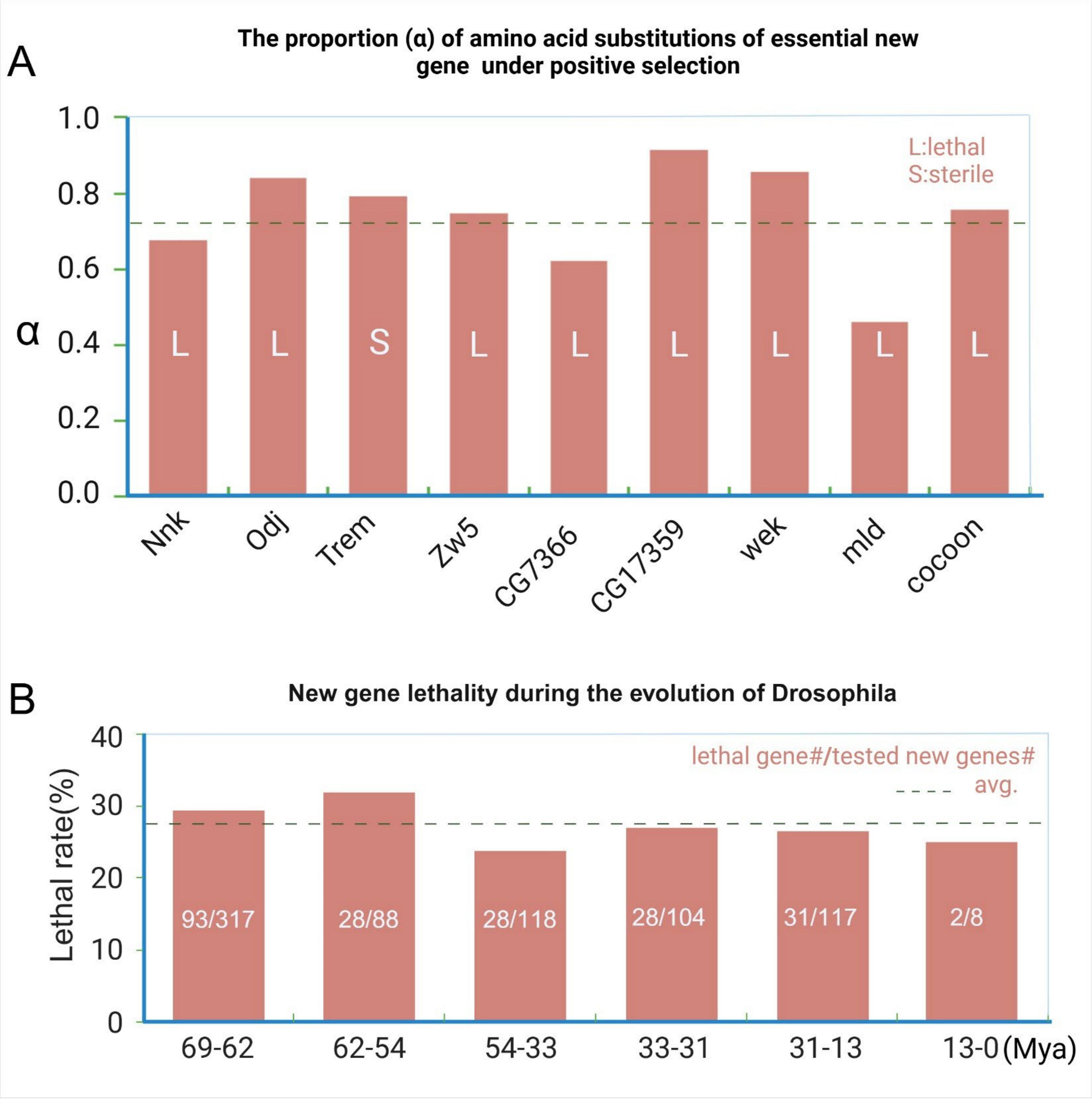
Supplementary Figure 1. The distribution of gene numbers and genome sizes across 6,794 extant species



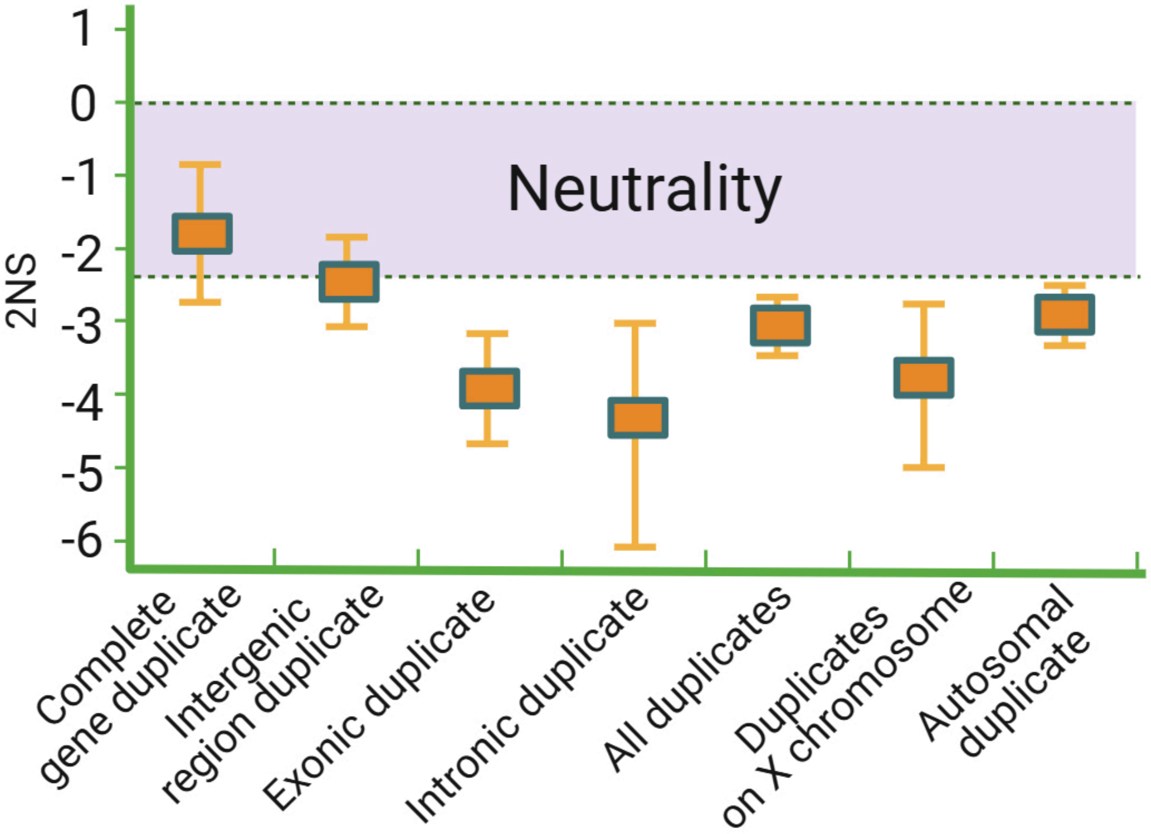
Supplementary Figure 2. Origination of new genes across life-origination scale evolutionary timescales. The two panels illustrate the origination of new genes on a long timescale (life-origination scale, ~4,250Mya) in humans[1] and Arabidopsis[2].



Supplementary Figure 3. New genes with documented essential phenotypes and measured selective forces. A. The proportion (α) of amino acid substitutions in a group of fixed new gene duplicates under positive selection that has evolved essential phenotypes[3] recalculated from the data[4,5] (L and S are the phenotypes of lethality and sterility detected by the knockdown experiments in the two studies). This suggests that a significant number of these substitutions are likely contributing to the development of essential functions of essential new genes. B. A genomic knockdown experiment in D. melanogaster revealed rapid evolution of functional essentiality in new genes[6]. For example, 317 genes which originated between 69 Mya and 62 Mya were tested if gene knock-down can cause lethal phenotype, eventually, 93 out of 317 genes show lethal/semi-lethal phenotype after KD. The ages of genes from GenTree[7]. "avg." means the ratio of total lethal gene number (210 genes) to total tested genes number (752 genes).



Supplementary Figure 4. The evolutionary fates of new gene via duplication and measured selective forces. The selective force of duplicate polymorphisms in various genomic regions in Drosophila: Complete gene duplicate; Intergenic region duplicate; exonic duplicate; Intronic duplicate; All duplicates; duplicates on X chromosome; Autosomal duplicate[8].



Supplementary Table 1. Fourteen molecular mechanisms of new gene structure formation.

	Mechanisms	Comments	Examples
Protein to protein	**1. Exon shuffling** Ectopic recombination of exons and domains from different genes or through retroposition[9].	A common process to create chimeric genes in Eukaryotes[10,11]	*jingwei*[12], *AK111451_3437611*[13], *CG18217*[14], *chirons*[15,] Chimeric TE-gene transcripts[10], *Sphinx*[16], Chimeric genes by truncated duplications[11], *wom*[17].
	2. Retroposition An RNA-based duplication creating new gene by reverse transcription[18].	A widely existing process in metazoans[19-21].	*PMCHL2*[22], *Sphinx*[23], *Pgk2*[24], *Zeus*[25], *HAPSTR2*[26].
	3. Gene duplication-divergence Chromosomal duplication, ectopic recombination, and genome duplication can duplicate genes[27,28].	A main process of new gene evolution, duplicates often evolved new and important functions[29,30].	*Apollo*[31], *Artemis*[31], *CGβ*[32], *Cid*[33], *Odysseus*[34], *COX4-like*[35].
	4.Horizontal gene transfer A gene is transferred between organisms[36-38].	Reported to frequently occur from bacteria to both eukaryotes and prokaryotes and play important role in genome innovation and evolution[39-45].	Vip3A[46], FHY3/FAR1 genes[47], tet(X) variants[48], Pismi_683611/Pismi_689601[49], LOC105383139[50]
	5.Gene fusion Two adjacent genes fuse into a single gene[51].	Involved formation of ~0.5% of prokaryotic genes[52].	fatty-acid synthesis enzymes[53,54], *Kua-UEV*[55], *Sdic*[56], *rpoB* and *rpoC*[57].
	6.Gene fission One single gene split into two genes[58]	Involved formation of ~0.5% of prokaryotic genes[59].	*kms1+*[60], *RPW8-NBS*[61].
	7.Isoform divergence The production of multiple, non-overlapping isoforms can drive distinct gene identities[62].	*NURF-1.D* and *NURF-1.B* are split isoforms from *NURF-1* in C. elegans with distinct roles in male and female gametogenesis, respectively[62]	NURF-1.D and NURF-1.B[62]
	8.Reading-frame shifting In a protein-coding gene, can generate novel proteins[63,64].	Frame shifts have been detected in a large number in human genomes and mosses[65,]	*YAN/AltYAN*[66], *CYP3A5*7 allele*[67], *MT-ATP6 frameshift mutation*[68], *SRY* gene[69], tRNA modification[70], Leucine-Rich Repeat Protein[71] ribosomal frameshift[72]. Cdkn2a and Rbm4[73].
Noncoding to protein	**9.De novo origination** A novel protein-coding gene generated from a noncoding ancestral sequence[74-79].	Detected in flowering plants, mammals, Drosophila, polar fish and yeast. Orphan genes may also be de novo gene candidates[80-87].	*C20orf203*[88],XLOC_175402[89], OsDR10[90], TWISTED1[91].
	10.Short repeat expansion Novel proteins of simple structures were created by repeating short repeats[92-94].	Antifreeze proteins essential for the viability of some polar fish were detected to be recently created by the expansion of short peptides[95].	*AFGP*[79,95-99], *Cb10*[100], *Bs_AFGP8*[101].
	11.TE domestication A transposable element (TE) sequence can be directly recruited by host genes to form a novel gene[102].	Generated 4% of new exons in human protein-coding genes and large proportion of promoters[103-106].	*HLA-DRβ1*[107], *hRED1*[108], *human DAF*[109], lungerkine mRNA[106], *mNSC1*[106], *PGBD1*[110].
	12.Bi-directional promoters It can drive expression both upstream and downstream, allowing for the creation of novel genes[111].	A large number of genes in S. cerevisiae are expressed under bi-directional promoters and inferred to be able to create new genes in upstream regions[112].	Pigt[113], peroxygenase[114], RPBSA promoter[115], designed bi-directional duplex promoters[116], catalase promoter system[117], TH2A and TH2B[118], PACRG[119].
Protein to noncoding	**13.Pseudogene regulators** Gene duplicates through recent pseudogenization regulate parental genes[120].	Mammalian pseudogenes were found to be transcribed and play a role as 'sponges' for microRNAs. It has been identified involving cancer in human[121-125] and rice[126,127].	*SUMO1P3*[128], *HBBP1*[129], *hYRNAs*[130].

Noncoding to noncoding	**14.Noncoding RNA genes** Noncoding RNAs can be generated from noncoding sequences[131].	These noncoding genes that do not encode proteins but can have various important functions[132,133] were detected in a large number in multiple species[134-140]. there are several mechanisms can promote noncoding RNA genes formation[141].	*Sphinx*[23,142,] HAR1[143-145], miR-941[146-148], Xist[149], roX[150], HOTAIR[151].

Supplementary Table 2. Lineage-specific origination rates of new genes across species.

Species/ lineage	Mechanism	Gene #/ Time (mya)/ origination rate
Vertebrate		
Homo sapiens (Homo)	Duplication & putative de novo genes	846/70/12.1[7]
	De novo	60/6.4/9.4[153]
	Retrogenes	76/63/1.2[155]
	Retrocopy with exon fusion	36/63/0.6[157]
Platichthys stellatus (Platichthys)	DNA duplication	1541/73.7/20.9[159]
	RNA duplication	96/73.7/1.3[159]
	De novo	128/73.7/1.7[159]
Mus musculus (Mus)	DNA/RNA duplication and de novo	2316/70/33.1[7]
D. rerio (Danio)	Orphan genes	66/320/0.2[164]
	Duplication	135/320/0.4[164]
Insecta		
Anopheles gambiae (malaria mosquitoes)	RNA-based duplication	28/100/0.3[167]
Bombyx mori (Bombyx)	Orphan genes	738/125/6[168]
Drosophila melangster (Sophophora)	Duplication & putative de novo genes	1848/62/29.8[134]
	Retrogenes	94/63/1.5[169]
	Exon shuffling	14/55/0.3[14]

Species/ Lineage	Mechanism	Gene #/ Time (mya)/ origination rate
Amoebozoa		
Dictyostelium discoideum (Dictyostelium)	DNA/RNA duplication and orphan	3304/520/6.4[152]
Fungi		
Saccharomyces cerevisiae (Saccharomyces)	De novo genes	415/82/5.1[154]
Lachancea fantastica (Lachancea)	De novo genes	288/82/3.5[154]
Escherichia coli	proto genes	9/0.05/1.4[156]
Nematoda		
Caenorhabditis elegans(Caenorhabditis)	Orphan genes and de novo genes	893/19/47[158]
	Orphan genes	5645/193/29[160]
Angiosperms		
Arabidopsis thaliana (Arabidopsis)	DNA/RNA duplication and de novo	2353/96/25[161]
Phyllostachys edulis (Phyllostachys)	Orphan genes	1622/46/35[162]
Citrus sinensis (Citrus)	Orphan genes	1039/8/129[163]
B. rapa (Brassica)	Orphan genes	529/26/20[165]
Triticum aestivum (Triticum)	Orphan genes	3812/19/200[166]
Oryza Sativa (Oryza)	RNA/DNA Duplication and de novo	7845/36/218[161]
	De novo genes	175/3.4/51[81]
	Gene fusion	310/4.9/63[119]
	Pack-Mule-mediated transduplication	76/63/1.2[170]
	Retropseudogene	337/53/6.4[13]

Supplementary Table 3 . References for new genes with documented phenotypes and functions.

Gene name	Function categories	phenotypical and functional studies references
Cocoon	*Development and morphogenesis*	[4]
Desr	*Development and morphogenesis*	[171]
ms(3)K81	*Development and morphogenesis*	[172]
Umbrea	*Development and morphogenesis*	[173,174]
bicoid[175]	*Development and morphogenesis*	[175]
ZAD-ZNF gene	*Development and morphogenesis*	[5]
EXOV	*Development and morphogenesis*	[176]
sORF2146	*Development and morphogenesis*	[177]
panish	*Development and morphogenesis*	[178]
chirons	*Development and morphogenesis*	[15]
XIST	*Development and morphogenesis*	[179-182]
Syncytin 1(ERVW1)	*Development and morphogenesis*	[183-185]
INSL4	*Development and morphogenesis*	[186]
FGF4	*Development and morphogenesis*	[187,188]
Prod1	*Development and morphogenesis*	[189]
JAUP1	*Development and morphogenesis*	[190]
jingwei	*Adaptation*	[191,192]
Adh-Finnega	*Adaptation*	[193]
CYP98A8/CYP98A9	*Adaptation*	[194,195]
CYP84A4	*Adaptation*	[196]
QQS	*Adaptation*	[197-204]
AtEWR	*Adaptation*	[2,205]
BrOG1	*Adaptation*	[206]
GapC	*Adaptation*	[207]
TCAF1/TCAF2	*Adaptation*	[208]
TRIMCyp	*Adaptation*	[209]
bcs-6	*Adaptation*	[210]
AFGP	*Adaptation*	[79]
GIG2	*Adaptation*	[211]
Xcbp1	*Brain and nervous system*	[171]
SRGAP2B/2C/2D	*Brain and nervous system*	[212-215]
ARHGAP11B	*Brain and nervous system*	[216-218]
NOTCH2NLA,B,C	*Brain and nervous system*	[219-221]
ENSG00000205704	*Brain and nervous system*	[222]
Olfactory receptor genes	*Brain and nervous system*	[223]
KRAB–ZNF family	*Brain and nervous system*	[224-226]
Primate opsin genes	*Brain and nervous system*	[227]
Odysseus (OdsH)	*Speciation*	[228-232]
JYAlpha	*Speciation*	[233]
iORF3 and iORF4	*Speciation*	[234,235]
HWS1 and HWS2	*Speciation*	[236]
YELLOW UPPER(YUP)	*Speciation*	[237]
Xmrk-2	*Speciation*	[238-242]
Apl	*Reproduction and Sexuality*	[31]
nsr	*Reproduction and Sexuality*	[243]
Sflc	*Reproduction and Sexuality*	[244]
Sdic family	*Reproduction and Sexuality*	[245,246]
p24-2	*Reproduction and Sexuality*	[173,247]
Zeus	*Reproduction and Sexuality*	[25,169,248,249]
Pros28.1A	*Reproduction and Sexuality*	[173,250]
Poseidon	*Reproduction and Sexuality*	[25]
sphinx	*Reproduction and Sexuality*	[16,251,252]
Tssor-3 and Tssor-4	*Reproduction and Sexuality*	[253]
POLDI/ Pldi	*Reproduction and Sexuality*	[254,255]
Bnams4b	*Reproduction and Sexuality*	[256-258]
INSL3	*Reproduction and Sexuality*	[259]
Goddard, Saturn, Atlas	*Reproduction and Sexuality*	[260-261]

Reference

1. Trigos, A.S., Pearson, R.B., Papenfuss, A.T. & Goode, D.L. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proceedings of the National Academy of Sciences* **114**, 6406-6411 (2017).
2. Arendsee, Z.W., Li, L. & Wurtele, E.S. Coming of age: orphan genes in plants. *Trends Plant Sci* **19**, 698-708 (2014).
3. Smith, N.G. & Eyre-Walker, A. Adaptive protein evolution in Drosophila. *Nature* **415**, 1022-1024 (2002).
4. Lee, Y.C.G. *et al.* Rapid Evolution of Gained Essential Developmental Functions of a Young Gene via Interactions with Other Essential Genes. *Mol Biol Evol* **36**, 2212-2226 (2019).
5. Kasinathan, B. *et al.* Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in *Drosophila*. *Elife* **9**, e63368 (2020).
6. Zeng, X.K. *et al.* Genome-wide RNAi screen identifies networks involved in intestinal stem cell regulation in *Drosophila*. *Cell Reports* **10**, 1226-1238 (2015).
7. Shao, Y. *et al.* GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Research* **29**, 682-696 (2019).
8. Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O. & Long, M. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* **320**, 1629-1631 (2008).
9. Gilbert, W. The exon theory of genes. *Cold Spring Harb Symp Quant Biol* **52**, 901-5 (1987).
10. Berthelier, J. *et al.* Long-read direct RNA sequencing reveals epigenetic regulation of chimeric gene-transposon transcripts in *Arabidopsis thaliana*. *Nature Communications* **14**, 3248 (2023).
11. Zhang, D. *et al.* Dosage sensitivity and exon shuffling shape the landscape of polymorphic duplicates in Drosophila and humans. *Nature Ecology & Evolution* **6**, 273-287 (2022).
12. Long, M. & Langley, C.H. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**, 91-95 (1993).
13. Wang, W. *et al.* High rate of chimeric gene origination by retroposition in plant genomes. *The Plant Cell* **18**, 1791-1802 (2006).
14. Rogers, R.L. & Hartl, D.L. Chimeric Genes as a Source of Rapid Evolution in *Drosophila melanogaster*. *Molecular biology and evolution* **29**, 517-529 (2012).
15. Fang, C., Gan, X., Zhang, C. & He, S. The new chimeric chiron genes evolved essential roles in zebrafish embryonic development by regulating NAD(+) levels. *Sci China Life Sci* **64**, 1929-1948 (2021).
16. Wang, W., Brunet, F.G., Nevo, E. & Long, M. Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*. *PNAS* **99**, 4448-4453 (2002).
17. Zou, Y., Geuverink, E., Beukeboom, L.W., Verhulst, E.C. & Van de Zande, L. A chimeric gene paternally instructs female sex determination in the haplodiploid wasp Nasonia. *Science* **370**, 1115-1118 (2020).
18. Brosius, J. Retroposons—seeds of evolution. *Science* **251**, 753-753 (1991).
19. Betrán, E., Thornton, K. & Long, M. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **12**, 1854-1859 (2002).
20. Emerson, J., Kaessmann, H., Betrán, E. & Long, M. Extensive gene traffic on the mammalian X chromosome. *Science* **303**, 537-540 (2004).
21. Kaessmann, H., Vinckenbosch, N. & Long, M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics* **10**, 19-31 (2009).
22. Courseaux, A. & Nahon, J.-L. Birth of two chimeric genes in the Hominidae lineage. *Science* **291**, 1293-1297 (2001).
23. Wang, W., Brunet, F.G., Nevo, E. & Long, M. Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* **99**, 4448-4453 (2002).
24. Boer, P.H., Adra, C.N., Lau, Y.-F. & McBURNEY, M.W. The testis-specific phosphoglycerate kinase gene pgk-2 is a recruited retroposon. *Molecular and cellular biology* **7**, 3107-3112 (1987).
25. Xia, S.Q. *et al.* Rapid Gene Evolution in an Ancient Post-transcriptional and Translational Regulatory System Compensates for Meiotic X Chromosomal Inactivation. *Molecular Biology and Evolution* **39**(2022).
26. Amici, D.R. *et al.* The HAPSTR2 retrogene buffers stress signaling and resilience in mammals. *Nature communications* **14**, 152 (2023).
27. Muller, H.J. Pilgrim Trust Lecture-The gene. *Proceedings of the Royal Society of London. Series B-Biological Sciences* **134**, 1-37 (1947).
28. Ohno, S. *Evolution by gene duplication*, (Springer Science & Business Media, 2013).

29. Conrad, B. & Antonarakis, S.E. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu. Rev. Genomics Hum. Genet.* **8**, 17-35 (2007).
30. Lynch, M. Gene duplication and evolution. *Science* **297**, 945-947 (2002).
31. VanKuren, N.W. & Long, M.Y. Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. *Nat. Ecol. Evol.* **2**, 705-712 (2018).
32. Talmadge, K., Vamvakopoulos, N.C. & Fiddes, J.C. Evolution of the genes for the β subunits of human chorionic gonadotropin and luteinizing hormone. *Nature* **307**, 37-40 (1984).
33. Malik, H.S. & Henikoff, S. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. *Genetics* **157**, 1293-1298 (2001).
34. Ting, C.-T. *et al.* Gene duplication and speciation in Drosophila: evidence from the *Odysseus* locus. *PNAS* **101**, 12232-12235 (2004).
35. Eslamieh, M., Mirsalehi, A., Markova, D.N. & Betrán, E. COX4-like, a nuclear-encoded mitochondrial gene duplicate, is essential for male fertility in *Drosophila melanogaster*. *Genes* **13**, 424 (2022).
36. Freeman, V.J. Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. *Journal of bacteriology* **61**, 675-688 (1951).
37. Ochman, H., Lawrence, J.G. & Groisman, E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304 (2000).
38. Syvanen, M. Cross-species gene transfer; implications for a new theory of evolution. *Journal of theoretical Biology* **112**, 333-343 (1985).
39. Brown, J.R. Ancient horizontal gene transfer. *Nature Reviews Genetics* **4**, 121-132 (2003).
40. Hotopp, J.C.D. Horizontal gene transfer between bacteria and animals. *Trends in genetics* **27**, 157-163 (2011).
41. Keeling, P.J. & Palmer, J.D. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**, 605-618 (2008).
42. Kurland, C.G., Canback, B. & Berg, O.G. Horizontal gene transfer: a critical view. *PNAS* **100**, 9658-9662 (2003).
43. Soucy, S.M., Huang, J. & Gogarten, J.P. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**, 472-482 (2015).
44. Van Elsas, J.D., Turner, S. & Bailey, M.J. Horizontal gene transfer in the phytosphere. *New phytologist* **157**, 525-537 (2003).
45. Gogarten, J.P. & Townsend, J.P. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**, 679-87 (2005).
46. Estruch, J.J. *et al.* Vip3A, a novel Bacillus thuringiensis vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects. *Proc Natl Acad Sci U S A* **93**, 5389-94 (1996).
47. Lin, R. *et al.* Transposase-derived transcription factors regulate light signaling in Arabidopsis. *Science* **318**, 1302-5 (2007).
48. He, T. *et al.* Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. *Nat Microbiol* **4**, 1450-1456 (2019).
49. Sahu, N. *et al.* Vertical and horizontal gene transfer shaped plant colonization and biomass degradation in the fungal genus Armillaria. *Nat Microbiol* **8**, 1668-1681 (2023).
50. Li, Y. *et al.* HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**, 2975-2987. e10 (2022).
51. Leffers, H., Gropp, F., Lottspeich, F., Zillig, W. & Garrett, R.A. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. *Journal of molecular biology* **206**, 1-17 (1989).
52. Zhou, Y. *et al.* Gene fusion as an important mechanism to generate new genes in the genus *Oryza*. *Genome biology* **23**, 1-23 (2022).
53. McCarthy, A.D. & Hardie, D.G. Fatty acid synthase—an example of protein evolution by gene fusion. *Trends in Biochemical Sciences* **9**, 60-63 (1984).
54. Wakil, S.J. Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* **28**, 4523-4530 (1989).
55. Thomson, T.M. *et al.* Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. *Genome research* **10**, 1743-1756 (2000).
56. Nurminsky, D.I., Nurminskaya, M.V., Aguiar, D.D. & Hartl, D.L. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**, 572-575 (1998).
57. Zakharova, N. *et al.* Fused and overlapping rpoB and rpoC genes in helicobacters, campylobacters, and related bacteria. *Journal of bacteriology* **181**, 3857-3859 (1999).

58. Riley, M. & Labedan, B. Protein evolution viewed through Escherichia coli protein sequences: introducing the notion of a structural segment of homology, the module. *Journal of molecular biology* **268**, 857-868 (1997).
59. Wang, W., Yu, H. & Long, M. Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. *Nature genetics* **36**, 523-527 (2004).
60. Shimanuki, M. *et al.* A novel fission yeast gene, kms1+, is required for the formation of meiotic prophase-specific nuclear architecture. *Molecular and General Genetics MGG* **254**, 238-249 (1997).
61. Zhong, Y. & Cheng, Z.-M. A unique RPW8-encoding class of genes that originated in early land plants and evolved through domain fission, fusion, and duplication. *Scientific reports* **6**, 32923 (2016).
62. Xu, W. *et al.* Evolution of Yin and Yang isoforms of a chromatin remodeling subunit precedes the creation of two genes. *Elife* **8**, e48119 (2019).
63. Advani, A. Reading-Frame Shift Mechanisms of Introducing Genetic Novelty to the Human Genome. (2020).
64. Roth, J.R. Frameshift mutations. *Annual review of genetics* **8**, 319-346 (1974).
65. Ardern, Z. Alternative reading frames are an underappreciated source of protein sequence novelty. *Journal of Molecular Evolution*, 1-11 (2023).
66. Guan, Y. *et al.* Gene refashioning through innovative shifting of reading frames in mosses. *Nature Communications* **9**, 1555 (2018).
67. Hu, J. & Ng, P.C. Predicting the effects of frameshifting indels. *Genome biology* **13**, 1-11 (2012).
68. Jackson, C.B. *et al.* A novel mitochondrial ATP6 frameshift mutation causing isolated complex V deficiency, ataxia and encephalomyopathy. *European journal of medical genetics* **60**, 345-351 (2017).
69. Wang, X.-B. *et al.* A de novo frameshift mutation of the SRY gene leading to a patient with 46, XY complete gonadal dysgenesis. *Asian Journal of Andrology* **21**, 522 (2019).
70. Brégeon, D., Colot, V., Radman, M. & Taddei, F. Translational misreading: a tRNA modification counteracts a+ 2 ribosomal frameshift. *Genes & development* **15**, 2295-2306 (2001).
71. Kondo, M. *et al.* Frameshift mutation confers function as virulence factor to leucine-rich repeat protein from *Acidovorax avenae*. *Frontiers in Plant Science* **7**, 1988 (2017).
72. Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244-247 (2006).
73. Okamura, K., Feuk, L., Marquès-Bonet, T., Navarro, A. & Scherer, S.W. Frequent appearance of novel protein-coding sequences by frameshift translation. *Genomics* **88**, 690-697 (2006).
74. Long, M., Betrán, E., Thornton, K. & Wang, W. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **4**, 865-875 (2003).
75. Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H.A. & Begun, D.J. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *PNAS* **103**, 9935-9939 (2006).
76. McLysaght, A. & Hurst, L.D. Open questions in the study of de novo genes: what, how and why. *Nature Reviews Genetics* **17**, 567-578 (2016).
77. Alba, M.M. & Castresana, J. Inverse relationship between evolutionary rate and age of mammalian genes. *Molecular Biology and Evolution* **22**, 598-606 (2005).
78. Carvunis, A.-R. *et al.* Proto-genes and de novo gene birth. *Nature* **487**, 370-374 (2012).
79. Zhuang, X., Yang, C., Murphy, K.R. & Cheng, C.-H.C. Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids. *PNAS* **116**, 4400-4405 (2019).
80. Dujon, B. The yeast genome project: what did we learn? *Trends in Genetics* **12**, 263-270 (1996).
81. Zhang, L. *et al.* Rapid evolution of protein diversity by de novo origination in Oryza. *Nature ecology & evolution* **3**, 679-690 (2019).
82. Murphy, D.N. & McLysaght, A. De novo origin of protein-coding genes in murine rodents. *PloS one* **7**, e48650 (2012).
83. Zhao, L., Saelao, P., Jones, C.D. & Begun, D.J. Origin and spread of de novo genes in *Drosophila melanogaster* populations. *Science* **343**, 769-772 (2014).
84. Vakirlis, N., Carvunis, A.R. & McLysaght, A. Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. *Elife* **9**(2020).
85. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**, 692-702 (2011).
86. Van Oss, S.B. & Carvunis, A.-R. De novo gene birth. *PLoS genetics* **15**, e1008160 (2019).
87. Grandchamp, A. *et al.* Population genomics reveals mechanisms and dynamics of de novo expressed open reading frame emergence in <i>Drosophila melanogaster</i>. *GENOME RESEARCH* **33**, 872-890 (2023).

88. Broeils, L.A., Ruiz-Orera, J., Snel, B., Hubner, N. & van Heesch, S. Evolution and implications of de novo genes in humans. *Nature ecology & evolution*, 1-12 (2023).
89. Ruiz-Orera, J. *et al.* Origins of De Novo Genes in Human and Chimpanzee. *PLoS Genet* **11**, e1005721 (2015).
90. Xiao, W.F. *et al.* A rice gene of De Novo origin negatively regulates pathogen-induced defense response. (2009).
91. Nayelli, M.-M. *et al.* Twisting development, the birth of a potential new gene. *iScience* **25**(2022).
92. Devries, A.L. Glycoproteins as biological antifreeze agents in Antarctic fishes. *Science* **172**, 1152-1155 (1971).
93. DeVries, A.L. Biological antifreeze agents in coldwater fishes. *Comparative Biochemistry and Physiology Part A: Physiology* **73**, 627-640 (1982).
94. O'Grady, S.M., Schrag, J.D., Raymond, J.A. & Devries, A.L. Comparison of antifreeze glycopeptides from Arctic and Antarctic fishes. *Journal of Experimental Zoology* **224**, 177-185 (1982).
95. Cheng, C.-H.C. & Zhuang, X. Molecular origins and mechanisms of fish antifreeze evolution. *Antifreeze proteins volume 1: environment, systematics and evolution*, 275-313 (2020).
96. Cheng, C.-H.C. & Chen, L. Evolution of an antifreeze glycoprotein. *Nature* **401**, 443-444 (1999).
97. Chen, L.B., DeVries, A.L. & Cheng, C.H.C. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *PNAS* **94**, 3817-3822 (1997).
98. Chen, L., DeVries, A.L. & Cheng, C.-H.C. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *PNAS* **94**, 3811-3816 (1997).
99. Zhuang, X., Yang, C., Fevolden, S.-E. & Cheng, C.C. Protein genes in repetitive sequence—antifreeze glycoproteins in Atlantic cod genome. *BMC genomics* **13**, 1-9 (2012).
100. Asano, T. *et al.* Identification of a cuticle protein with unique repeated motifs in the silkworm, Bombyx mori. *Insect biochemistry and molecular biology* **43**, 344-351 (2013).
101. Zhuang, X. & Cheng, C.-H.C. Propagation of a De Novo Gene under Natural Selection: Antifreeze Glycoprotein Genes and Their Evolutionary History in Codfishes. *Genes* **12**, 1777 (2021).
102. Makałowski, W., Mitchell, G.A. & Labuda, D. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends in Genetics* **10**, 188-193 (1994).
103. Chu, C. *et al.* Comprehensive identification of transposable element insertions using multiple sequencing technologies. *Nature communications* **12**, 3836 (2021).
104. Etchegaray, E., Naville, M., Volff, J.-N. & Haftek-Terreau, Z. Transposable element-derived sequences in vertebrate development. *Mobile DNA* **12**, 1-24 (2021).
105. González, J., Martínez, J. & Makalowski, W. Lack of population differentiation patterns of previously identified putatively adaptive transposable element insertions at microgeographic scales. *Biology Direct* **10**, 1-6 (2015).
106. Nekrutenko, A. & Li, W.-H. Transposable elements are found in a large number of human protein-coding genes. *Trends in genetics* **17**, 619-621 (2001).
107. Cairns, J. *et al.* Sequence polymorphism of HLA DRβ1 alleles relating to T-cell-recognized determinants. *Nature* **317**, 166-168 (1985).
108. Gerber, A., O'Connell, M.A. & Keller, W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *Rna* **3**, 453 (1997).
109. Caras, I.W. *et al.* Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature* **325**, 545-9 (1987).
110. Raskó, T. *et al.* A novel gene controls a new structure: PiggyBac Transposable Element-derived 1, unique to mammals, controls mammal-specific neuronal paraspeckles. *Molecular biology and evolution* **39**, msac175 (2022).
111. Beck, C. & Warren, R. Divergent promoters, a common form of gene organization. *Microbiological reviews* **52**, 318-326 (1988).
112. Wei, W., Pelechano, V., Järvelin, A.I. & Steinmetz, L.M. Functional consequences of bidirectional promoters. *Trends in Genetics* **27**, 267-276 (2011).
113. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360-363 (2013).
114. Besleaga, M., Ebner, K., Bürgler, M., Spadiut, O. & Kopp, J. Bi-directional promoter systems facilitate expression of unspecific peroxygenase in *K. phaffii* (known as *P. pastoris*) BSYBG11. (2022).
115. He, K., Rad, S.A.H., Poudel, A. & McLellan, A.D. Compact bidirectional promoters for dual-gene expression in a Sleeping Beauty transposon. *International journal of molecular sciences* **21**, 9256 (2020).

116. Li, Z.T., Jayasankar, S. & Gray, D. Bi-directional duplex promoters with duplicated enhancers significantly increase transgene expression in grape and tobacco. *Transgenic research* **13**, 143-154 (2004).
117. Rajamanickam, V., Metzger, K., Schmid, C. & Spadiut, O. A novel bi-directional promoter system allows tunable recombinant protein production in *Pichia pastoris*. *Microbial Cell Factories* **16**, 1-7 (2017).
118. Singh, I. & Parte, P. Heterogeneity in the epigenetic landscape of murine testis-specific histone variants TH2A and TH2B sharing the same bi-directional promoter. *Frontiers in Cell and Developmental Biology* **9**, 755751 (2021).
119. West, A.B., Lockhart, P.J., O'Farell, C. & Farrer, M.J. Identification of a novel gene linked to parkin via a bi-directional promoter. *Journal of molecular biology* **326**, 11-19 (2003).
120. Jacq, C., Miller, J. & Brownlee, G. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* **12**, 109-120 (1977).
121. Ahmad, H.I. *et al.* Behind the scene: surprises and snags of pseudogenes. *Critical Reviews™ in Eukaryotic Gene Expression* **31**(2021).
122. Bernhardt, M., Galach, M., Novak, D. & Utikal, J. Mediators of induced pluripotency and their role in cancer cells–current scientific knowledge and future perspectives. *Biotechnology journal* **7**, 810-821 (2012).
123. Gonçalves, I., Duret, L. & Mouchiroud, D. Nature and structure of human genes that generate retropseudogenes. *Genome research* **10**, 672-678 (2000).
124. Hu, X., Yang, L. & Mo, Y.-Y. Role of pseudogenes in tumorigenesis. *Cancers* **10**, 256 (2018).
125. Zhang, Z., Harrison, P. & Gerstein, M. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome research* **12**, 1466-1482 (2002).
126. Guo, X., Zhang, Z., Gerstein, M.B. & Zheng, D. Small RNAs originated from pseudogenes: cis-or trans-acting? *PLoS computational biology* **5**, e1000449 (2009).
127. Xiao, J. *et al.* Pseudogenes and their genome-wide prediction in plants. *International Journal of Molecular Sciences* **17**, 1991 (2016).
128. Mei, D. *et al.* Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association. *Medical oncology* **30**, 1-5 (2013).
129. Ma, Y. *et al.* Genome-wide analysis of pseudogenes reveals HBBP1's human-specific essentiality in erythropoiesis and implication in β-thalassemia. *Developmental Cell* **56**, 478-493. e11 (2021).
130. Perreault, J. *et al.* Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. *Nucleic acids research* **33**, 2032-2041 (2005).
131. Holley, R.W. *et al.* Structure of a ribonucleic acid. *Science* **147**, 1462-1465 (1965).
132. Eddy, S.R. Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics* **2**, 919-929 (2001).
133. Ruiz-Orera, J., Villanueva-Cañas, J.L. & Albà, M.M. Evolution of new proteins from translated sORFs in long non-coding RNAs. *Experimental cell research* **391**, 111940 (2020).
134. Dong, C. *et al.* New gene evolution with subcellular expression patterns detected in PacBio-sequenced genomes of *Drosophila* genus. *bioRxiv*, 2022.11. 30.518489 (2022).
135. Eddy, S.R. Noncoding RNA genes. *Current opinion in genetics & development* **9**, 695-699 (1999).
136. Gingeras, T.R. Origin of phenotypes: genes and transcripts. *Genome research* **17**, 682-690 (2007).
137. Jia, H. *et al.* Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *Rna* **16**, 1478-1487 (2010).
138. Mattick, J.S. & Makunin, I.V. Non-coding RNA. *Human molecular genetics* **15**, R17-R29 (2006).
139. Waititu, J.K., Zhang, C., Liu, J. & Wang, H. Plant non-coding RNAs: origin, biogenesis, mode of action and their roles in abiotic stress. *International Journal of Molecular Sciences* **21**, 8401 (2020).
140. Gao, G. *et al.* A long-term demasculinization of X-linked intergenic noncoding RNAs in *Drosophila melanogaster*. *Genome Research* **24**, 629-638 (2014).
141. Palazzo, A.F. & Koonin, E.V. Functional long non-coding RNAs evolve from junk transcripts. *Cell* **183**, 1151-1161 (2020).
142. Dai, H. *et al.* The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. *PNAS* **105**, 7478-7483 (2008).
143. Pollard, K.S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167-72 (2006).
144. Waters, E. *et al.* HAR1: an insight into lncRNA genetic evolution. *Epigenomics* **13**, 1831-1843 (2021).
145. Zhang, L. *et al.* Effects of HAR1 on cognitive function in mice and the regulatory network of HAR1 determined by RNA sequencing and applied bioinformatics analysis. *Frontiers in Genetics* **14**, 947144 (2023).

146. Hu, H. *et al.* Recently Evolved Tumor Suppressor Transcript TP73-AS1 Functions as Sponge of Human-Specific miR-941. *Mol Biol Evol* **35**, 1063-1077 (2018).
147. Bai, R. *et al.* miR-941 as a promising biomarker for acute coronary syndrome. *BMC Cardiovascular Disorders* **17**, 1-11 (2017).
148. Zhao, Q. *et al.* Serum exosomal miR-941 as a promising oncogenic biomarker for laryngeal squamous cell carcinoma. *Journal of Cancer* **11**, 5329 (2020).
149. McHugh, C.A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).
150. Quinn, J.J. *et al.* Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes & development* **30**, 191-207 (2016).
151. De Kumar, B. & Krumlauf, R. HOX s and lincRNAs: Two sides of the same coin. *Science advances* **2**, e1501402 (2016).
152. Luna, S.K. & Chain, F.J.J. Lineage-Specific Genes and Family Expansions in Dictyostelid Genomes Display Expression Bias and Evolutionary Diversification during Development. *Genes (Basel)* **12**(2021).
153. Wu, D.D., Irwin, D.M. & Zhang, Y.P. De Novo Origin of Human Protein-Coding Genes. *PLOS Genetics* **7**(2011).
154. Vakirlis, N. *et al.* A Molecular Portrait of De Novo Genes in Yeasts. *Molecular Biology and Evolution* **35**, 631-645 (2018).
155. Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A. & Kaessmann, H. Emergence of young human genes after a burst of retroposition in primates. *PLOS Biology* **3**, 1970-1979 (2005).
156. Uz-Zaman, M.H., D'Alton, S., Barrick, J.E. & Ochman, H. Promoter recruitment drives the emergence of proto-genes in a long-term evolution experiment with Escherichia coli. *PLoS Biology* **22**, e3002418 (2024).
157. Vinckenbosch, N., Dupanloup, I. & Kaessmann, H. Evolutionary fate of retroposed gene copies in the human genome. *PNAS* **103**, 3220-3225 (2006).
158. Zhang, W.Y., Gao, Y.X., Long, M.Y. & Shen, B.R. Origination and evolution of orphan genes and *de novo* genes in the genome of *Caenorhabditis elegans*. *Sci. China Life Sci.* **62**, 579-593 (2019).
159. Li, H. *et al.* Pattern of New Gene Origination in a Special Fish Lineage, the Flatfishes. *Genes (Basel)* **12**(2021).
160. Prabh, N. & Rodelsperger, C. De Novo, Divergence, and Mixed Origin Contribute to the Emergence of Orphan Genes in *Pristionchus Nematodes*. *G3 (Bethesda)* **9**, 2277-2286 (2019).
161. Cui, X. *et al.* Young Genes out of the Male: An Insight from Evolutionary Age Analysis of the Pollen Transcriptome. *Molecular plant* **8**, 935-945 (2015).
162. Jin, G.H. *et al.* New Genes Interacted With Recent Whole-Genome Duplicates in the Fast Stem Growth of Bamboos. *Molecular Biology and Evolution* **38**, 5752-5768 (2021).
163. Xu, Y.T. *et al.* Identification, characterization and expression analysis of lineage-specific genes within sweet orange (*Citrus sinensis*). *BMC Genomics* **16**(2015).
164. Yang, L.D., Zou, M., Fu, B.D. & He, S.P. Genome-wide identification, characterization, and expression analysis of lineage-specific genes within zebrafish. *BMC Genomics* **14**(2013).
165. Jiang, M.L. *et al.* Mining of Brassica-Specific Genes (BSGs) and Their Induction in Different Developmental Stages and under *Plasmodiophora brassicae* Stress in *Brassica rapa*. *International journal of molecular sciences* **19**(2018).
166. Ma, S., Yuan, Y., Tao, Y., Jia, H. & Ma, Z. Identification, characterization and expression analysis of lineage-specific genes within Triticeae. *Genomics* **112**, 1343-1350 (2020).
167. Miller, D. *et al.* Retrogene Duplication and Expression Patterns Shaped by the Evolution of Sex Chromosomes in *Malaria Mosquitoes*. *Genes (Basel)* **13**(2022).
168. Sun, W., Zhao, X.W. & Zhang, Z. Identification and evolution of the orphan genes in the domestic silkworm, Bombyx mori. *FEBS Lett* **589**, 2731-8 (2015).
169. Bai, Y., Casola, C., Feschotte, C. & Betran, E. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in Drosophila. *Genome Biol* **8**, R11 (2007).
170. Campbell, M.A. *et al.* Identification and characterization of lineage-specific genes within the Poaceae. *Plant physiology* **145**, 1311-1322 (2007).
171. Chen, S.D. *et al.* Frequent Recent Origination of Brain Genes Shaped the Evolution of Foraging Behavior in *Drosophila*. *Cell Reports* **1**, 118-132 (2012).
172. Loppin, B., Lepetit, D., Dorus, S., Couble, P. & Karr, T.L. Origin and neofunctionalization of a *Drosophila* paternal effect gene essential for zygote viability. *Current Biology* **15**, 87-93 (2005).
173. Chen, S.D., Zhang, Y.E. & Long, M.Y. New Genes in *Drosophila* Quickly Become Essential. *Science* **330**, 1682-1685 (2010).

174. Ross, B.D. *et al.* Stepwise Evolution of Essential Centromere Function in a *Drosophila* Neogene. *Science* **340**, 1211-1214 (2013).
175. Lemke, S. *et al. bicoid* occurrence and Bicoid-dependent *hunchback* regulation in lower cyclorrhaphan flies. *Evolution & Development* **10**, 413-420 (2008).
176. Huang, Y. *et al.* Species-specific partial gene duplication in *Arabidopsis thaliana* evolved novel phenotypic effects on morphological traits under strong positive selection. *Plant Cell* **34**, 802-817 (2022).
177. Takeda, T. *et al.* A de novo gene originating from the mitochondria controls floral transition in Arabidopsis thaliana. *Plant Molecular Biology* **111**, 189-203 (2023).
178. Klomp, J. *et al.* A cysteine-clamp gene drives embryo polarity in the midge Chironomus. *Science* **348**, 1040-1042 (2015).
179. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653-1655 (2006).
180. Dror, I. *et al.* XIST directly regulates X-linked and autosomal genes in naive human pluripotent cells. *Cell* **187**, 110-129 e31 (2024).
181. Engreitz, J.M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
182. Simon, M.D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465-469 (2013).
183. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785-789 (2000).
184. Esnault, C., Cornelis, G., Heidmann, O. & Heidmann, T. Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the EnvV syncytin, captured for a function in placentation. *PLoS Genet* **9**, e1003400 (2013).
185. Dang, Q. *et al.* Nuclear Binding Protein 2/Nesfatin-1 Affects Trophoblast Cell Fusion during Placental Development via the EGFR-PLCG1-CAMK4 Pathway. *Int J Mol Sci* **25**(2024).
186. Arroyo, J.I., Hoffmann, F.G., Good, S. & Opazo, J.C. INSL4 pseudogenes help define the relaxin family repertoire in the common ancestor of placental mammals. *J Mol Evol* **75**, 73-8 (2012).
187. Parker, H.G. *et al.* An Expressed *Fgf4* Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. *Science* **325**, 995-998 (2009).
188. Brown, E.A. *et al.* FGF4 retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. *PNAS* **114**, 11476-11481 (2017).
189. Kumar, A., Gates, P.B., Czarkwiani, A. & Brockes, J.P. An orphan gene is necessary for preaxial digit formation during salamander limb development. *Nature Communications* **6**, 8684 (2015).
190. Muzaffar, A. *et al.* A newly evolved rice-specific gene JAUP1 regulates jasmonate biosynthesis and signalling to promote root development and multi-stress tolerance. *Plant Biotechnology Journal* (2024).
191. Long, M.Y. & Langley, C.H. Natural Selection and the Origin of *jingwei*, a Chimeric Processed Functional Gene in *Drosophila*. *Science* **260**, 91-95 (1993).
192. Zhang, J.M., Dean, A.M., Brunet, F. & Long, M.Y. Evolving protein functional diversity in new genes of *Drosophila*. *PNAS* **101**, 16246-16250 (2004).
193. Jones, C.D. & Begun, D.J. Parallel evolution of chimeric fusion genes. *Proc Natl Acad Sci U S A* **102**, 11373-8 (2005).
194. Matsuno, M. *et al.* Evolution of a Novel Phenolic Pathway for Pollen Development. *Science* **325**, 1688-1692 (2009).
195. Liu, Z. *et al.* Evolutionary interplay between sister cytochrome P450 genes shapes plasticity in plant metabolism. *Nature Communications* **7**, 13026 (2016).
196. Weng, J.K., Li, Y., Mo, H.P. & Chapple, C. Assembly of an Evolutionarily New Pathway for α-Pyrone Biosynthesis in *Arabidopsis*. *Science* **337**, 960-964 (2012).
197. Li, L. *et al.* Identification of the novel protein QQS as a component of the starch metabolic network in Arabidopsis leaves. *The Plant Journal* **58**, 485-498 (2009).
198. Silveira, A.B. *et al.* Extensive Natural Epigenetic Variation at a *De Novo* Originated Gene. *PLoS genetics* **9**(2013).
199. Li, L. & Wurtele, E.S. The *QQS* orphan gene of Arabidopsis modulates carbon and nitrogen allocation in soybean. *Plant biotechnology journal* **13**, 177-187 (2015).
200. Li, L. *et al.* QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. *PNAS* **112**, 14734-14739 (2015).
201. Jones, D.C. *et al.* A Clade-Specific *Arabidopsis* Gene Connects Primary Metabolism and Senescence. *Frontiers of plant science* **7**(2016).

202. Qi, M.S. *et al.* *QQS* orphan gene and its interactor *NF-YC4* reduce susceptibility to pathogens and pests. *Plant biotechnology journal* **17**, 252-263 (2019).
203. Tanvir, R. *et al.* *AtQQS* orphan gene and *NtNF-YC4* boost protein accumulation and pest resistance in tobacco (*Nicotiana tabacum*). *Plant science journal* **317**(2022).
204. Lee, S.W. *et al.* PHYTOCHROME-INTERACTING FACTORS are involved in starch degradation adjustment via inhibition of the carbon metabolic regulator QUA-QUINE STARCH in *Arabidopsis*. *The Plant Journal* **114**, 110-123 (2023).
205. Yadeta, K.A., Valkenburg, D.J., Hanemian, M., Marco, Y. & Thomma, B. The Brassicaceae-Specific *EWR1* Gene Provides Resistance to Vascular Wilt Pathogens. *PLOS one* **9**(2014).
206. Jiang, M.L. *et al.* *Brassica rapa* orphan genes largely affect soluble sugar metabolism. *Horticulture research* **7**(2020).
207. Long, M., de Souza, S.J., Rosenberg, C. & Gilbert, W. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *PNAS* **93**, 7727-31 (1996).
208. Hsieh, P. *et al.* Evidence for opposing selective forces operating on human-specific duplicated TCAF genes in Neanderthals and humans. *Nat Commun* **12**, 5118 (2021).
209. Sayah, D.M., Sokolskaja, E., Berthoux, L. & Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569-573 (2004).
210. Han, Z. *et al.* New genes helped acorn barnacles adapt to a sessile lifestyle. *Nat Genet* (2024).
211. Rives, N., Lamba, V., Cheng, C.-H.C. & Zhuang, X. Diverse origins of near-identical antifreeze proteins in unrelated fish lineages provide insights into evolutionary mechanisms of new gene birth and protein sequence convergence. *bioRxiv* (2024).
212. Charrier, C. *et al.* Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* **149**, 923-935 (2012).
213. Dennis, M.Y. *et al.* Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* **149**, 912-922 (2012).
214. Schmidt, E.R., Kupferman, J.V., Stackmann, M. & Polleux, F. The human-specific paralogs SRGAP2B and SRGAP2C differentially modulate SRGAP2A-dependent synaptic development. *Scientific reports* **9**, 18692 (2019).
215. Meng, X. *et al.* Brain developmental and cortical connectivity changes in transgenic monkeys carrying the human-specific duplicated gene SRGAP2C. *National Science Review* **10**, nwad281 (2023).
216. Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465-1470 (2015).
217. Florio, M., Namba, T., Pääbo, S., Hiller, M. & Huttner, W.B. A single splice site mutation in human-specific ARHGAP11B causes basal progenitor amplification. *Science advances* **2**, e1601941 (2016).
218. Heide, M. *et al.* Human-specific ARHGAP11B increases size and folding of primate neocortex in the fetal marmoset. *Science* **369**, 546-550 (2020).
219. Fiddes, I.T. *et al.* Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. *Cell* **173**, 1356-1369. e22 (2018).
220. Suzuki, I.K. *et al.* Human-specific NOTCH2NL genes expand cortical neurogenesis through Delta/Notch regulation. *Cell* **173**, 1370-1384. e16 (2018).
221. Florio, M. *et al.* Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *elife* **7**, e32332 (2018).
222. An, N.A. *et al.* De novo genes with an lncRNA origin encode unique human brain developmental functionality. *Nat Ecol Evol* **7**, 264-278 (2023).
223. Niimura, Y. & Nei, M. Extensive Gains and Losses of Olfactory Receptor Genes in Mammalian Evolution. *PLOS One* **2**(2007).
224. Nowick, K., Hamilton, A.T., Zhang, H.M. & Stubbs, L. Rapid Sequence and Expression Divergence Suggest Selection for Novel Function in Primate-Specific KRAB-ZNF Genes. *Molecular Biology and Evolution* **27**, 2606-2617 (2010).
225. Imbeault, M., Helleboid, P.Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550-554 (2017).
226. Turelli, P. *et al.* Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene expression in human neurons. *Sci Adv* **6**, eaba3200 (2020).
227. Surridge, A.K., Osorio, D. & Mundy, N.I. Evolution and selection of trichromatic vision in primates. *Trends in Ecology & Evolution* **18**, 198-205 (2003).
228. Nunes, W.V.B. *et al.* A comprehensive evolutionary scenario for the origin and neofunctionalization of the Drosophila speciation gene Odysseus (OdsH). *G3: Genes, Genomes, Genetics*, jkad299 (2023).

229. Bayes, J.J. & Malik, H.S. Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. *Science* **326**, 1538-41 (2009).
230. Ting, C.T. *et al.* Gene duplication and speciation in Drosophila: evidence from the *Odysseus* locus. *PNAS* **101**, 12232-5 (2004).
231. Sun, S., Ting, C.T. & Wu, C.I. The normal function of a speciation gene and its hybrid sterility effect. *Science* **305**, 81-83 (2004).
232. Ting, C.T., Tsaur, S.C., Wu, M.L. & Wu, C.I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501-1504 (1998).
233. Masly, J.P., Jones, C.D., Noor, M.A.F., Locke, J. & Orr, H.A. Gene transposition as a cause of hybrid sterility in Drosophila. *Science* **313**, 1448-1450 (2006).
234. Wang, C. *et al.* A natural gene drive system confers reproductive isolation in rice. *Cell* **186**, 3577-3592 e18 (2023).
235. Wang, D. *et al.* Two complementary genes in a presence-absence variation contribute to indica-japonica reproductive isolation in rice. *Nat Commun* **14**, 4531 (2023).
236. Liao, B. *et al.* Dysfunction of duplicated pair rice histone acetyltransferases causes segregation distortion and an interspecific reproductive barrier. *Nat Commun* **15**, 996 (2024).
237. Liang, M. *et al.* Taxon-specific, phased siRNAs underlie a speciation locus in monkeyflowers. *Science* **379**, 576-582 (2023).
238. Hirai, H., Maru, Y., Hagiwara, K., Nishida, J. & Takaku, F. A novel putative tyrosine kinase receptor encoded by the eph gene. *Science* **238**, 1717-20 (1987).
239. Schartl, M. A sex chromosomal restriction-fragment-length marker linked to melanoma-determining Tu loci in *Xiphophorus*. *Genetics* **119**, 679-685 (1988).
240. Wittbrodt, J. *et al.* Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in Xiphophorus. *Nature* **341**, 415-21 (1989).
241. Wittbrodt, J., Lammers, R., Malitschek, B., Ullrich, A. & Schartl, M. The Xmrk Receptor Tyrosine Kinase Is Activated in Xiphophorus Malignant-Melanoma. *Embo Journal* **11**, 4239-4246 (1992).
242. Adam, D., Dimitrijevic, N. & Schartl, M. Tumor suppression in Xiphophorus by an accidentally acquired promoter. *Science* **259**, 816-9 (1993).
243. Ding, Y. *et al.* A Young *Drosophila* Duplicate Gene Plays Essential Roles in Spermatogenesis by Regulating Several Y- Linked Male Fertility Genes. *PLOS Genetics* **6**(2010).
244. Chen, S.D., Yang, H.W., Krinsky, B.H., Zhang, A. & Long, M.Y. Roles of young serine-endopeptidase genes in survival and reproduction revealed rapid evolution of phenotypic effects at adult stages. *FLY* **5**, 345-351 (2011).
245. Nurminsky, D.I., Nurminskaya, M.V., De Aguiar, D. & Hartl, D.L. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**, 572-575 (1998).
246. Yeh, S.D. *et al.* Functional evidence that a recently evolved Drosophila sperm-specific gene boosts sperm competition. *PNAS* **109**, 2043-2048 (2012).
247. Saleem, S. *et al. Drosophila melanogaster* p24 trafficking proteins have vital roles in development and reproduction. *Mechanisms of Development* **129**, 177-191 (2012).
248. Chen, S.D. *et al.* Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *EMBO Journal* **31**, 2798-2809 (2012).
249. Krinsky, B.H. *et al.* Rapid Cis-Trans Coevolution Driven by a Novel Gene Retroposed from a Eukaryotic Conserved CCR4-NOT Component in *Drosophila*. *Genes* **13**(2022).
250. Kondo, S. *et al.* New genes often acquire male-specific functions but rarely become essential in. *Genes & Development* **31**, 1841-1846 (2017).
251. Chen, Y., Dai, H.Z., Chen, S.D., Zhang, L.Y. & Long, M.Y. Highly Tissue Specific Expression of *Sphinx* Supports Its Male Courtship Related Role in *Drosophila melanogaster*. *PLOS one* **6**(2011).
252. Dai, H.Z. *et al.* The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. *PNAS* **105**, 7478-7483 (2008).
253. Li, T.P., Zhang, L.W., Li, Y.Q., You, M.S. & Qian, Z. Functional analysis of the orphan genes *Tssor-3* and *Tssor-4* in male *Plutella xylostella*. *Journal of integrative agriculture* **20**, 1880-1888 (2021).
254. Heinen, T., Staubach, F., Häming, D. & Tautz, D. Emergence of a New Gene from an Intergenic Region. *Curr. Biol.* **19**, 1527-1531 (2009).
255. Linnenbrink, M. *et al.* Experimental evaluation of a direct fitness effect of the de novo evolved mouse gene *Pldi*. *bioRxiv*, 2024.01. 13.575362 (2024).
256. Zhang, Z. *et al.* Two young genes reshape a novel interaction network in Brassica napus. *New Phytol* **225**, 530-545 (2020).

257. Xia, S. *et al.* Altered transcription and neofunctionalization of duplicated genes rescue the harmful effects of a chimeric gene in Brassica napus. *The Plant Cell* **28**, 2060-2078 (2016).
258. Schuhmann, P. *et al.* Two wrongs make a right: heat stress reversion of a male-sterile *Brassica napus* line. *Journal of Experimental Botany* **73**, 3531-3551 (2022).
259. Park, J.I. *et al.* Origin of INSL3-mediated testicular descent in therian mammals. *Genome research* **18**, 974-985 (2008).
260. Gubala, A.M. *et al.* The Goddard and Saturn Genes Are Essential for Drosophila Male Fertility and May Have Arisen De Novo. *Mol Biol Evol* **34**, 1066-1082 (2017).
261. Rivard, E.L. *et al*. A putative de novo evolved gene required for spermatid chromatin condensation in Drosophila melanogaster. *PLoS Genet* **17**: e1009787 (2021).